

# Integrative Analysis of ATAC-Seq and Whole-Transcriptome Sequencing in the Trilogy of Gastric Carcinogenesis: From Normal Mucosa to Precancerous Lesions and to Gastric Cancer

**Nianzhi Chen**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Qingsong Liu**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Xiao Ma**

Chengdu University of Traditional Chinese Medicine Wenjiang Campus: Chengdu University of Traditional Chinese Medicine

**Ziyi Zhao**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Jundong Wang**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Yu Chen**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Maoyuan Zhao**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Qiao Zheng**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Gang Zhang**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Ting Xia**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Yu Guo**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Yun Yuan**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Shangbin Lv**

Chengdu University of Traditional Chinese Medicine Wenjiang Campus: Chengdu University of Traditional Chinese Medicine

**Yulin Qi**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Zhongzhen Zeng**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Jinhao Zeng**

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

**Jianyuan Tang** (✉ [tangjy@cdutcm.edu.cn](mailto:tangjy@cdutcm.edu.cn))

Chengdu University of Traditional Chinese Medicine Affiliated Hospital

---

## Research

**Keywords:** early gastric cancer, gastric precancerous lesions, biomarkers

**Posted Date:** September 28th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-910498/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background:

Many specific biomarkers and target genes of gastric cancer have been identified, but the research areas on the characterization of epigenetic and transcriptional landscape in the process of gastric carcinogenesis, known as the trilogy from normal mucosa to precancerous lesions and finally to gastric cancer, remains a black box.

## Results:

We performed whole-transcriptome sequencing and ATAC-seq based on human gastric tumours (T), gastric precancerous lesions (P) and normal gastric mucosa (N) and obtained chromatin open regions, circRNAs, lncRNAs, miRNAs, and mRNAs. Subsequently, differential expression analysis was performed on the N, P, and T groups, followed by GO, KEGG, and PPI analyses to investigate cancer-related hub genes (e.g., VEGFA, FN1, CDK1, and MYC) and extracellular matrix related signalling pathways that might serve as potential therapeutic targets in precancerous lesions and early gastric cancer. A core competing endogenous RNA (ceRNA) network was then constructed that contained 17 mRNAs, 11 miRNAs, 11 lncRNAs and two circRNAs, and the RNAs were validated by TCGA database and RT-PCR to prove they play crucial roles in the trilogy of gastric carcinogenesis; notably, hsa-miR-1226-5p, hsa-miR-6720-5p, lncRNA H19, ARID3A and GPR161 were significantly enriched in the complex network, which might be specificity tumour markers. Finally, ROC curve analysis suggested that nine mRNAs and five lncRNAs may be novel biomarkers with clinical implications for the precise diagnosis of early gastric cancer. Survivorship curve analysis further identified four mRNAs and one lncRNA as potential prognostic indicators.

## Conclusions:

These results reveal a panel of premalignant and malignant specific marker genes in the trilogy of gastric carcinogenesis, potentially aiding the early identification of precancerous lesions and early gastric cancer. Meanwhile, the present study identifies the key molecular targets in the dynamic process of gastric carcinogenesis for experimental research on drugs and molecular mechanisms.

## Background

Gastric cancer (GC) is one of the most common cancers in the world. The incidence rate of gastric cancer ranks fifth, and the mortality rate ranks third among common cancers globally, with increasing incidence and mortality in China. [1, 2] According to the Correa model, human gastric carcinogenesis is a multistep process. Chronic atrophic gastritis (CAG), intestinal metaplasia (IM) and dysplasia are the main premalignant lesions of intestinal-type gastric cancer (GC). [3, 4] These premalignant lesions are collectively called gastric precancerous lesions (GPLs), which have the highest risk of progression to early gastric cancer (EGC).[5] Generally, this is a trilogy of the complete process of transition from normal

gastric mucosa to precancerous lesions and EGC. It is important to investigate how progression from GPLs to EGC can be inhibited via prevention and treatment. Additionally, elucidating the molecular mechanisms underlying GPLs is essential. Therefore, it is vitally important for researchers to unveil authentic molecular mechanisms representing early neoplastic stages, such as gastric precancerous lesions, and to unveil the mechanisms of transitioning from normal gastric mucosa to premalignant lesions and to early gastric cancer, to improve the cure rate of EGC. However, specific molecular biomarkers of the overall process for predicting the transition from normal gastric mucosa to GPLs and to EGC remain unclear despite the molecular mechanisms of GC being widely studied.[6]

A large number of noncoding RNAs (ncRNAs) have been found to be involved in the control of gene expression and activation pathways, including the processes of activation and inhibition of the development of cancer.[7] In recent years, multiple studies have reported the aberrant expression of ncRNAs in many human diseases, including gastric cancer.[8] For instance, microRNAs are small noncoding RNAs that regulate gene expression by controlling target mRNAs.[9] MicroRNAs (miRNAs) have been proven to be the key in regulating important growth regulatory pathways of cancer pathogenesis in many studies.[10] Growing evidence shows that miRNAs participate in crucial cell processes of gastric tumours, including gastric tumour cell differentiation, proliferation, apoptosis, migration, invasion and metastasis.[11] Moreover, the dysregulation function of long noncoding RNAs (lncRNAs) has also been implicated in gastric cancer according to emerging evidence.[12] For instance, lncRNA-MT1JP acts as a potential therapeutic target by competitively binding to miRNA in gastric cancer and has become an essential epigenetic regulator that plays a critical role in gastric cancer.[13, 14] Circular RNAs (circRNAs) are a new class of regulatory RNAs that have been explored in recent years.[15] As a rising star, circRNAs show great potential as novel biomarkers of gastric cancer due to their unique structures and features.[16] However, the "miRNA sponge" is the most common phenomenon, in which circRNAs and lncRNAs play crucial roles in the progression of tumours. These ncRNAs can act as ceRNAs to affect the expression levels of target genes and eventually act as tumour promoters in GC tumours. [17–19] Competing endogenous RNAs (ceRNAs) represent a regulatory mechanism by which ncRNAs can naturally competitively bind miRNAs, thereby inhibiting the regulation of target genes by miRNAs, such as lncRNAs and circRNAs, and forming complex miRNA-mediated ceRNA networks.[20, 21] The underlying mechanism of ceRNAs was determined by bioinformatics analysis. The ceRNA network of GC has been reported in many types of research, but the mechanism of GPLs and EGC is rarely reported. Thus, it is of enormous biological significance to explore the mechanisms of ceRNA in GPLs and gastric carcinogenesis.[22]

Currently, the Assay of Transposase Accessible Chromatin sequencing (ATAC-seq), as a next-generation technology, is available to predict the molecular targets of disease. Meanwhile, RNA-seq methods are available for studying many different aspects of RNA biology with the development of next-generation sequencing.[23] In addition, ATAC-seq has been widely used in studying chromatin biology, and it is an approach developed to decipher the epigenetic landscape and systematically analyse chromatin open regions throughout the genome by Tn5 transposase.[24] The combination of these two methods can

more comprehensively identify molecular biomarkers from epigenetics and genetics and more accurately identify the target genes and molecular mechanisms of EGC.

To identify potential novel clinical therapeutic targets for EGC and gastric precancerous lesions and systematically complete the data of the whole process of transition from normal gastric mucous membrane to GPLs and EGC to gain more insights into the molecular mechanisms of this cancer, we performed whole-transcriptome sequencing and ATAC-seq based on tumours (T) from 30 early gastric cancer patients, the gastric mucosa (P) from gastric precancerous lesion patients, and the normal gastric mucous membrane (N) from 30 participants. Then, differential expression analysis was performed between the T and P groups and between the N and T groups. Subsequently, we analysed functional interaction prediction to investigate gene-regulatory circuits. Furthermore, The Cancer Genome Atlas (TCGA) for gastric cancer data and real-time PCR were used to validate the results. Overall, our research results may provide novel perspectives to clarify the mechanisms of the transition from premalignant gastric tissue to EGC for the early diagnosis of gastric cancer and suggest some genes as potential therapeutic targets in GPLs and EGC.

## Results

### 1. Differential Expression Analysis

According to the screening criteria and method, a total of 2,457 differentially expressed mRNAs (dif-mRNAs) in GPLs compared with EGC (P versus T) were obtained, along with 2,857 dif-mRNAs in normal tissues compared with EGC (N versus T), of which 2,548 were upregulated and 2,766 were downregulated (Figure 1A-1B). The overlapping dif-mRNAs between P versus T and N versus T were filtered out using Venn diagrams (Figure 1C). To further integrate epigenetic elements into the results of dif-mRNAs, we obtained high-quality reads from the ATAC-seq with alignment of the reference genome, obtained the position information of the reads on the genome and selected an intersection for the dif-mRNAs and the aligned genes by ATAC-seq. Finally, the DEGs were formed. Among a total of 2,205 DEGs in P versus T and 2,578 DEGs in N versus T, the overlapping DEGs between the two pairwise groups were presented by Venn diagrams and hierarchical cluster heatmaps (Figure 1E-1H), and 1,705 DEGs overlapped in total. In particular, 11 DEGs were upregulated in NvsT but downregulated in PvsT, including CPS1, CYP2D6, APOB, CHST5, ANPEP, CLCA1, SLC6A19, MME, CDHR5, SLC5A1, and MALL. Most of these mRNAs maybe play an essential role in the occurrence and development of tumours, which is worthy of further exploration to determine whether these are the target mechanisms of EGC.

A total of 352 differentially expressed miRNAs (dif-miRNAs), 3,715 differentially expressed mRNAs (dif-lncRNAs), and 117 differentially expressed circRNAs (dif-circRNAs) were obtained in P versus T, and 455 dif-miRNAs, 4,404 dif-lncRNAs, and 208 dif-circRNAs were obtained in N versus T (Figures 2A–2F). The overlapping dif-miRNAs (207), dif-lncRNAs (2,235), and dif-circRNAs (59) between P versus T and N versus T were filtered out using Venn diagrams (see Figure 2G-2I). The overlapping dif-miRNAs, dif-lncRNAs, and dif-circRNAs are shown in hierarchical cluster heatmaps (Figure 2J-2L), from which we

found significant separation between the three groups, indicating that the results of the differential expression analysis were available. The primary point here is that hsa-miR-215–5p levels were elevated in normal versus EGC tissues, whereas levels were decreased in both GPLs versus EGC samples which might be a potential target microRNA.

