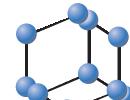


RESEARCH ARTICLE

BENTHAM
SCIENCE

Investigating Full-Length circRNA Transcripts to Reveal circRNA-Mediated Regulation of Competing Endogenous RNAs in Gastric Cancer



Jingjing Liu^{1,#}, Quan Yuan^{2,#}, Runqiu Cai^{3,#}, Jian Zhao¹, Juan Chen⁴, Meng Zhang¹, Yulan Wang¹, Minhui Zhuang¹, Tianyi Xu^{5,*}, Xiaofeng Song^{1,*} and Jing Wu^{2,*}

¹Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 211106, China; ²School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, Jiangsu, 211166, China; ³Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, Jiangsu, 210029, China; ⁴Children's Hospital of Nanjing Medical University, Nanjing, Jiangsu, 210008, China; ⁵National Genomics Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing, 100101, China

Abstract: **Background:** Circular RNAs (circRNAs) play important regulatory roles in the progression of gastric cancer (GC), but the exact mechanisms governing their regulation remain incompletely understood. Prior studies typically used back-spliced junctions (BSJs) to represent a range of circRNA isoforms, overlooking the prevalence of alternative splicing (AS) events within circRNAs, which could lead to unreliable or even incorrect conclusions in subsequent analyses, hindering our comprehension of the specific functions of circRNAs in GC.

Objective: This study aimed to explore the potential functional roles of the dysregulated circRNA transcripts in GC and provide new biomarkers and effective novel therapeutic strategies for GC treatment.

Methods: RNA-seq data with rRNA depletion and RNase R treatment was employed to characterize the expression profiles of circRNAs in GC, and RNA-seq data only with rRNA depletion was employed to identify differentially expressed mRNAs in GC. Based on the full-sequence information and accurate isoform-level quantification of circRNA transcripts calculated by the CircAST tool, we performed a series of bioinformatic analyses. A circRNA-miRNA-hub gene regulatory network was constructed to reveal the circRNA-mediated regulation of competing endogenous RNAs in GC, and then the protein-protein interaction (PPI) network was built to identify hub genes.

Results: A total of 18,398 circular transcripts were successfully reconstructed in the samples. Herein, 351 upregulated and 177 downregulated circRNA transcripts were identified. Functional enrichment analysis revealed that their parental genes were strongly associated with GC. After several screening steps, 19 dysregulated circRNA transcripts, 40 related miRNAs, and 65 target genes (mRNAs) were selected to construct the ceRNA network. Through PPI analysis, five hub genes (COL5A2, PDGFRB, SPARC, COL1A2, and COL4A1) were excavated. All these hub genes may play vital roles in gastric cancer cell proliferation and invasion.

Conclusion: Our study revealed a comprehensive profile of full-length circRNA transcripts in GC, which could provide potential prognostic biomarkers and targets for GC treatment. The results would be helpful for further studies on the biological roles of circRNAs in GC and offer new mechanistic insights into the pathogenesis of GC.

ARTICLE HISTORY

Received: August 12, 2024
Revised: August 29, 2024
Accepted: September 06, 2024

DOI:
10.2174/0115748936346422240930081839



Keywords: Biomarker, ceRNA network, circRNA, full-length transcript, gastric cancer, endogenous RNAs.

1. INTRODUCTION

Gastric cancer (GC) is a prevalent malignancy of the digestive system originating from the gastric mucosal

epithelium. Globally, GC ranks fifth in terms of morbidity and fourth in mortality among all cancers, leading to approximately 800,000 deaths annually [1]. Despite a decrease in incidence and mortality over the past five decades, GC remains a significant contributor to cancer-related deaths worldwide, with a 5-year survival rate of less than 30% after diagnosis [2]. The pathogenesis of human GC is a multifaceted process, with *Helicobacter pylori* (*H. pylori*) infection considered a major predisposing factor, although the exact mechanism remains incompletely understood [3-5]. GC can be categorized into three subtypes: adenocarcinoma, signet ring-cell carcinoma, and undifferentiated carcinoma [6]. Among these, stomach adenocarcinoma (STAD) is the most

*Address correspondence to these authors at the School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, Jiangsu 211166, China; E-mail: wujing@njmu.edu.cn (J. Wu); Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China; E-mail: xfong@nuaa.edu.cn (X. Song); National Genomics Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China; E-mail: xuty@big.ac.cn (T. Xu)

[#]These authors contribution equal to this work.

prevalent pathological type, accounting for approximately 95% of all diagnosed cases [7, 8]. Currently, the primary treatment for GC is surgical resection [9, 10]. However, due to the lack of specific clinical manifestations at an early stage, most patients are diagnosed at advanced stages characterized by extensive invasion and lymphatic metastasis, leading to a poor prognosis [11]. Furthermore, conventional cancer therapies, including chemotherapy and radiotherapy, entail inevitable toxic and side effects for GC patients, and the efficacy of chemotherapy diminishes over time due to drug resistance [12, 13]. In recent years, there has been increasing exploration of the molecular mechanisms underlying GC, with molecular therapeutics gradually being integrated into GC treatment [14-16]. Therefore, in-depth investigations into the molecular mechanisms involved in GC progression and the identification of novel biomarkers or therapeutic targets are urgently needed to develop more effective therapeutic strategies.

Circular RNAs (circRNAs), a newly discovered class of endogenous ncRNAs, have garnered considerable attention in cancer research due to their significant biological functions [17]. Defined by covalently closed continuous loop structures lacking 5' and 3' polarities and polyadenylate tails, circRNAs were initially considered a by-product of aberrant RNA splicing or splicing errors [18]. Advancements in high-throughput sequencing technologies have led to the identification of tens of thousands of circRNAs in various species and cell lines [19]. Subsequent research has revealed that circRNAs exhibit cell- or tissue-specific patterns and are highly conserved across species [20]. It has been demonstrated that circRNAs play crucial biological roles by acting as a microRNA (miRNA) sponge to modulate miRNA-mediated gene expression [19, 21], regulating gene transcription or alternative splicing [22, 23], or aiding in protein translation [24-26]. In comparison to their linear counterparts, circRNAs exhibit more stable expression and are predominantly localized in the cytoplasm, suggesting their potential to function as competitive endogenous RNAs (ceRNAs) [27]. Growing evidence has shown that circRNAs play significant roles in various human diseases, including cancer, and their aberrant expression may contribute to tumorigenesis, cancer progression, and metastasis, indicating their potential as promising biomarkers or therapeutic targets for cancer therapy [28, 29].