## 2. Functional Enrichment Analysis of DEGs

Functional enrichment analysis of the overlapping DEGs was performed to explore the biological functions. Upregulated genes and downregulated genes were enriched in the top 30 GO terms and top 20 KEGG pathways, respectively (Figure 1I-1L). According to the GO analysis enriched by the upregulated genes, we found that many pathways of gastric carcinogenesis involve abnormal physiological function of the extracellular matrix, which indicates that extracellular matrix structural constituents and collagen manifest the critical processes of transition from normal to GPLs and further to EGC. Among the top 20 KEGG pathways based on the upregulated genes, the PI3K-Akt signalling pathway, proteoglycans in cancer, cell cycle, p53 signalling pathway, etc., were related to gastric carcinogenesis. Similarly, the KEGG pathways of downregulated genes suggest that the calcium signalling pathway, gastric acid secretion and chemical carcinogenesis-DNA adducts may be involved in the occurrence of EGC.

## 3. Protein–Protein Interaction (PPI) Analysis of DEGs and dif-ncRNAs

To identify important hub genes among the overlapping DEGs, dif-miRNAs, dif-lncRNAs, and dif-circRNAs, we performed protein–protein interaction (PPI) analysis of all the differentially expressed RNAs. A total of 1,555 DEGs were generated, the degree of each node was calculated using CytoNCA and visualized by the R packages, and the nodes with high topological scores can be regarded as vital nodes of the network, with the hub-gene determined according to the number of adjacent nodes. A bar chart including the top 30 hub genes (e.g., VEGFA, FN1, CDK1, MYC, etc.) is shown in (Figure 3B). Meanwhile, PPI networks of the target genes in DEGs corresponding to dif-miRNAs, dif-lncRNAs, and dif-circRNAs were also established (Figure 4A-4C). The PPI network based on dif-miRNAs consisted of 1,390 genes and a total of 999 genes based on dif-lncRNAs, while only 13 DEGs corresponded to dif-circRNAs. We generated bar charts of the top 20 target genes (Figure 4D-4F). The target genes (e.g., FN1, MYC, VEGFA, CDK1, etc.) corresponding to the dif-miRNAs and dif-lncRNAs with a significantly higher degree (e.g., DNMT3B, HIST1H2BJ, HIST1H2N, etc.) were found to be associated with dif-circRNAs. We then extracted five clustered modules from the PPI network and formed five subnetworks, the details of the representative genes ranked by degree were shown in Table 1. The five modules contained 96 (Module 1), 82 (Module 2), 77 (Module 3), 48 (Module 4), and 49 (Module 5) nodes, among which the upregulated genes were most often included. Among the 17 genes in Module 1 with degree=95 (e.g., PBK, KIF23, TOP2A, CDK1, CDC45, etc.), FN1 with the highest degree in Module 2 (degree=48), ICAM1 with degree=23 in Module 3 with the same circumstance, THBS1 with twice the degree of the second gene in Module 4 (degree=30), VEGFA is the highest in Module 5 (degree=21).

Moreover, genes in the five modules were subjected to GO and KEGG enrichment analysis (Figure 3H-3I). According to the KEGG analysis, four main pathways may be involved in the occurrence and development

of gastric carcinogenesis: the cell cycle, DNA replication, PI3K-AKT signalling pathway, and ECM-receptor interaction. We believed that changes in the extracellular matrix and extracellular structure organization were probably involved in the pathological processes of GPLs and EGC.

#### 4. ceRNA Network Construction

First, we constructed a circRNA-related and lncRNA-related ceRNA network. Based on the theory of ceRNA and the regulatory relationships of dif-circRNA-dif-miRNA, dif-lncRNA-dif-miRNA, and dif-mRNA-dif-miRNA, the differentially expressed circRNA, lncRNA, and mRNA regulated by the same miRNA were screened. By using circRNA and lncRNA as decoys, circRNA-miRNA-mRNA and lncRNA-miRNA-mRNA networks were formed (Figure 4G-4H). The circRNA-miRNA-mRNA ceRNA network contained 17 mRNAs, two circRNAs, and 11 miRNAs (Figure 4G). In this network, the core circRNA is an unknown gene expressed as gene loci information, and nodes and edges represent genes and interactions, respectively. The lncRNA-miRNA-mRNA ceRNA network contained 5,916 edges, 366 mRNAs, 312 lncRNAs, and 55 miRNAs (Figure 4H). Since most of the lncRNAs (approximately 283 lncRNA transcripts) were expressed as gene loci information, we created a core ceRNA network that filtered out the unknown lncRNAs, which contained 2,762 edges, 366 mRNAs, 29 lncRNAs, and 55 miRNAs (Figure 4I).

According to the two subnetworks and the regulatory relationship of these differentially expressed ncRNAs, we constructed a circRNA-lncRNA-miRNA-mRNA ceRNA network containing 1,013 edges, 17 mRNAs, 11 miRNAs, 242 lncRNAs, and 2 circRNAs (Figure 5A) and then filtered out 223 unknown lncRNAs to form a core network containing 17 mRNAs, 11 miRNAs, 11 lncRNAs, and two circRNAs (Figure 5B), the representative RNAs in the three big ceRNA networks were shown in Table 2.

In all of the above networks, we found that in both the whole network and the core network, hsa-miR-1226-5p and hsa-miR-6720-5p were the most important miRNAs. Furthermore, we found two mRNAs of GPR161 and ARID3A: lncRNA H19-202 had the highest degree (degree=12) in the core ceRNA network. We performed GO and KEGG enrichment analysis of all the genes in the core ceRNA network. The results showed that the smoothed signalling pathway was significantly enriched in GO-BP terms, and cytokine receptor, immune receptor, G protein-coupled peptide receptor, and peptide receptor activity were significant in GO-MF terms (Figure 6A). The results of KEGG enrichment showed that the classical MAPK signalling pathway and cytokine-cytokine receptor interaction pathway were significant (Figure 6B).

#### 5. Validation of The Cancer Genome Atlas (TCGA) Database

To validate our findings in the core ceRNA network, the RNA expression levels were compared with the TCGA database. Because the circRNAs represented gene locus information that could not be verified, we were more concerned with mRNA, miRNA, and lncRNA validation. Because H19-201, H19-202, and H19-203 were the isomers of lncRNA-H19, the validation of only nine lncRNAs was performed. The validation results are shown in Figure 6C-6E. Most of the ceRNA expression patterns in the network were consistent with the TCGA data. There were ten mRNAs (Figure 6C), eight lncRNAs (Figure 6D), and four miRNAs (Figure 6E) that were significantly different, indicating participation in the development of EGC.

## 6. Survivorship Curve Analysis and ROC Curve Analysis

To further determine the roles of these genes in the ceRNA network with respect to survival and prognosis, we performed a survivorship curve analysis of mRNAs, miRNAs, and lncRNAs. The results showed that four mRNAs and one lncRNA were significant (Figure 7A-7E). For example, CCR10 (Figure 7A), GSDME (Figure 7C), LRRN1 (Figure 7D), and AC136475.3 (Figure 7E) had higher expression, and GPC3 (Figure 7B) had lower expression, the overall survival rate and prognosis were better. Literature retrieval revealed that most of these RNAs play a role in cancer progression, suggesting potential targets for the above RNAs in GPLs and EGC. Therefore, we performed ROC curve analysis to evaluate the diagnostic value of the candidate ceRNAs in GPL and EGC. The results showed that the AUC values of nine mRNAs (e.g., UBAP2, NOL6, LGR5, ASCL2, IL1RAP, etc.) and five lncRNAs (e.g., DLGAP1-AS2, CYTOR, AC136475.3, AGAP2-AS1, and AP003774.1.) were more significant than 0.7 (Figure 6F-6G). These results demonstrated that the above mRNAs and lncRNAs have good diagnostic value in GPLs and EGC, and thus serve as early biomarkers to predict GPLs and EGC.

## 7. Real Time-PCR Validation

Based on the results of the ceRNA network, most miRNAs except hsa-miR-3180-3p and hsa-miR-3180-5p existed in the TCGA database, and we found that hsa-miR-1226-5p, hsa-miR-6720-5p, lncRNA H19, ARID3A and GPR161 were in critical core positions, but hsa-miR-1226-5p was not significantly different in the TCGA database. The three mRNAs of GSDME, LRRN1, and GPC3 had certain significance with respect to the overall survival rate and prognosis, but they were not significant in TCGA. We further validated the three miRNAs and the three mRNAs by real-time PCR in normal, GPLs and EGC tissues to prove our data meaningful. In the results depicted in (Figure 7F-7K), there are significant differences in the expression of GPC3 (Figure 7F), GSDEM (Figure 7G), LRRN1 (Figure 7H), miR-3180-3p (Figure 7I), miR-3180-5p (Figure 7J) and miR-1226-5p (Figure 7K) between the three groups of normal gastric mucosal tissues, GPL tissues, EGC tissues, which confirms that the RNA sequencing and real-time PCR results were consistent.

## Discussion

The molecular targets of gastric cancer are continually increasing; nonetheless, the prognosis and mortality remain dismal in Western countries despite declining incidences.[25, 26] Gastric cancer is not a “jump” from normal epithelial cells to cancer cells: the pathogenesis of gastric cancer is a multistep carcinogenesis process.[27] Gastric precancerous lesions (GPLs) are a classic precancerous state of the stomach, and GPLs provide a critical window to prevent and treat gastric cancer. Therefore, specific molecules are urgently needed to represent and elucidate the causes of GPLs. In recent years, epigenetics has become popular, indicating that gene function has undergone heritable changes that eventually lead to phenotypic changes without changes in gene DNA sequences. However, little is known about epigenetic alterations influencing GPLs pathogenesis and gastric carcinogenesis.[28] Thus, the main focus of this study is to identify the specific biomarkers that can be employed to predict this global

process of transition from normal gastric mucosa to GPLs and further to EGC from multiple perspectives and in the context of numerous methods.

ATAC-seq can identify open chromatin regions and capture regulatory sequence information. Open chromatin regions can bind regulatory proteins and regulate gene expression, resulting in epigenetic changes.[29] Therefore, this study established whole-transcriptome sequencing and ATAC-seq on GPLs, EGC, and normal gastric mucosa and obtained differential expression profiles such as those of lncRNAs, circRNAs, miRNAs, and mRNAs. In the present study, we integrated gene sequence changes and epigenetic information into the differential mRNAs. Therefore, the differential mRNAs and the reference genome produced by ATAC-seq were cross-screened, and the final DEGs were the target genes with two simultaneous factors. Subsequently, integrated functional enrichment analysis and the establishment of protein–protein interactions (PPIs) were used to identify candidate genes. We identified several candidate genes by focusing on the DEG degree values in PPI analysis, which commonly changed in the three groups, including classic oncogenes such as VEGFA, CDK1, FN1, MYC, AURKA, TOP2A, etc. Combining the PPI networks of the genes targeted by dif-lncRNAs and dif-miRNAs, we found that four genes, VEGFA, FN1, CDK1 and MYC, were the key promoter genes from the comprehensive gene degree data. According to the genes targeted by dif-circRNAs, we found that DNMT3B, HIST1H2BJ, and HIST1H2N might be involved in the process of gastric carcinogenesis. In addition to the cell cycle, p53 signalling pathway, and MAPK and PI3K-AKT classical signalling pathways, the mechanism of abnormal physiological function of the extracellular matrix is closely related to EGC and GPLs according to functional enrichment analysis.

Vascular endothelial growth factor A (VEGFA) can promote neovascularization and increase vascular permeability, acting as a key factor in angiogenesis.[30] Many tumour angiogenesis events primarily rely on VEGFA-driven responses, and angiogenesis is a vital process in the progression of gastric adenocarcinoma and is necessary for gastric tumour growth and metastasis.[31] VEGF-A inhibition has become a therapeutic strategy for many cancers, and several anti-VEGF-A strategies have been examined for gastric cancer. However, very few drugs can be used in the relevant clinical setting.[32] In addition, there is also evidence indicating that VEGFA expressed in gastric mucosa is related to intestinal metaplasia or dysplasia, so VEGFA should be represented as a critical gene in the complete process of gastric carcinogenesis and necessitates further development of relevant drugs.[33, 34] Fibronectin 1 (FN1), as a marker of EMT, plays an essential role in the cell adhesion, invasion, migration and metastasis of tumours.[35] Mechanistically, there are also a great deal of academic studies suggesting that FN1 is a novel target for many types of cancers, such as gastrointestinal cancer.[36, 37] FN1 can promote the malignant progression of gastric cancer cells through ECM remodelling and suppress the proliferation, migration and invasion of GC cells by downregulation in gastric cancer. It is significantly related to poor prognosis of gastric adenocarcinoma, which is the final cancerous result of gastric precancerous lesions.[38–40] The MYC gene is a key classical gene in tumorigenesis in various types of tissues that contributes to gastric cancer progression, and it can be mediated by ncRNAs and may be associated with cell proliferation, survival, and drug resistance.[41–43] Significantly higher MYC mRNA expression was observed in atrophy and dysplasia and intestinal metaplasia compared to gastritis

samples.[44, 45] CDK1 is a cyclin-dependent kinase and often mediates cell cycle defects associated with many tumours, and emerging evidence suggests that it is mainly involved in prostate cancer, small cell lung cancer, and GC.[46, 47] Moreover, based on the results of the genes targeted by dif-circRNAs, the DNA methyltransferase 3B (DNMT3B) gene has been associated with the occurrence or decrease of various tumours. Some studies have revealed that DNMT3B promotes GC tumorigenesis as an oncogene.[48, 49] Some studies provide insights into the molecular mechanisms that regulate the HIST1H2BJ gene, which may be related to the progression of cancer, but the HIST1H2N gene has rarely been reported.[50, 51] Moreover, many studies have examined the expression patterns of ECM components and signalling pathways in normal gastric epithelium, premalignant lesions, and gastric cancer. ECM-associated molecules can be used as predictive biomarkers of disease or as potential targets in gastric cancer.[52] We believe that these four genes (e.g., VEGFA, FN1, CDK1, and MYC), ECM components and signalling pathways play vital roles in the comprehensive processes of gastric cancer. More related studies are merited. Our findings may contribute to the development of targeted drugs for EGC and GPLs.

In the present study, we constructed a core circRNA-lncRNA-miRNA-mRNA ceRNA network using candidate RNAs that overlapped between normal tissues, GPLs and EGC. Most of these representative RNAs in the network were validated by TCGA database and RT-PCR to prove the reliability of our data. Although little is known about the above two circRNAs, their related network miRNAs, lncRNAs, and mRNAs have been revealed to serve potential roles in EGC. Through ROC curve analysis, most of these lncRNAs and mRNAs in this core ceRNA network were found to be significant in predictive diagnosis and useful as novel diagnostic markers and therapeutic targets. Survivorship curve analysis also demonstrated that several indicators (e.g., CCR10, GSDME, LRRN1, AC136475.3, and GPC3) can act together as a prognostic index to evaluate EGC. LRRN1 and GSDME expression represent loss of regulation in EGC tissues, and high LRRN1 expression is associated with poor prognosis, which plays an important role in EGC. Glypican-3 (GPC3) is expressed in liver cancers, and antibody-based therapies targeting glypicans are being investigated in preclinical and clinical studies, which may be further applicable to gastric cancer.[53–55] We also note that hsa-miR-1226–5p and hsa-miR-6720–5p represent the centre of the network, and that GPR161, ARID3A, and lncRNA-H19 are of great importance in both GPLs and EGC. Evidence in the literature shows that miR-1226, which acts as a specific tumour suppressor, affects many tumours, regulates the expression of oncoproteins and induces cell death.[56–58] MiR-6720 might affect cancer via the hedgehog signalling pathway and calcium signalling pathway.[59] The G-protein-coupled receptor GPR161 is overexpressed in cancer and is a promoter of cell proliferation and invasion that can suppress the Sonic hedgehog pathway.[60, 61] ARID3A, which is a paralogue from the AT-rich interactive domain (ARID) family, has been found to be involved in cancer and stem cell processes and associated with colorectal carcinoma.[62, 63] H19 is a lncRNA that is described as an oncofoetal transcript and has been associated with the risk of hepatocellular carcinoma, breast cancer, bladder cancer, gastric cancer and colorectal cancer.[64, 65]

Our research clarifies the mechanism of the entire process of gastric transition from normal to gastric premalignant tissue and cancer from a new perspective, contributing to the early diagnosis of gastric cancer by identifying transcripts (e.g., hsa-miR-1226–5p, hsa-miR-6720–5p, H19, etc.) as potential

therapeutic targets in early gastric cancer and gastric precancerous lesions. The above results combine epigenetic and genetic factors, which can provide researchers with multiple directions and perspectives for further experimental research on drugs and molecular mechanisms. Furthermore, our findings suggest that most of these biomarkers are associated with the development of EGC, although research on modification, modulation, and mutual mediation between these ncRNAs remains insufficient. Further research could build on our findings from data analysis to further unveil the mechanisms of these ncRNAs in EGC and precancerous lesions from the perspective of epigenetics, determine how these genes interact with each other, and reveal their relationships with EGC. Similarly, few experiments have investigated the molecular mechanisms of precancerous gastric lesions, and these genes and pathways can be further used as novel targets for the treatment of GPLs to prevent both their occurrence and the development of gastric carcinogenesis.

In conclusion, these results reveal the key promoter biomarkers in the entire dynamic process of gastric carcinogenesis and will help us acquire a more comprehensive understanding of the development of EGC. This study provides novel and reliable biomarkers for EGC and GPLs, which will be useful for further clinical applications in diagnosis, prognosis and targeted therapy.

## Methods

### 1. Participants and sample collection

A total of 90 samples were collected (30 early gastric cancer tissues from patients who underwent operative treatment, 30 gastric precancerous lesion tissues and 30 normal gastric mucous membrane tissues from mucosal biopsy specimens from the gastric tract) by gastroscopy (N group: 14 males and 16 females, P group: 16 males and 14 females, T group: 18 males and 12 females, age<60: N 24 people, P 19 people, T 17 people, age≥60: N 6 people, P 11 people, T 13 people, mean age: N 53.5 years, P 56 years, T 56.5 years) at the Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, between 2018.12–2020.9 from patients with no preoperative radiation or chemotherapy or drug treatment. All specimens were snap frozen in liquid nitrogen and stored at -80 °C. This study was approved by the Ethics Committee of the Hospital of Chengdu University of Traditional Chinese Medicine (approval no. 2018KL-023), and informed consent was obtained from all participants.

### 2. RNA Extraction and quality control

Because sufficient RNA was required for RNA sequencing but the tissues were small, 3 tissues were used as one sample for sequencing, and total RNA was extracted from three tissues of three groups by using TRIzol reagent (Life Technologies CA, USA) following the manufacturer's instructions. RNA degradation and contamination were monitored on 1.5% agarose gels. RNA concentration and purity were measured using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). RNA integrity was assessed by using the RNA Nano 6000 Assay Kit for the Agilent Bioanalyzer 2100 System (Agilent Technologies, CA, USA).

### 3. RNA-Seq and ATAC-seq data

RNA sequencing libraries were generated using the NEBNextR Ultra™ Directional RNA Library Prep Kit for Illumina R (NEB, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. To preferentially select insert fragments 150-200 bp in length, the library fragments were purified with AMPure XP Beads (Beckman Coulter, Beverly, USA). Finally, library quality was assessed on an Agilent Bioanalyzer 2100 and through qPCR. After cluster generation, the library preparations were sequenced on an Illumina platform, and reads were generated. ATAC was performed according to the ATAC-seq protocol by Biomarker Technologies Biotechnology Ltd. The Nextera DNA Library Preparation Kit (Illumina) was used to perform transposition according to the manufacturer's manual. The transposed DNA fragments were purified immediately with a MinElute PCR Purification Kit (Qiagen). Subsequent libraries were subjected to sequencing on an Illumina NovaSeq 6000 using PE150. Finally, RNA sequencing and ATAC-seq data were acquired. In brief, we prepared a sequencing database for the gene regulatory sequences of chromatin open regions, lncRNAs, circRNAs, miRNAs and mRNAs.

### 4. Screening of differentially expressed RNAs

mRNAs, miRNAs, lncRNAs, and circRNAs were divided into pairwise groups (P versus T and N versus T). Differential expression analysis of these groups was performed using the EdgeR package (4.0.2). Genes with an adjusted P value <0.05 and an absolute value of  $\log_2$  (fold change) >1 found by EdgeR were assigned as differentially expressed. Then, we used Bowtie2 software to compare the high-quality reads obtained from the ATAC-seq of each sample with the reference genome to determine the alignment efficiency of the sample reads and the position information of the reads on the genome and performed peak extraction by MACS2 v2.1.1 software. To identify the openness of chromatin and ensure high-quality data, DeepTools v2.07 was used to map the density distribution of sequencing reads in the 3 kb intervals upstream and downstream of the TSS of each gene: the results are presented as heatmaps. For the purpose of further integrating epigenetic elements into the results of dif-mRNAs, we selected an intersection for dif-mRNAs and the aligned genes based on the ATAC-seq data, and then the final DEGs were formed. The differentially expressed RNAs between pairwise groups were visualized on volcano plots and structured by plotting  $-\log_{10}$  (p value) on the y-axis and  $\log_2$  (fold change) on the x-axis. Overlapping differentially expressed RNAs between the pairwise groups were determined by Venn diagrams and subjected to further analysis. Heatmaps were generated by the heatmap R package to exhibit the expression patterns in pairwise groups of DEGs, dif-miRNAs, dif-lncRNAs, and dif-circRNAs.