Numerous studies have delved into the molecular functions of circRNAs in the initiation and progression of GC. For example, circ-DONSON was found to enhance the growth and invasive potential of GC cells by binding to the NURF complex to activate SOX4 transcription [30]. Additionally, circNHS1 was identified as an oncogenic circRNA that exacerbates the advancement and metastasis of GC through the ceRNA network circNHS1/miR-1306-3p/SIX1/Vimentin [31]. On the other hand, circLARP4 was observed to suppress the aggressiveness of GC by sequestering miR-424-5p and increasing LATS1 expression [32]. However, the internal structure of circRNA and the expression levels of full-length circRNA transcripts in GC are not

yet elucidated, hindering our comprehension of the specific functions of circRNAs in GC.

Circular RNA and mRNA share high sequence similarity, which presents a major challenge to accurately reconstruct circRNA full-length sequences and quantify circRNA at the isoform level. Prior studies have predominantly relied on a straightforward approach, sequentially concatenating known mRNA exons to generate complete circRNA sequences and quantifying circRNA expression based on the number of back-spliced junction (BSJ) reads [33]. However, these methods overlook the prevalence of alternative splicing events within circRNAs, potentially resulting in the generation of multiple circRNA transcripts with distinct sequences. Incorrect or incomplete circRNA sequences, along with imprecise expression estimates, may lead to unreliable or even erroneous conclusions in subsequent analyses. In recent years, several novel computational methods have been developed for assembling and quantifying circular RNA transcripts. For instance, CircAST employs multiple splice graphs and the expectation-maximization (EM) algorithm to reconstruct and quantify full-length circRNAs [34]. CIRI-full combines reverse overlap (RO) and back-splice junction (BSJ) features to reconstruct circRNA sequences and evaluate their abundance [35]. Some other tools like circRNAfull and JCCirc have also made efforts to assemble full-length circRNA sequences. All these tools have deepened circRNA studies to the isoform-level resolution, enhancing our understanding of their biogenesis and functions [36, 37]. However, most of the existing studies on GC are still based on the BSJ level rather than the isoform level.

In the present study, we employed RNA-seq data to explore the expression profiles of circRNA transcripts in GC and matched adjacent normal tissues (Fig. 1). We utilized the established tool CircAST to reconstruct full-length circRNA transcripts in GC and quantify their expression at the isoform level. Complete information on circRNA full-length sequence aids in precisely identifying miRNA-mRNA interaction pairs, while accurate quantification of circRNA at the isoform level assists in filtering differentially expressed circRNA transcripts for subsequent functional analysis. We constructed a circRNA-miRNA-mRNA network, which was in line with the ceRNA theory, and screened hub genes based on the protein-protein interaction (PPI) network (Fig. S1). The objective of this study was to elucidate the molecular mechanisms of circRNA in the pathogenesis of GC and to identify novel biomarkers and effective therapeutic targets for the treatment of GC.

2. MATERIALS AND METHODS

2.1. Data Sources

A total of 26 high-throughput RNA-sequencing (RNA-seq) datasets of human GC were obtained from the NCBI Sequence Read Archive database (SRA, <http://www.ncbi.nlm.nih.gov/sra/>). Among these, six datasets underwent rRNA depletion and RNase R treatment, comprising three GC tissues and three paired adjacent normal tissues (SRA:

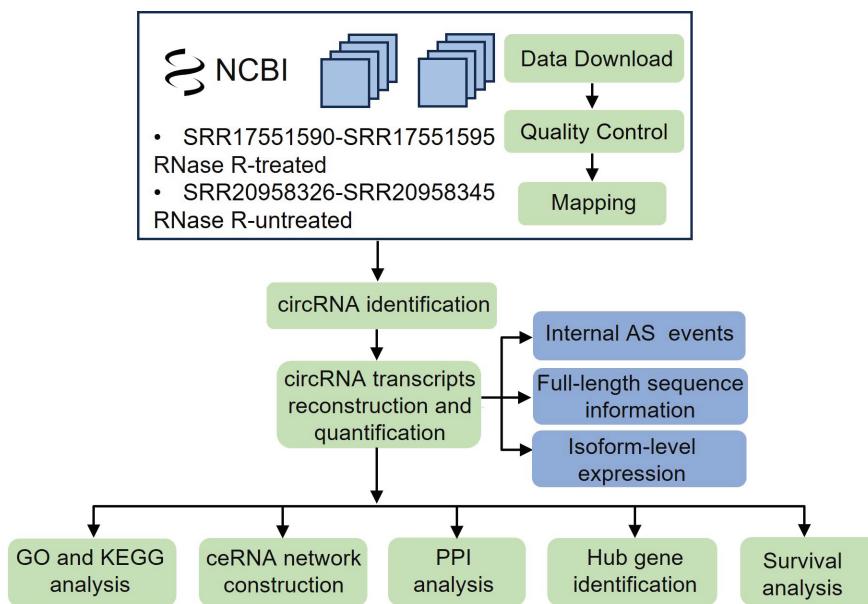


Fig. (1). The overall flow chart of this study. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

SRR17551590-SRR17551595) [38], which were utilized to identify circRNAs linked to GC progression. The remaining 20 datasets, exclusively subjected to rRNA depletion, included 10 tumor samples and 10 matched adjacent normal tissue samples from the same GC patients (SRA: SRR20958326-SRR20958345) and were employed to identify differentially expressed mRNAs between GC and normal samples.

The miRNA expression profile of GC in TCGA was acquired from the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>). The dataset comprised 491 individuals, consisting of 446 tumor samples and 45 normal samples.