### 5. DEGs functional enrichment analysis

The screening of differentially expressed mRNAs (DEGs) was based on ATAC-seq data. To further clarify the biological mechanism of DEGs in the course of disease development, Gene Ontology (GO) [66] and KEGG pathway enrichment [67, 68] analyses of the up- and downregulated DEGs were carried out by the cluster Profiler R packages (<http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html>), respectively, and the

threshold  $p < 0.05$  was considered significant. The GO enrichment analysis included three parts: BP (biological process), CC (cellular component) and MF (molecular function). The terms were selected for display according to the significance order, and we constructed bubble charts for the top 10 GO items and the top 20 KEGG items of the upregulated and downregulated DEGs. Subsequently, when we established the PPI network of DEGs and constructed the core ceRNA network, we performed GO and KEGG analysis on the mRNAs in the networks, where the threshold was set to  $P$  less than 0.05.

## 6. PPI and Module Analysis of DEGs, dif-miRNAs, dif-lncRNAs and dif-circRNAs

The PPI analysis of DEGs was performed using the STRING database (version 11.5, <https://www.string-db.org/>). [69] The DEGs were imported into the STRING database, the species was set as human, the confidence level was set to 0.4, dissociated nodes were hidden, the TSV format file was downloaded, and the file was imported into Cytoscape software V3.8.0 (<https://cytoscape.org/>). We then constructed a protein interaction network (PPI) by Cytoscape software, and R software was used to count the number of adjacent nodes of each gene in the PPI network. Then, we generated a histogram of the top 30 hub genes from the network. In addition, a score  $\geq 6$  was set up in the MCODE plug-in (<https://apps.cytoscape.org/apps/MCODE>) to cluster the DEGs of the PPI network and obtain five subnetworks. GO and KEGG enrichment analyses were also performed on the genes in the significantly clustered modules, and the threshold  $p < 0.05$  was considered significant. At the same time, we also performed the PPI networks of the targeted DEGs corresponding to dif-miRNAs, dif-lncRNAs and dif-circRNAs, and plotted the top 20 hub genes in DEGs corresponding to dif-miRNAs and dif-lncRNAs, while the DEGs corresponding to dif-circRNAs only numbered 13.

## 7. Construction of ceRNA Network

It is generally known that miRNAs can lead to gene silencing by binding to mRNA, while some ncRNAs can act as ceRNAs that can counteract the repressive activity of miRNAs: a ceRNA network is then formed. [70] To construct a ceRNA network of molecular targets throughout the entire dynamic process of gastric carcinogenesis. First, we find the predicted DEGs-bound miRNAs. We also integrated the dif-lncRNA regulatory relationships to find the predicted miRNAs for targeted binding by dif-lncRNAs. Similarly, the miRNAs bound to dif-circRNAs were predicted at the same time.

Subsequently, we devoted more attention to the correlation expression of lncRNA-mRNA and circRNA-mRNA, so we focused on miRNAs that can simultaneously regulate mRNA and lncRNA or circRNA. The relationships of lncRNA-miRNA-mRNA and circRNA-miRNA-mRNA were extracted and introduced into Cytoscape V3.5.0 (San Diego, CA, USA), and the correlation coexpression relationship between these miRNA-regulated mRNAs and ncRNAs was further obtained. Cytoscape was used to construct the lncRNA-miRNA-mRNA and circRNA-miRNA-mRNA complex networks.

Alternatively, these two complex networks of circRNA-miRNA-mRNA and lncRNA-miRNA-mRNA were integrated, and then we focused on screening miRNAs that can simultaneously regulate circRNA, lncRNA, and mRNA. Correlation coexpression relationships of mRNA, lncRNA, and circRNA regulated by these

miRNAs were further obtained. Finally, the circRNA-lncRNA-miRNA-mRNA ceRNA network was constructed by Cytoscape. Among them, the spliced transcripts were compared with known lncRNA databases (lncRNADB (<http://lncrnadb.org>) and NONCODE (<http://www.noncode.org/>) to identify the known lncRNAs in the sample. Then we filtered out unknown lncRNAs to form a core ceRNA network.

## 8. TCGA Gastric Cancer Database Validation and RT-PCR Validation

Two circRNAs in the ceRNA network could not be verified by TCGA (The Cancer Genome Atlas) because all the circRNAs were expressed as unknown gene loci, so we were more concerned with the TCGA validation of mRNAs, miRNAs and lncRNAs. The mRNA, miRNA and lncRNA expression matrix and annotation information were loaded in the TCGA database for data verification in this study. The mRNA and lncRNA data of gastric cancer contained 375 tumour samples and 32 adjacent normal samples, and the miRNA data of gastric cancer contained 446 tumour samples and 45 adjacent normal samples. We examined the expression levels of all the mRNAs, miRNAs and lncRNAs in the ceRNA network in TCGA gastric cancer data by using the online tool starBase and then generated a visual box diagram of differential RNAs by the R software ggpubr package, where  $P \leq 0.05$  was considered a significant difference.[71] Among them, the mRNAs, miRNAs and lncRNAs, which were valuable but not significantly different or did not exist in the TCGA database, were further validated by real-time PCR. The RT-PCR validation was applied for newly collected human EGC tissues, precancerous lesions tissues, and normal gastric mucous tissues. Total RNA from three newly collected human tissues as one group was isolated by using the Trizol reagent (Life Technologies CA, USA), and then reverse-transcribed into cDNA according to the manufacturer's instruction. The data represent three times experiments. The primers were purchased by the company (GeneCopoeia, USA), the item codes of these RNAs were GPC3 (Cat#HQP007587), GSDME (Cat#HQP004226), LRRN1 (Cat#HQP015766), GAPDH (Cat#HQP006940), miR-3180-3p (Cat#HmiRQP1638), miR-3180-5p (Cat#HmiRQP1511), miR-1226-5p (Cat#HmiRQP0061), RNU6-2 (Cat#HmiRQP9001), respectively.

## 9. Survivorship Curve Analysis and ROC Curve Analysis

Survivorship curve analysis of the mRNAs, miRNAs and lncRNAs of the ceRNA network was performed by R software survival and the survminer package, and the generation curve of  $P < 0.05$  was visualized.[72] When the ROC curves were generated, the AUC was calculated, and an  $AUC > 0.7$  was considered as meaningful. Then, the ROC curve of the prognostic mRNAs in TCGA was drawn by the R software pROC package.[73] We also analysed the ROC curves of miRNAs and lncRNAs in the network to assess the diagnostic value of ceRNA expression for discriminating GPL/EGC from normal tissues, however, the miRNAs were not significant.

## Declarations

### Acknowledgements:

We thank Xian Chen and Daoyin Gong (Department of Pathology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China) for assistance with pathological diagnosis of human tissues. We are grateful to Xiaodong Chen (Department of Gastrointestinal surgery, Sichuan Cancer Hospital, Chengdu, China) for provide methods for collecting human tissues.

### **Funding:**

This work was supported by the National Natural Science Foundation of China (grant nos. 82174346, 81804066 and 81904178), and by Xinglin Scholar Research Promotion Project of Chengdu University of TCM (grant nos. QNXZ2019017 and QNXZ2020003) and “Hundred Talents Program” of the Hospital of Chengdu University of Traditional Chinese Medicine (grant nos. 20-L01, 20-Q03, and 21-Y17), and by Research Project of Sichuan Provincial Administration of TCM (grant no. 2021MS104).

### **Ethics approval and consent to participate**

There are no direct clinical activities associated with the study and only uses primary human samples for analysis, and this study was approved by the Ethics Committee of the Hospital of Chengdu University of Traditional Chinese Medicine (approval no. 2018KL-023), and informed consent was obtained from all participants.

### **Consent for publication**

Not applicable

### **Availability of data and materials**

In this study, the public data used and analyzed are summarized in “Methods”.

All the source code used in this article to analyze data and generate the results and figures can be obtained on GitHub by visit <https://github.com/lqscptp/ceRNA.git>.

### **Authors' contributions**

JT, XM, YC and JZ conceived and directed the study. NC, QL, ZZ analyzed the data, interpreted the results, and wrote the manuscript. JW, MZ and TX processed the data and conducted bioinformatics analysis. QZ, GZ, YG, YY, YQ and ZZ collected all patient samples and helped with the whole-transcriptome sequencing and RT-PCR experiments. All authors read and approved the manuscript.

### **Competing interests**

The authors declare that they have no competing interests.

### **Authors' information**

Nianzhi Chen, Qingsong Liu, Xiao Ma and Ziyi Zhao contributed equally to this work.

## **Affiliations:**

**Department of Oncology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China**

Nianzhi Chen & Yu Chen

**TCM Regulating Metabolic Diseases Key Laboratory of Sichuan Province, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China**

Ziyi Zhao, Maoyuan Zhao, Qiao Zheng, Gang Zhang, Yun Yuan, Zhongzhen Zeng & Jianyuan Tang

**State Key Laboratory of Southwestern Chinese Medicine Resources, School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China**

Xiao Ma

**Geriatric Department, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China**

Jinhao Zeng

**Department of Gastroenterology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China**

Jundong Wang & Ting Xia

**Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China**

Qingsong Liu, Yu Guo & Yulin Qi

**School of Basic Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu, China**

Shangbin Lv

## **Corresponding authors**

Correspondence to Jinhao Zeng or Jianyuan Tang.

## **References**

1. Freddie, Bray, Jacques, Ferlay, Isabelle, Soerjomataram, Rebecca, Siegel, Lindsey: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 2018.
2. Etemadi A, Safiri S, Sepanlou S, Ikuta KS, Malekzadeh R: The global, regional, and national burden of stomach cancer in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease study 2017. *The Lancet Gastroenterology & Hepatology* 2020, 5:42-54.