2.2. Expression Profiles of Full-length circRNA Isoforms in GC

CircRNAs were initially identified using CIRI2 [39] with default parameters. Unlike previous related analytical approaches, to conduct our subsequent analysis at the isoform level, candidate circRNAs derived from exonic regions were selected as input for CircAST to perform full-length assembly and quantify alternatively spliced circRNA transcripts. The sequence details of each circRNA transcript, along with their abundance estimates measured in FPKM, were obtained from the CircAST output file. The human reference genome (GRCh37/hg19) obtained from the UCSC Genome Browser (<http://genome.ucsc.edu/>) was utilized in both CIRI2 and CircAST.

2.3. Identification of Differentially Expressed circRNA Transcripts

Unlike previous studies, we performed isoform-level differential expression analyses. After normalizing the relative abundance values, fold change (FC) values were computed as the ratio of average transcript expression in each group,

followed by performing Student's *t*-test for significance testing. The screening criteria for differentially expressed circRNA transcripts were adjusted as *p*-value < 0.05 and $|\log_2 \text{FC}| > 2$. All bioinformatics analyses were carried out using R and Python software.

2.4. GO Enrichment and KEGG Pathway Analyses

Previous research indicated that circRNAs may exert their functions through their parental genes [22]. To gain deeper insights into the potential functions of host genes corresponding to dysregulated circRNA transcripts in GC, GO enrichment and KEGG pathway analyses were conducted using the Database for Annotation, Visualization and Integrated Discovery (DAVID) online analysis tool (<http://david.abcc.ncifcrf.gov/>) [40]. The enrichment test was performed with default parameters across three categories: BP (biological process), MF (molecular function), and CC (cellular component). The associated GO and pathway terms were arranged in descending order based on the enrichment score-log (*p*-value).

2.5. Prediction of circRNA-miRNA Interactions and miRNA Target Genes

The prediction of circRNA-miRNA interactions heavily relies on their sequence composition. Given that we had reconstructed the full-length sequence of circRNA transcripts, we proceeded by obtaining the sequence data of human mature miRNAs from the miRBase database (version 22)[41] and utilized miRanda software [42] to predict miRNA targets for differentially expressed circular transcripts. Subsequently, these miRNAs were further refined using miRTarbase (<https://mirtarbase.cuhk.edu.cn/>) [43] based on miRNA-disease associations. Simultaneously, leveraging the miRNA expression profiles of GC patients from TCGA, we identified potential differentially expressed miRNAs in GC. By

cross-referencing the target miRNAs with differentially expressed miRNAs obtained from TCGA, we ultimately confirmed GC-associated target miRNAs.

The prediction of miRNA target genes was conducted using three online databases: TargetScan (www.targetscan.org/) [44], miRDB (<http://www.mirdb.org/>) [45], and miRTarBase. Target genes present in these three databases were considered. Meanwhile, differentially expressed mRNAs in GC were identified using the edgeR package with RNA-seq datasets from NCBI. The intersection of the target mRNAs and the differentially expressed mRNAs was determined to ascertain the potential target genes of the miRNAs.

2.6. Construction of ceRNA Network and Identification of Hub Genes

The dysregulated circRNA transcripts, interacting miRNAs, and associated differentially expressed mRNAs in GC were utilized to establish a circRNA-miRNA-mRNA regulatory network based on the ceRNA theory. Subsequently, Cytoscape software (v3.8.0) was employed to visualize the network [46].

To further identify genes closely associated with GC, the target genes were extracted to construct a PPI network using the online database STRING (<https://cn.string-db.org/>) [47]. Direct physical interactions and indirect functional correlations between proteins were determined based on co-localization, co-expression, text mining, and other factors. Hub genes were identified from the most significant modules using the MCODE plug-in of Cytoscape 3.8.0.

2.7. Survival Analysis and Expression of Hub Genes

To confirm and ascertain the potential prognostic significance of the identified hub genes in GC, survival analysis was conducted using the Kaplan-Meier method. Clinical data of GC patients, encompassing molecular subtype, survival status, stage, grade, and overall survival time, were obtained from the TCGA database using the R package ‘TCGAbiolinks’ [48]. The log-rank test was employed to evaluate significant differences between survival curves, with a *p*-value < 0.05 deemed statistically significant.

Subsequently, we compared the expression levels of hub genes across different clinical T staging groups (T1–T4). GC data from TCGA was acquired from the Broad Institute’s Firehose pipeline using the getFirehoseData function from the R package RTCGAToolbox [49]. The Wilcoxon rank-sum test was applied to identify statistically significant differences, with a *p*-value < 0.05 considered significant, and boxplots were utilized for visualization.

3. RESULTS

3.1. Characterization of Back-spliced Circular RNA Transcripts in Human GC

In our systematic analysis of circRNAs associated with gastric cancer, we examined RNA-seq datasets sequenced from three pairs of gastric cancer and adjacent normal tis-

sues. Using CIRI2 software, with the criterion of having a minimum of five supporting reads on each head-to-tail junction in at least one sample, we identified a total of 24,068 candidate circRNAs. According to the source statistics, 87% were exonic, 11% were intronic, and 2% were intergenic circRNAs (Fig. 2A). Focusing on exon-derived circRNAs, we further reconstructed a total of 18,398 circular transcripts using CircAST. Among the successfully reconstructed back-spliced circRNAs, approximately 85% produced only one isoform, 11% produced two isoforms, and the remaining 4% produced three or more isoforms (Fig. 2B). These reconstructed full-length circRNAs were found to be unevenly distributed across all human chromosomes, with chromosomes 1, 2, 3, 12, and 10 containing the largest number of unique circRNA transcripts, while only a few transcripts were located on chromosome Y (Fig. 2C).