3. Correa PA: A human model of gastric carcinogenesis. *Cancer Research* 1988, 48:3554-3560.
4. Correa P: Human Gastric Carcinogenesis: A Multistep and Multifactorial Process—First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Research* 1992, 52:6735-6740.
5. Genta RM, Rugge M: Gastric precancerous lesions: heading for an international consensus. *Gut* 1999, 45 Suppl 1:I5-8.
6. Chang WJ DY, Zhao X, Ma LY, Cao GW: Inflammation-related factors predicting prognosis of gastric cancer. *World Journal of Gastroenterology* 2014:4586-4596.
7. M. E: Non-coding RNAs in human diseases. *Nature Reviews Genetics* 2011, 12:861-874.
8. Pei-Fei L, Sheng-Can C, Tian X, Xiao-Ming J, Yong-Fu S, Bing-Xiu X, Jun-Ming G: Non-coding RNAs and gastric cancer. *World journal of gastroenterology* 2014, 20.
9. Lu TX, Rothenberg ME: MicroRNA. *Journal of Allergy & Clinical Immunology* 2017:S0091674917315932.
10. St A, Tjz A, Aga B: Non-coding RNAs and potential therapeutic targeting in cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 2020, 1875.
11. Jiang C, Chen X, Alattar M, Wei J, Liu H: MicroRNAs in tumorigenesis, metastasis, diagnosis and prognosis of gastric cancer. *Cancer Gene Therapy* 2015, 22:291-301.
12. Wei L, Sun J, Zhang N, Zheng Y, Yang M: Noncoding RNAs in gastric cancer: Implications for drug resistance. *Molecular Cancer* 2020, 19.
13. Mi-die, Xu, Yiqin, Wang, Weiwei, Weng, Ping, Wei, Peng, Qi: A Positive Feedback Loop of lncRNA-PVT1 and FOXM1 Facilitates Gastric Cancer Growth and Invasion. *Clinical Cancer Research* 2016, 23:2071.
14. Zhang G, Li S, Lu J, Ge Y, Wang Q, Ma G, Zhao Q, Wu D, Gong W, Du M: LncRNA MT1JP functions as a ceRNA in regulating FBXW7 through competitively binding to miR-92a-3p in gastric cancer. *Molecular Cancer* 2018, 17:87.
15. Ea A, Cs A, Lm A, Li B, Do B, Chlac D: CircRNAs and cancer: Biomarkers and master regulators. *Seminars in Cancer Biology* 2019, 58:90-99.
16. Li R, Jiang J, Shi H, Qian H, Xu W: CircRNA: a rising star in gastric cancer. *Cellular and Molecular Life Sciences* 2020, 77.
17. Zhang X, Wang S, Wang H, Cao J, Huang X, Chen Z, Xu P, Sun G, Xu J, Lv J: Circular RNA circNRIP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway. *Molecular Cancer* 2019, 18.
18. Yang XZ, Cheng T, He QJ, Lei ZY, Chi J, Zhen T, Liao QX, Hong Z, Zeng LS, Cui SZ: LINC01133 as ceRNA inhibits gastric cancer progression by sponging miR-106a-3p to regulate APC expression and the Wnt/ $\beta$ -catenin pathway. *Molecular Cancer* 2018, 17:126-.
19. Huang X, Li Z, Zhang Q, Wang W, Li B, Wang L, Xu Z, Zeng A, Zhang X, Zhang X: Circular RNA AKT3 upregulates PIK3R1 to enhance cisplatin resistance in gastric cancer via miR-198 suppression.

*Molecular Cancer* 2019, 18.

20. Tay Y, Rinn J, Pandolfi PP: The multilayered complexity of ceRNA crosstalk and competition. *Nature* 2014, 505:344-352.
21. Karreth FA, Pandolfi PP: ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discovery* 2013, 3:1113-1121.
22. Guo LL: Competing endogenous RNA networks and gastric cancer. *World Journal of Gastroenterology* 2015, 21:11680.
23. Stark R, Grzelak M, Hadfield J: RNA sequencing: the teenage years. *Nature Reviews Genetics* 2019.
24. Feng Y, Powell DR, Curtis DJ, Wong NC: From reads to insight: A hitchhiker's guide to ATAC-seq data analysis. *Genome Biology* 2020, 21.
25. Christoph R: Molecular classification of gastric cancer. *Expert review of molecular diagnostics* 2017, 17.
26. Maleki SS, R?cken C: Chromosomal Instability in Gastric Cancer Biology. *Neoplasia* 2017, 19:412-420.
27. Hartgrink HH, Jansen E, Grieken N, Velde C: Gastric cancer. *Lancet* 2009, 374:477-490.
28. Harvey ZH, Chen Y, Jarosz DF: Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Molecular Cell* 2017:S1097276517308079.
29. Zhijian, Li, Marcel, Schulz, Thomas, Look, Matthias, Begemann, Martin: Identification of transcription factor binding sites using ATAC-seq. *Genome Biology* 2019.
30. Claesson-Welsh L WM: VEGFA and tumour angiogenesis. *Journal of Internal Medicine* 2013, 273.
31. Park DJ, Thomas NJ, Yoon C, Yoon SS: Vascular Endothelial Growth Factor A Inhibition in Gastric Cancer. *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 2015, 18:33-42.
32. Roviello G PR, Marano L, Polom K, Marrelli D, Perrella A, Roviello F. : Angiogenesis inhibitors in gastric and gastroesophageal junction cancer. *Gastric Cancer* 2016, 19:31-41.
33. Liu D, He Q, Liu C: Correlations among Helicobacter pylori infection and the expression of cyclooxygenase-2 and vascular endothelial growth factor in gastric mucosa with intestinal metaplasia or dysplasia. *Journal of Gastroenterology & Hepatology* 2010, 25:795-799.
34. W-X Z, Z-F L, X-L L, Z L: Correlations of serum homocysteine, VEGF and gastrin 17 with gastric cancer and precancerous lesions. *European review for medical and pharmacological sciences* 2019, 23.
35. Liu Y, Xue M, Du S, Feng W, Zhang K, Zhang L, Liu H, Jia G, Wu L, Hu X: Competitive endogenous RNA is an intrinsic component of EMT regulatory circuits and modulates EMT. *Nature Communications* 2019, 10.
36. Cai X, Liu C, Zhang TN, Zhu YW, Dong X, Xue P: Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration and invasion. *Journal of Cellular Biochemistry* 2017.

37. Hiroshima Y, Kasajima R, Kimura Y, Komura D, Miyagi Y: Novel targets identified by integrated cancer-stromal interactome analysis of pancreatic adenocarcinoma. *Cancer Letters* 2019, 469.
38. Li L, Zhu Z, Zhao Y, Zhang Q, Fei S: FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. *Scientific Reports* 2019, 9.
39. Zhang HS, Zhiguo Li, Yan Fan, Dong Jiang, Hao MicroRNA-200c binding to FN1 suppresses the proliferation, migration and invasion of gastric cancer cells. *Biomedicine & Pharmacotherapy* 2017, 88:285-292.
40. Wu C, Zeng MH, Liao G, Qian K, Li H: Neuropilin-1 Interacts with Fibronectin-1 to Promote Epithelial-Mesenchymal Transition Progress in Gastric Cancer. *OncoTargets and Therapy* 2020, Volume 13:10677-10687.
41. Xu TP, Ma P, Wang WY, Shuai Y, Wang YF, Yu T, Xia R, Shu YQ: KLF5 and MYC modulated LINC00346 contributes to gastric cancer progression through acting as a competing endogenous RNA and indicates poor outcome. *Cell Death & Differentiation* 2019.
42. Du W LD, Guo X, Li P, Li X, Tong S, Tong J, Kuang L, Liang D.: Circ-PRMT5 promotes gastric cancer progression by sponging miR-145 and miR-1304 to upregulate MYC. *Artificial Cells* 2019, 47:4120-4130.
43. Yang Z, Jiang X, Li D, Jiang X: HBXIP promotes gastric cancer via METTL3-mediated MYC mRNA m6A modification. *Aging* 2020, 12.
44. Silva TCR, Leal MF, Calcagno DQ, Souza CRTd, Khayat AS, Santos NPCd, Montenegro RC, Rabenhorst SHB, Nascimento MQ, Assumpção PP, et al: hTERT, MYC and TP53 deregulation in gastric preneoplastic lesions. *BioMed Central* 2012, 12.
45. Danli, Cai, Jianshun, Jing, Qiu, Beihui, Zhiyun, Chen, Maoxiang, Yan: Dynamic changes of Sonic Hedgehog signaling pathway in gastric mucosa of rats with MNNG-induced gastric precancerous lesions. *Journal of cellular physiology* 2018.
46. Zhang X, Ma H, Zou Q, Wu J: Analysis of Cyclin-Dependent Kinase 1 as an Independent Prognostic Factor for Gastric Cancer Based on Statistical Methods. *Frontiers in Cell and Developmental Biology* 2020, 8.
47. Barbacid MMM: Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer* 2009.
48. Li HL, Wen Liu, Shanshan Zong, Shaoqi Wang, Weibing Ren, Jianlin Li, Qi Hou, Fenggang Shi, Qi DNMT1, DNMT3A and DNMT3B Polymorphisms Associated With Gastric Cancer Risk: A Systematic Review and Meta-analysis. *EBioMedicine* 2016, 13.
49. Li Y, Yan J, Wang Y, Wang C, Zhang C, Li G: LINC00240 promotes gastric cancer cell proliferation, migration and EMT via the miR-124-3p / DNMT3B axis. *Cell Biochemistry and Function* 2020, 38.
50. Du L, Wang X, Yin Y, Zhang Y, Qi J: Identification of a potentially functional circRNA-miRNA-mRNA ceRNA regulatory network in bladder cancer by analysis of microarray data. *Translational Andrology and Urology* 2021, 10:24-36.