We conducted an analysis of the sequences of the full-length circular transcripts. Consistent with previous studies, both upstream and downstream flanking introns exhibited significantly longer lengths compared to randomly selected introns (Fig. 2D). Long flanking introns are known to contain repetitive and reverse complement elements that can facilitate circularization, making them a characteristic feature of circRNA biogenesis [50]. Among all circRNA transcripts, nearly 58% were approximately 300–600 bp in length, while only 4% exceeded 1500 bp (Fig. 2E). Additionally, although 684 out of 18,398 reconstructed exonic circRNA transcripts contained only one annotated exon, the majority of circular transcript sequences consisted of multiple exons. Specifically, around 60% of the circular transcripts contained 2–4 exons, while less than 5% were composed of more than 10 exons (Fig. 2F). Furthermore, as depicted in Fig. (2G), exons from circRNA transcripts with a single back-spliced exon were notably longer than the circularized exons from circRNA transcripts with multiple exons, indicating that exon length could potentially influence exon(s) circularization efficiency.

We also delved into the internal alternative splicing events within these circRNAs. As shown in Fig. (2H), approximately 29.1% of the circular transcripts underwent exon skipping, implying that AS events within circRNAs are prevalent in gastric tissues. Further examination revealed that a significant portion (43.4%) of the AS events were exclusive to circRNAs and did not appear in their linear mRNAs, suggesting that circRNAs do not share the same splicing mechanism as their linear counterparts. Interestingly, alternatively spliced exons exhibited a shorter length compared to other exons in circRNAs, indicating that exon length might also influence AS in circRNAs (Fig. 2I).

We conducted a series of analyses to characterize the genomic features of these back-spliced circular transcripts. The reconstructed circRNA transcripts originated from 4,439 different genes, and it was observed that these genes exhibited a high number of exons compared to genes expressing only linear transcripts (Fig. 3A). The majority of the circular transcripts (18,386 out of 18,398) consisted of exons located in the middle of RefSeq genes, with only a small fraction containing the first or last exons (Fig. 3B), indicating that

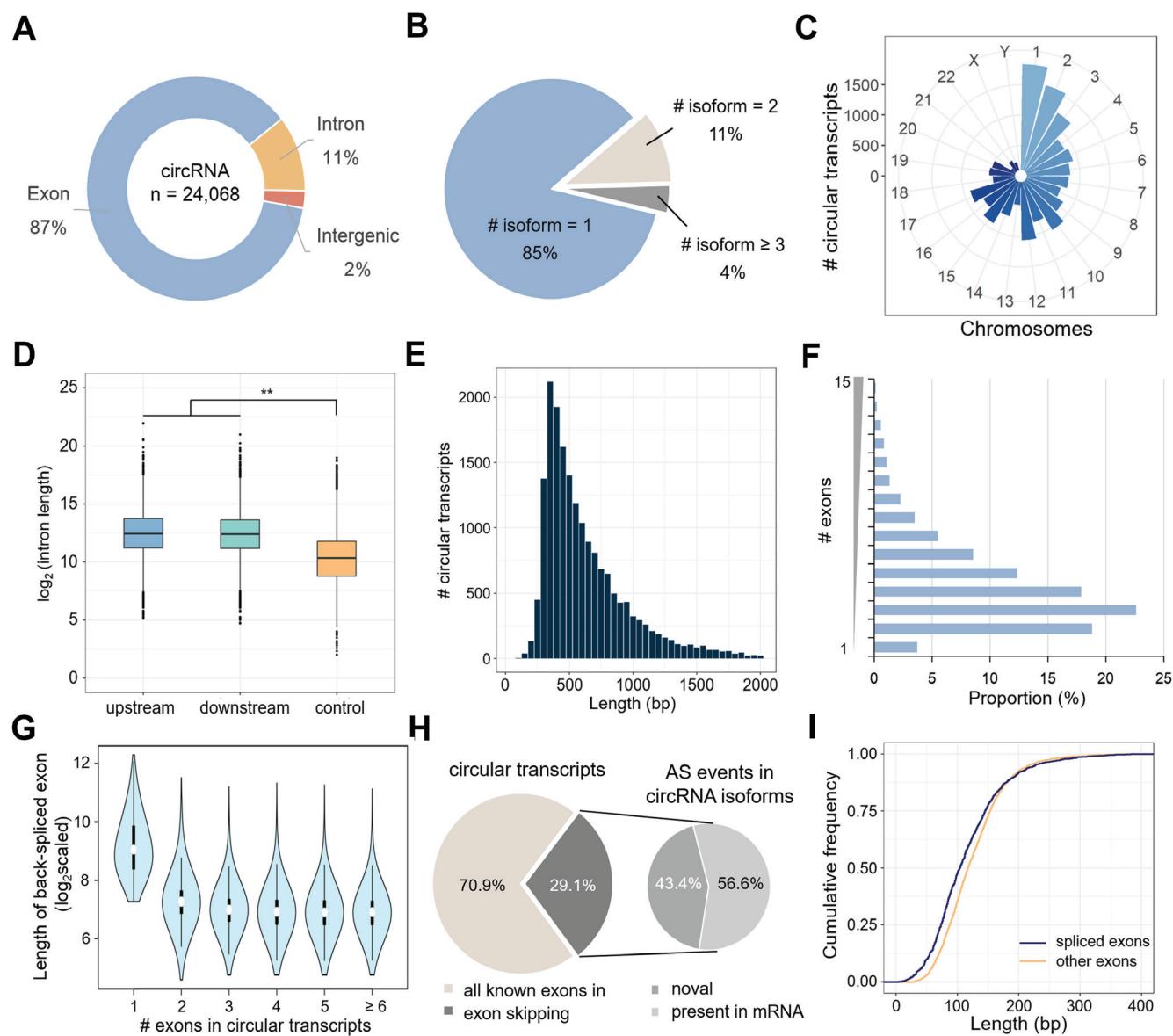


Fig. (2). The landscape of reconstructed full-length circRNA transcripts in human gastric cancer (GC). (A) Genomic origin of circRNAs in GC. (B) Number distribution of circRNA isoforms in each back-splicing event. (C) Number distribution of circular transcripts on human chromosomes. (D) Length distribution of the flanking introns that bracket exonic circRNAs. Both upstream (blue box) and downstream (green box) flanking introns are significantly longer than control introns (yellow box). ** $p < 0.01$, Wilcoxon rank-sum test. (E) Length distribution of circRNA transcripts. (F) Number distribution of exons in circular transcripts. More than 95% of circRNAs contained multiple exons. (G) Length distribution of back-spliced exons (y-axis) from circular transcripts composed of different numbers of exons (x-axis). (H) Reconstructed circRNA transcripts can be divided into isoforms composed of all known exons and isoforms with exon skipping. For circular isoforms with exon skipping, 43.4% of AS events in these transcripts were present only in circRNAs, not found in mRNAs. (I) Cumulative frequency of exon length in circRNA transcripts. The alternatively spliced exons (blue line) are shorter than other exons (yellow line) in circRNAs. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