51. Li X, Tian R, Gao H, Yang Y, Williams B, Gantier MP, Mcmillan N, Xu D, Hu Y, Gao Y: Identification of a histone family gene signature for predicting the prognosis of cervical cancer patients. *Scientific Reports* 2017, 7:16495.
52. Moreira AM, Pereira J, Melo S, Fernandes MS, Carneiro P, Seruca R, Figueiredo J: The Extracellular Matrix: An Accomplice in Gastric Cancer Development and Progression. *Cells* 2020, 9.
53. Liu B, Zhang Y, Fan Y, Wang S, Li Z, Deng M, Li C, Wang J, Ma R, Wang X, et al: Leucine-rich repeat neuronal protein-1 suppresses apoptosis of gastric cancer cells through regulation of Fas/FasL. *Cancer Science* 2019, 110.
54. Li N, Gao W, Zhang Y-F, Ho M: Glypicans as Cancer Therapeutic Targets. *Trends in Cancer* 2018.
55. J I, E DS, K OdB: GSDME: A Potential Ally in Cancer Detection and Treatment. *Trends in cancer* 2021, 7.
56. Jin C, Rajabi H, Kufe D: miR-1226 targets expression of the mucin 1 oncoprotein and induces celldeath. *International Journal of Oncology* 2010, 37.
57. Xiao-Yu L, Li-Ye Z, Hao L, Qi Z, Liang Z, Gu-Yue L, Chu F, Jun-Yong Z, Yuan-Yuan Z, Xue L: The long noncoding RNA MIR210HG promotes tumor metastasis by acting as a ceRNA of miR-1226-3p to regulate mucin-1c expression in invasive breast cancer. *Aging* 2019, 11.
58. Zahra M, Zahra G, M. SB, Parisa H: Cell Specific Tumor Suppressor Effect of Hsa-miR-1226-3p through Downregulation of HER2, PIK3R2, and AKT1 genes. *The International Journal of Biochemistry & Cell Biology* 2021, 134.
59. Yi W, Liu J, Qu S, Fan H, Lv Z: An 8 miRNA-Based Risk Score System for Predicting the Prognosis of Patients With Papillary Thyroid Cancer. *Technology in Cancer Research & Treatment* 2020, 19:153303382096559.
60. E FM, Bin X, C HM, K MS: G-protein-coupled receptor GPR161 is overexpressed in breast cancer and is a promoter of cell proliferation and invasion. *Proceedings of the National Academy of Sciences of the United States of America* 2014, 111.
61. Shimada IS, Hwang S-H, Somatilaka BN, Wang X, Skowron P, Kim J, Kim M, Shelton JM, Rajaram V, Xuan Z, et al: Basal Suppression of the Sonic Hedgehog Pathway by the G-Protein-Coupled Receptor Gpr161 Restricts Medulloblastoma Pathogenesis. *Cell Reports* 2018, 22.
62. Song M, Kim H, Kim WK, Hong SP, Lee C, Kim H: High Expression of AT-Rich Interactive Domain 3A (ARID3A) is Associated with Good Prognosis in Colorectal Carcinoma. *Annals of Surgical Oncology* 2014, 21.
63. Dausinas P, Pulakanti K, Rao S, Cole JM, Dahl R, Dahl KDC: ARID3A and ARID3B induce stem promoting pathways in ovarian cancer cells. *Gene* 2020, 738.
64. Soudeh G-F, Mohammadhosein E, Mohammad T: H19 lncRNA: Roles in tumorigenesis. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 2020, 123.
65. Jing R, Liang D, Dongya Z, Guoping S, Qianyun X, Sunan S, Yaping W, Tingting W, Yayi H: Carcinoma-associated fibroblasts promote the stemness and chemoresistance of colorectal cancer by transferring exosomal lncRNA H19. *Theranostics* 2018, 8.

66. Gene Ontology Consortium: going forward. *Nucleic acids research* 2015, 43.
67. Minoru K, Michihiro A, Susumu G, Masahiro H, Mika H, Masumi I, Toshiaki K, Shuichi K, Shujiro O, Toshiaki T, Yoshihiro Y: KEGG for linking genomes to life and the environment. *Narnia* 2008, 36.
68. M K, S G: KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 2000, 28.
69. Vella D, Marini S, Vitali F, Silvestre DD, Mauri G, Bellazzi R: MTGO: PPI Network Analysis Via Topological and Functional Module Identification. *Scientific Reports* 2018, 8.
70. W TD, E DM: Endogenous microRNA sponges: evidence and controversy. *Nature reviews Genetics* 2016, 17.
71. Hutter C, Zenklusen JC: The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* 2018, 173.
72. Tanujit D, Anish M, Sounak C: A Practical Overview and Reporting Strategies for Statistical Analysis of Survival Studies. *Chest* 2020, 158.
73. Obuchowski NA, Bullen JA: Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Physics in Medicine & Biology* 2018, 63.

## Tables

**Table 1 The Degree of the Representative RNAs in the Five Modules**

Cluster	Score	ID	Degree
Module 1	82.547	PBK	95
		KIF23	95
		KIF2C	95
		NCAPG	95
		ASPM	95
Module 2	28.198	FN1	48
		TIMP1	40
		FBN1	37
		CXCL8	37
		FSTL1	36
Module 3	9.658	ICAM1	23
		LY6E	20
		GBP1	19
		OASL	18
		IRF4	17
Module 4	9.617	THBS1	30
		THBS2	15
		SPARC	14
		F2R	14
		CCKBR	13
Module 5	6.875	VEGFA	21
		CD44	17
		FOXP3	11
		NT5E	10
		MAP3K11	10

**Table 2 Representative RNAs in the three Big ceRNA Networks**

Network	ID	Degree
lncRNA-miRNA-mRNA	hsa-miR-1226-5p	234
	hsa-miR-6720-5p	209
	novel-miR-1309	157
	hsa-miR-615-5p	145
	hsa-miR-3127-3p	99
	hsa-miR-3180-3p	93
	hsa-miR-10396b-3p	82
	novel-miR-495	82
	hsa-miR-3180-5p	80
	MSTRG.29081.1	77
circRNA-miRNA-mRNA	4:39508566 39552158	24
	hsa-miR-6720-5p	9
	hsa-miR-1226-5p	9
	ARID3A	6
	GPR161	6
	14:105587554 105707562	6
	hsa-miR-3180-3p	5
	hsa-miR-615-5p	5
	hsa-miR-3180-5p	5
	CACNA1E	4
ceRNA	hsa-miR-1226-5p	124
	hsa-miR-6720-5p	99
	hsa-miR-615-5p	55
	hsa-miR-3127-3p	43
	hsa-miR-1304-3p	42
	hsa-miR-3180-5p	41
	hsa-miR-3180-3p	38
	GPR161	29

Figures

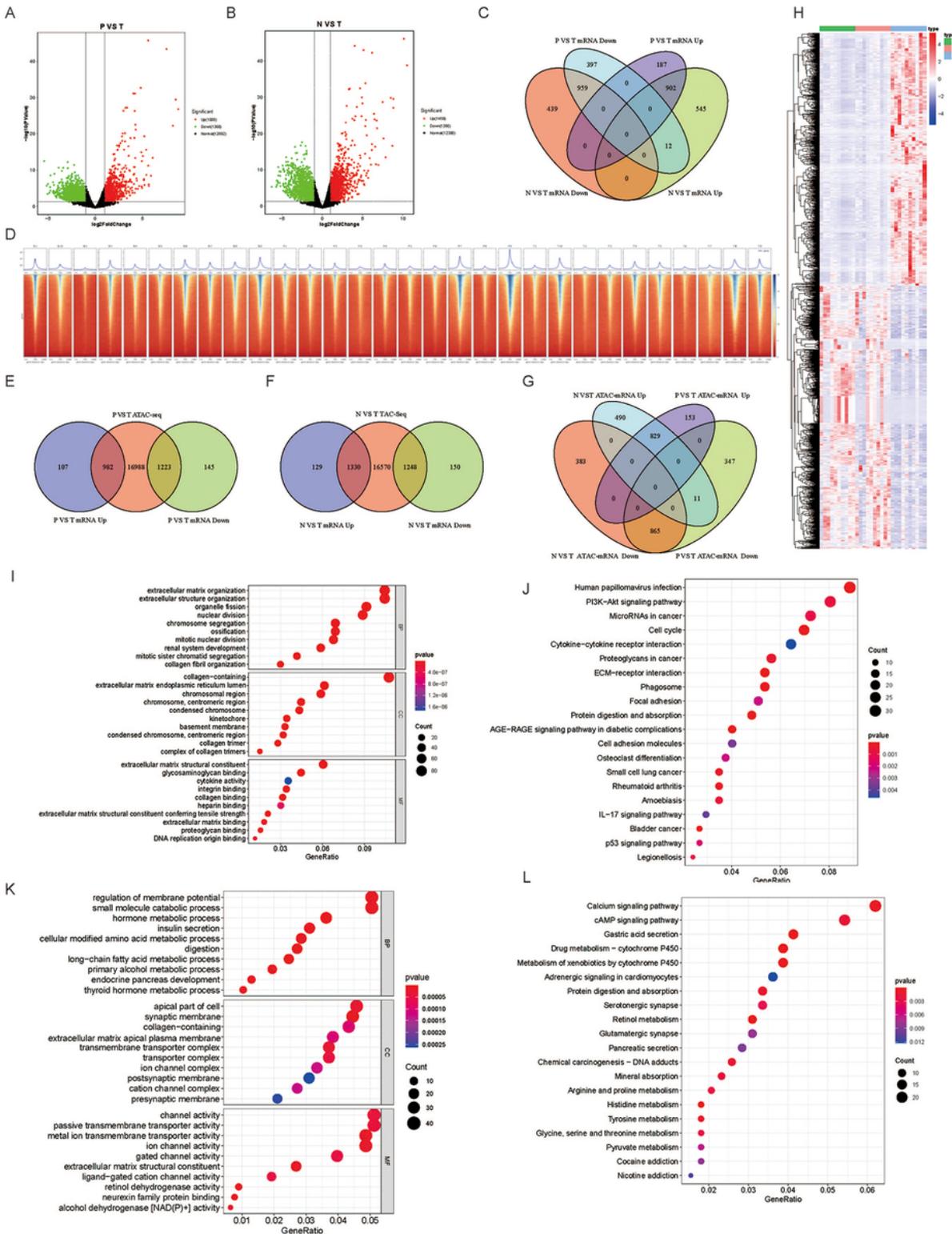
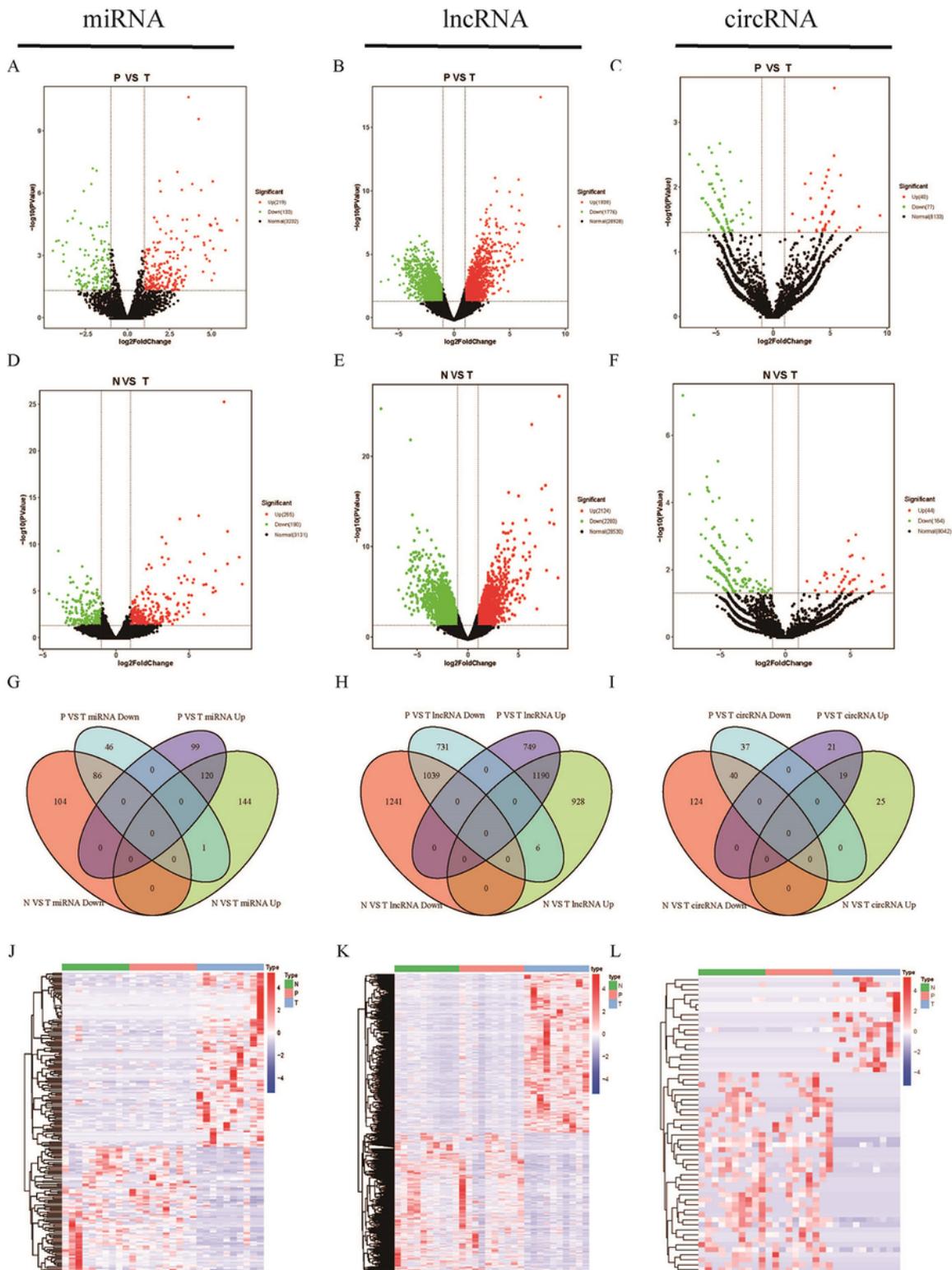