circRNA formation is typically closely linked to RNA splicing. Among the circRNA-producing genes, approximately 34.6% (1536) generated only a single circular transcript, while notably, 9.0% (399) yielded ten or more circular isoforms (Fig. 3C). One notable example was the gene DHTKD1, situated on chromosome 10, which generated 88 distinct circular transcripts. Furthermore, we observed a proportional increase in the number of circular transcripts with the number of annotated exons per gene (Fig. 3D), leading to significant diversity in the number of circular isoforms.

3.2. Characteristics of the circRNA Expression Profiles at Isoform Level Between GC Tissues and Adjacent Paracancerous Tissues

Many circRNA transcripts exhibited significant differences in their expression levels among individuals. In order to pinpoint the circular transcripts that might play crucial roles in the genesis and progression of gastric cancer tumors, we focused solely on back-spliced junctions supported by a minimum of five reads in at least two samples of gastric can-

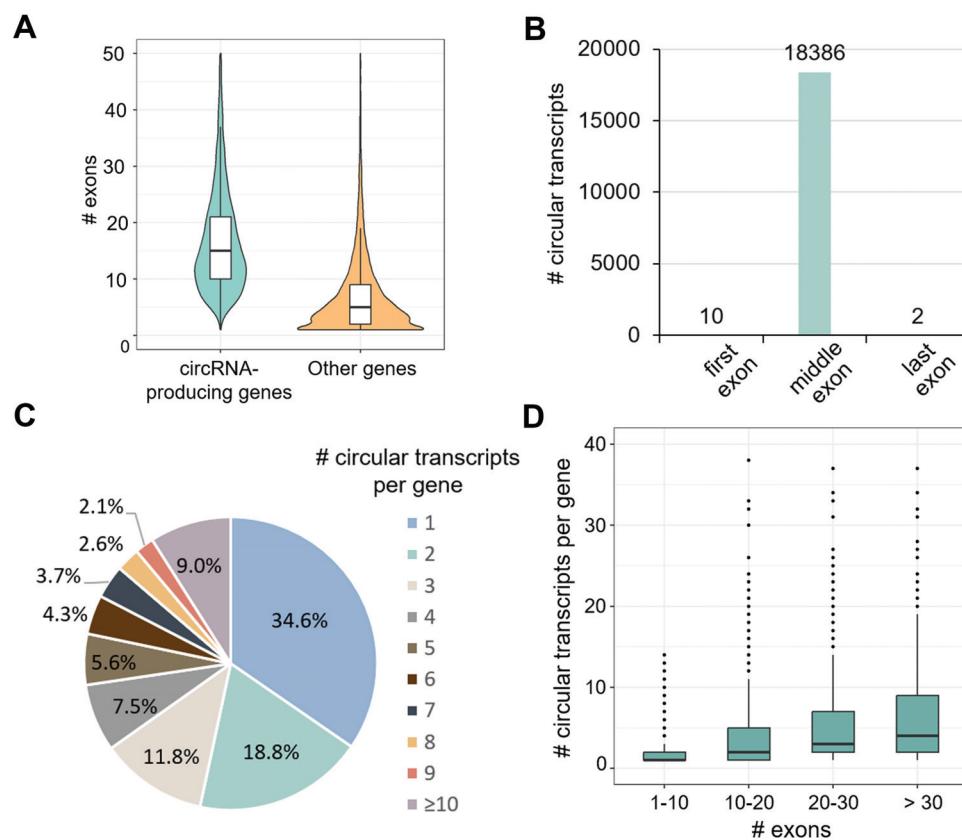


Fig. (3). Characteristics of circRNA-producing genes in GC. **(A)** Number distribution of exons in circRNA-producing genes. **(B)** Distribution of back-spliced exons across genomic regions. The vast majority of back-spliced exons are situated in the middle region of genes, whereas the first and last exons are rarely included in circRNAs. **(C)** Number distribution of circRNA transcripts produced per gene. **(D)** Number distribution of circular transcripts (y-axis) from circRNA-producing genes composed of different numbers of exons (x-axis). Genes with more exons tend to generate more circRNA isoforms. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

cer tissues and their paired adjacent normal tissues. Consequently, a total of 8024 circRNA transcripts, considered to be high-confidence, were included in our analysis (Fig. 4A). Among these, 2200 transcripts were exclusively identified in gastric cancer tissues, 1562 were exclusively identified in adjacent paracancerous tissues, and 4262 were detected in both groups (Fig. 4B). These circular transcripts were derived from 2607 genes in cancer tissues and 2374 genes in adjacent normal tissues, indicating that, obviously, more genes are involved in generating circRNAs in tumor tissues (Fig. 4C).

We investigated the expression patterns of these circRNA transcripts. In general, the expression levels of most circular transcripts were relatively low, with predominant FPKM values below 10. While a significant portion of circRNA-producing genes generated only one circular transcript, we observed that these transcripts were expressed at significantly higher levels than others in both normal and cancer groups (Fig. 4D). To ascertain whether a circRNA host gene harbors a circular transcript that exerts dominant expression, we selected genes with multiple circular transcripts and calculated the expression ratio between the two most abundant circRNA transcripts (referred to as the major and minor circular transcripts). As depicted in Fig. (4E), nearly half of the studied genes in both groups had a major circular transcript that

was expressed at least twice as much as the other circular transcript, and 5% were even ten times higher, suggesting that for a substantial fraction of genes, despite their tendency to produce two or more alternatively spliced circRNA isoforms, they typically have a dominant major circular transcript.