Figure 1

Differentially Expressed Analysis and Functional Enrichment of Candidate DEGs In the volcano plots, the significant changes of dif-mRNAs after comparing the two groups were visualized. Red and green dots indicate upregulated and downregulated genes, respectively(A-B). (A) Volcano plots of differentially expressed mRNAs in GPLs (P group) compared with EGC (T group). (B) and in normal tissues (N group) compared with EGC (T group). (C) Venn diagram of the dif-mRNAs in two groups of PvsT and NvsT analysis. (D) Heatmaps of ATAC-seq of the density distribution in TSS of each gene to testify the chromatin is open well. (E) The intersection of dif-mRNAs and the alignment genes by ATAC-seq in P group compared with T group. (F) and in N group compared with T group. (G) Venn diagram of DEGs in pairwise groups. (H) Heatmaps of overlapped DEGs in Venn diagram G, colours indicate expression values, with brighter red indicating higher values and brighter purple lower values. (I-J) Top 10 pathways of GO-BP, GO-CC, GO-MF terms and top 20 KEGG pathways enriched by upregulated DEGs, respectively. (K-L) Downregulated DEGs were enriched by the same way as the former.



**Figure 2**

Identification and Analysis of Differentially Expressed ncRNAs (A)Volcano plots of differentially expressed miRNAs in PvsT. (B) and in NvsT. (C) Volcano plots of differentially expressed lncRNAs in PvsT. (D) and in NvsT. (E) Volcano plots of differentially expressed circRNAs in PvsT. (F) and in NvsT. (G) Venn diagram of the overlapped dif-miRNAs in pairwise groups. (H) and Venn diagram of dif-lncRNAs. (I)

and Venn diagram of the dif-circRNAs. (J) Heatmaps of dif-miRNAs. (K) Heatmaps of dif-lncRNAs. (L) Heatmaps of dif-circRNAs.



**Figure 3**

PPI Networks of Candidate DEGs In PPI networks of DEGs, red and green indicate upregulated and downregulated genes, respectively. (A) PPI network of candidate DEGs. (B) A bar chart with top 30 hub-

genes according to the number of interactions. (C-G) Five significant clustered modules extracted from PPI network. (H-I) GO and KEGG enrichment analysis of the genes in the five modules.

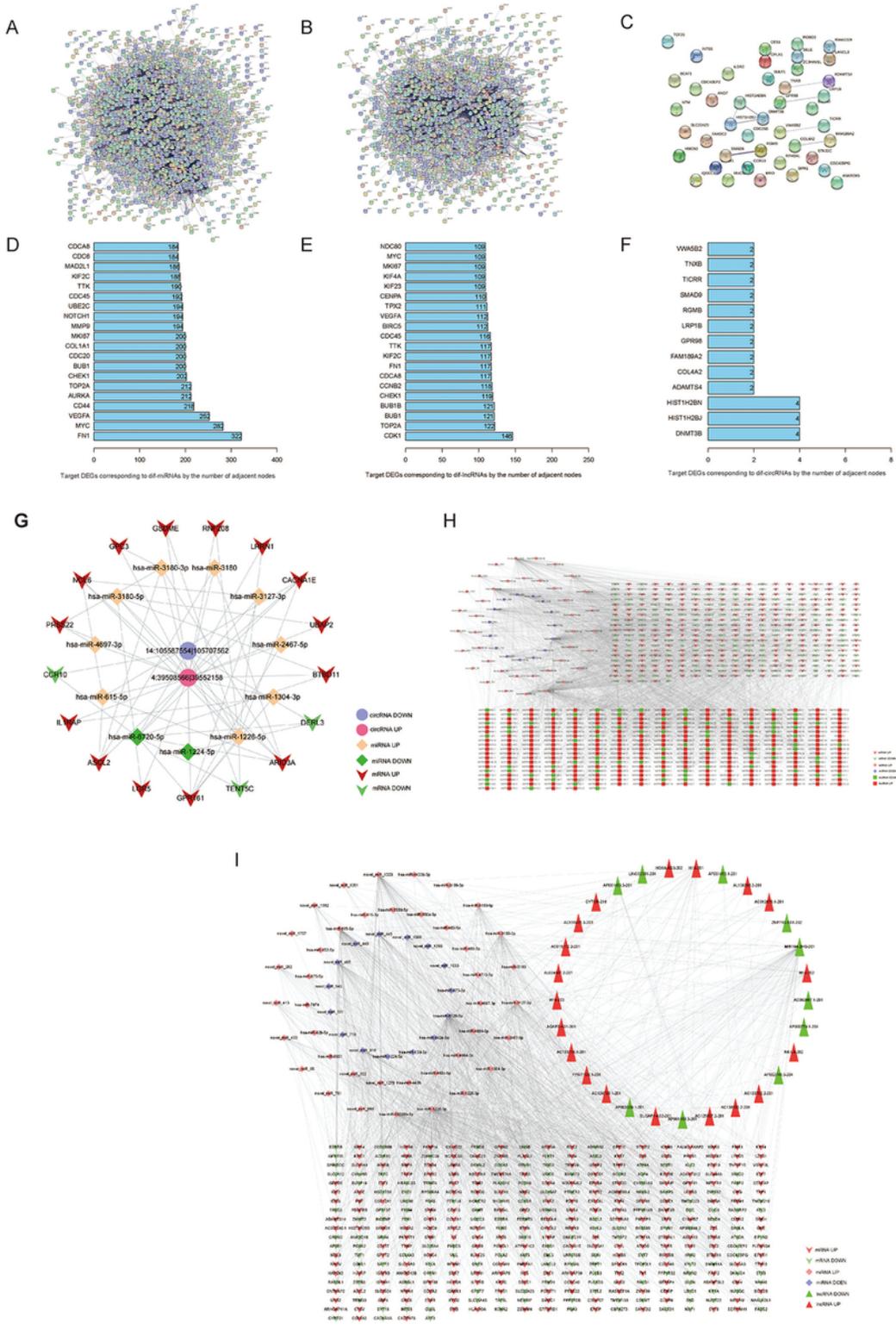
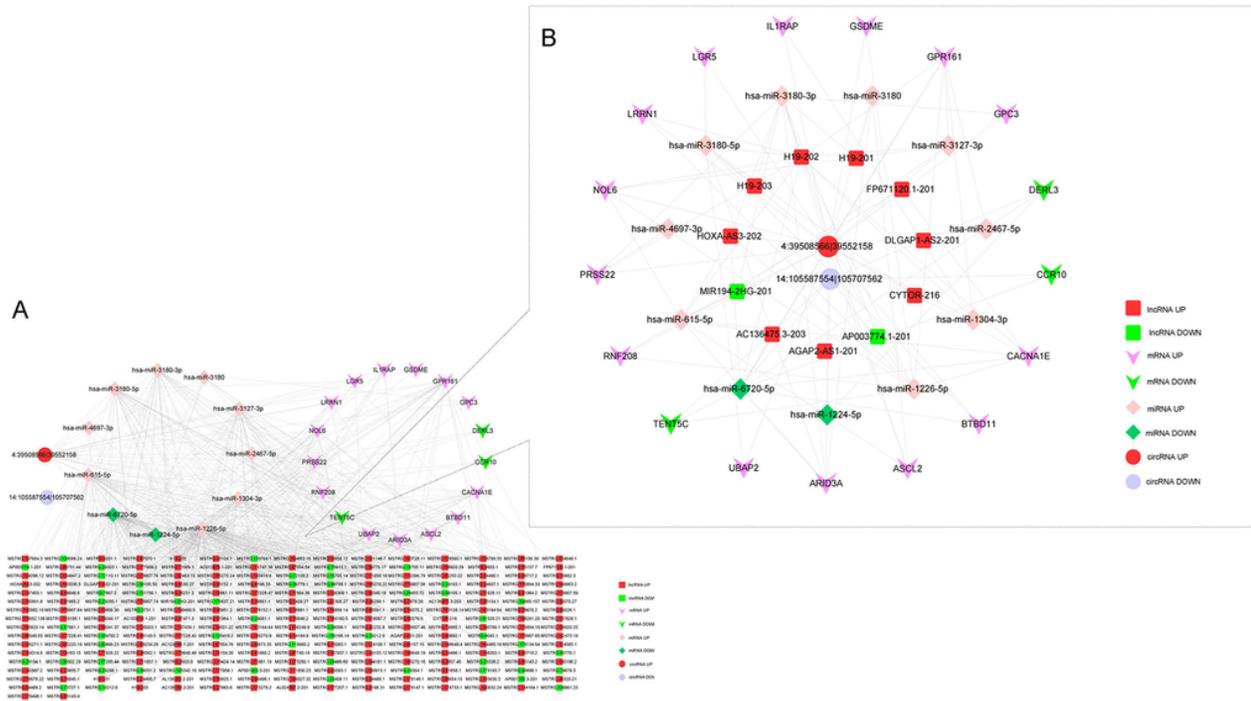