We further scrutinized the full-length sequences of these most abundant circRNA transcripts and found that approximately 8.9% and 9.7% of the circular transcripts underwent exon skipping in the normal group and the cancer group, respectively (Fig. 4F). This indicates that alternative splicing events are present in circRNAs, highlighting the necessity to identify their precise full-length sequences when conducting functional analysis of circRNAs in human gastric cancer. Additionally, in both groups, major circular transcripts tended to contain fewer exons (Fig. 4G) and were consequently shorter than others (Fig. 4H), suggesting that long circular transcripts are less likely to maintain high expression levels. When comparing the major circular transcripts between normal and cancer groups, we observed that approximately 66% of these transcripts, which were dominantly expressed in the normal group, maintained their dominance in the cancer group, while 12% of these transcripts transitioned to a minor circular form (Fig. 4I). This observation indicates that

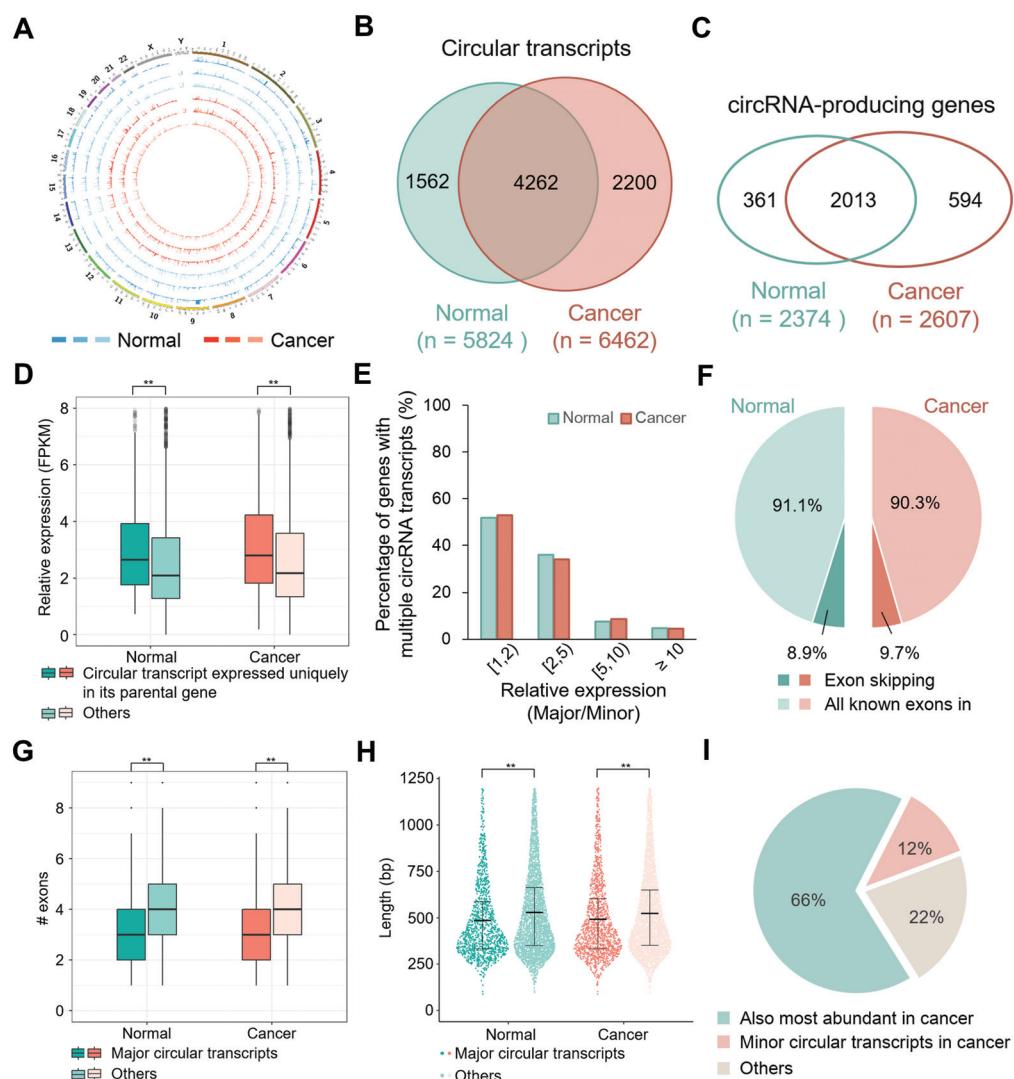


Fig. (4). Expression profiles of circRNA transcripts in GC. (A) Reconstructed circular transcripts from 3 GC tissues and 3 paired adjacent normal tissues were derived from different chromosomes. Normalized expression (FPKM) of circRNA isoforms is shown in the circus plot. (B) Venn diagrams of unique or common circRNA transcripts between normal and cancer tissues. (C) Overlap of circRNA-producing genes in normal and cancer tissues. (D) Expression distribution of circRNA transcripts in normal and cancer groups. Circular transcripts in each group were divided into two categories based on whether the single circRNA transcript is uniquely expressed in its parental gene. (E) Percentage of genes (y-axis) categorized by the relative expression ratio of the major to the minor circRNA transcript (x-axis). (F) Percentage of most abundant circRNA isoforms containing all known exons between back splice sites in normal and cancer groups. (G) Number distribution of exons contained in major circRNA transcripts in both groups. (H) Length distribution of major circRNA transcripts in both groups. (I) Tracking investigation on expression of most abundant circRNA transcripts in normal tissues. ** $p < 0.01$, Wilcoxon rank-sum test. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

circRNA transcripts with high abundance may have a significant functional impact on gastric tissue.

3.3. Differential Expression Analysis of Circular RNA Transcripts in Human Gastric Cancer

In order to identify key circRNA transcripts associated with GC tumorigenesis, we conducted a differential expression analysis between the tumor and adjacent normal tissues. Utilizing normalized relative abundance \log_2 ($FPKM + 1$), the expression variation of the circRNA transcripts between the two sample sets was depicted via scatter plots (Fig. 5A). Applying criteria of $|\log_2 FC| > 2$ and an adjusted p -value < 0.05 , we identified a total of 528 differentially expressed

circRNA transcripts, comprising 351 upregulated and 177 downregulated transcripts. The screening results, demonstrating significant statistical differences, were visually represented in a volcano plot (Fig. 5B). The full-length sequence information of the top 10 significantly upregulated and top 10 significantly downregulated circRNA transcripts, ranked by FC, is listed in Table S1. Additionally, we examined the distribution of differentially expressed circular transcripts across human chromosomes and observed that the 528 circRNA transcripts were widely dispersed across all chromosomes (Fig. 5C). Notably, chromosome 19 contained the highest number of dysregulated circRNA transcripts, while the Y chromosome had the fewest. To gain a clearer understanding of the similarities and differences among the

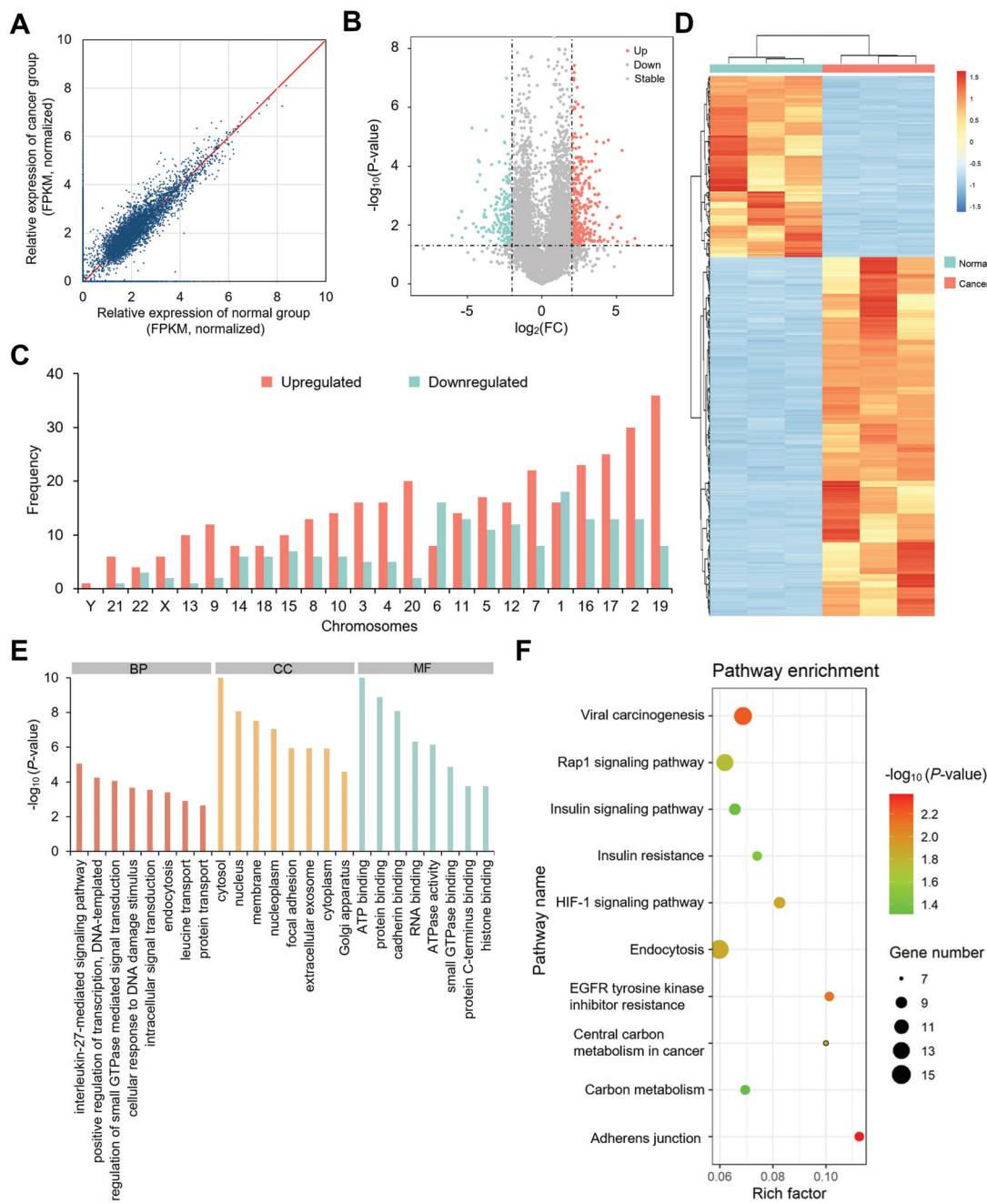


Fig. (5). Differential expression analysis of circRNA transcripts and pathway enrichment analysis of their host genes. **(A)** Scatter plots show the relative expression of circRNA transcripts in normal and cancer tissues. The points above and below the red line represent up-regulated and down-regulated circRNA transcripts, respectively. **(B)** Volcano plots of differentially expressed circRNA transcripts. The significantly up-regulated and down-regulated circRNA transcripts are indicated with red and green dots, respectively (adjusted p -value < 0.05 , raw p -value derived from Student's t -test, and $|\log_2 \text{FC}| > 2$). **(C)** Frequency distribution of the dysregulated circRNA transcripts between normal and cancer groups on chromosomes. **(D)** Heat map of the differentially expressed circRNA transcripts. Each row represents a circRNA transcript, and each column corresponds to one individual sample. **(E)** GO function enrichment analysis on host genes of dysregulated circRNA transcripts based on the DAVID GO analysis result. **(F)** Results of KEGG pathway enrichment analysis. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

six samples, all the differentially expressed circRNA transcripts were clustered and visually presented in a heat map. As illustrated in Fig. (5D), the three gastric cancer samples clustered together and were distinctly separate from the three normal samples, indicating that the circRNA expression profiles were distinguishable between these two groups.

Given the regulatory potential of circRNAs on the expression of their parental genes, we conducted GO and KEGG pathway enrichment analyses of the parental genes to elucidate the biological functions associated with differentially expressed circRNA transcripts. As illustrated in Fig. (5E), based on the prediction terms exported by DAVID, the

top three significantly enriched BP (biological process) terms were the interleukin-27-mediated signaling pathway, positive regulation of transcription, and regulation of small GTPase-mediated signal transduction. Additionally, the top three significantly enriched CC (cellular component) terms were associated with the cytosol, nucleus, and membrane, while the top three significantly enriched MF (molecular function) terms included ATP binding, protein binding, and cadherin binding. These enriched GO terms are largely implicated in the occurrence and progression of gastric cancer.