Figure 4

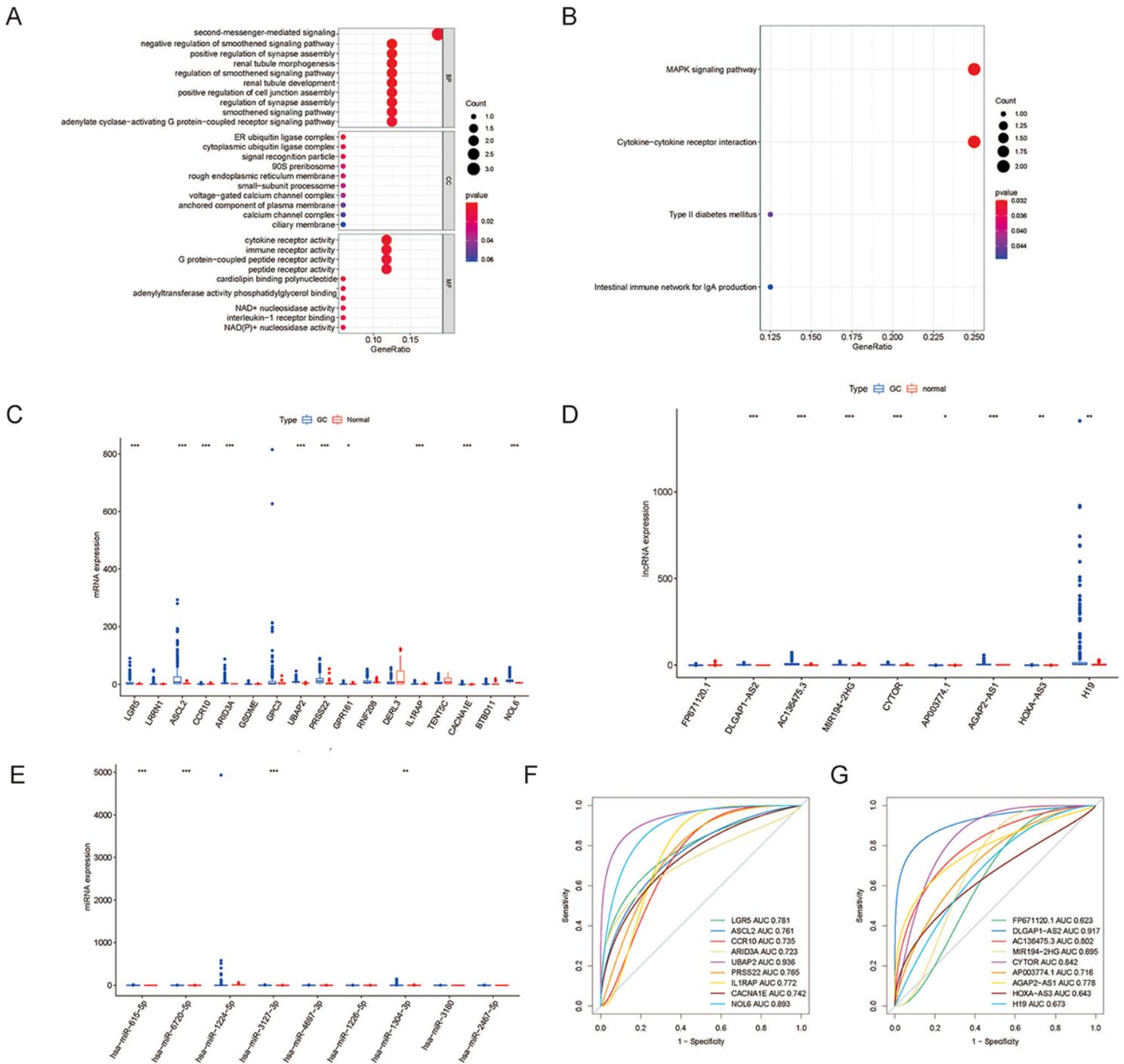
PPI networks of dif-ncRNAs and lncRNA-Related, circRNA-Related ceRNA Network (A-C) PPI networks of the target genes in DEGs which corresponding to dif-miRNAs, dif-lncRNAs and dif-circRNAs. (D-F) A bar chart of the top 20 target genes according to the number of adjacent nodes. (G) The differentially

expressed circRNA-miRNA-mRNA network. (H) The differentially expressed lncRNA-miRNA-mRNA network. (I) and then based on the lncRNA-miRNA-mRNA network formed a core sub-network that filters out unknown lncRNAs.



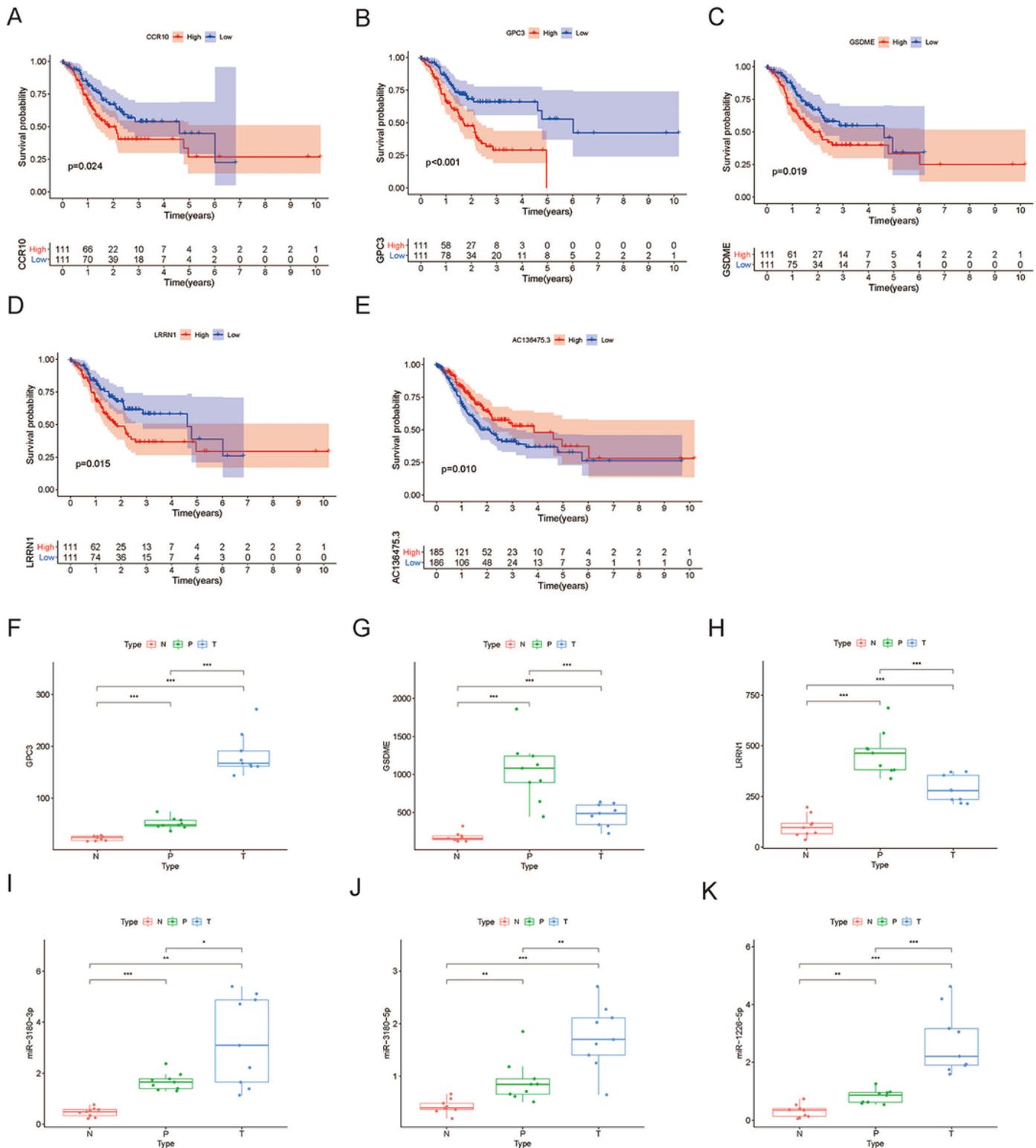
**Figure 5**

The Competing Endogenous RNA (ceRNA) Network (A) The competing endogenous RNA (ceRNA) network with differentially expressed circRNA-lncRNA-miRNA-mRNA. (B) A core ceRNA network that filters out unknown lncRNAs which cannot be verified.



**Figure 6**

Functional Enrichment of the genes in ceRNA network, Validation of The Cancer Genome Atlas (TCGA) Database and ROC Curve Analysis (A-B) GO and KEGG enrichment analysis of the genes in the ceRNA network. (C-E) mRNA, miRNA and lncRNA expression levels in gastric cancer and normal samples according to TCGA database, \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ . (C) mRNA. (D) miRNA. (E) lncRNA. (F) ROC curve analysis with nine mRNAs which AUC  $\geq 0.7$  represent diagnostic values for discriminating cancer from normal tissues. (G) ROC curve analysis with nine lncRNAs.



**Figure 7**

Survivorship Curve Analysis and RT-PCR Validation (A-E) Four mRNAs and one lncRNA with significant differences in survivorship curve analysis, including (A) CCR10, (B) GPC3, (C) GSDME, (D) LRRN1, (E) AC136475.3. (F-K) RT-PCR validation of (F) GPC3, (G) GSDME, (H) LRRN1, (I) hsa-miR-3180-3p, (J) hsa-miR-3180-5p, (K) hsa-miR-1226-5p between N (normal gastric mucous), P (precancerous lesions tissues), T (EGC tissues) groups, \*\*\*P<0.001, \*\*P<0.01, \*P<0.05.