Furthermore, the KEGG pathway analysis revealed a total of 10 significantly enriched pathways. Among these, the top enriched pathways encompassed those related to adherens junction, viral carcinogenesis, and EGFR tyrosine kinase inhibitor resistance (Fig. 5F). The adherens junction pathway is involved in maintaining homeostatic cell signaling, and its disruption has been found to drive tumorigenesis [51]. Given that gastric carcinogenesis is closely linked to infectious events, the viral carcinogenesis pathway assumes obvious significance for GC development. Additionally, the EGFR tyrosine kinase inhibitor resistance pathway holds notable relevance to cancer. These signaling pathways collectively play pivotal roles in tumor development and the pathogenesis of GC.

3.4. Construction of the ceRNA Network

Previous reports have highlighted the role of circRNAs as miRNA sponges, thereby influencing gene expression and subsequently impacting GC progression [31, 32, 52]. However, due to the lack of accurate circRNA sequences in prior studies, there may have been potential errors in predicting miRNA binding sites within circRNA sequences. To accu-

rately explore the biological function of circRNAs in GC, we predicted circRNA-miRNA interactions based on the reconstructed full-length sequences of dysregulated circRNA transcripts and subsequently predicted the miRNA target genes (mRNAs). Following several screening steps, we identified 19 dysregulated circRNA transcripts, 40 related miRNAs, and 65 target genes (mRNAs) to construct the ceRNA network. Illustrated in Fig. (6), the dysregulated circRNA transcripts circNIBAN2 and circTNRC18 exhibited the largest number of predicted interactions with miRNAs, suggesting their potentially significant regulatory roles in the progression of gastric cancer.

Conducting GO enrichment analysis of the target genes revealed significantly enriched BP terms related to collagen fibril organization, extracellular matrix organization, and positive regulation of transcription from RNA polymerase II promoter, among others. Additionally, the KEGG pathway analysis results demonstrated pathways of significance, including protein digestion and absorption, the PI3K-Akt signaling pathway, and microRNAs in cancer, among others. These identified terms and pathways are closely linked to GC progression [53-55].

3.5. Identification of Hub Genes Based on PPI Network and Potential Value to GC Prognosis

We generated a PPI network by inputting the aforementioned potential target genes into the STRING online database and visualizing it using Cytoscape (Fig. 7A). Genes exhibiting high connectivity within the network, often referred to as "hub genes", typically hold functional significance. Subsequently, we conducted an analysis of the total nodes using the MCODE plugin in Cytoscape and identified

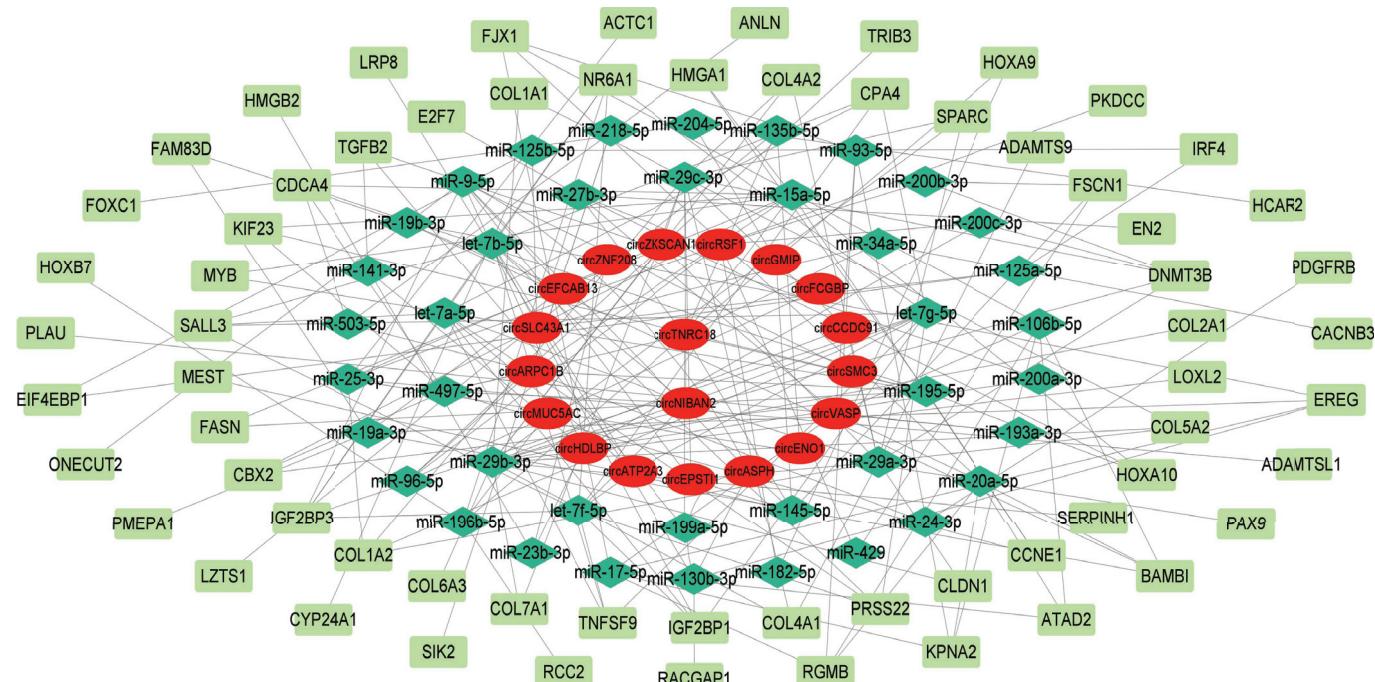


Fig. (6). A circRNA-miRNA-mRNA regulatory network. The ellipses represent dysregulated circRNA transcripts, the diamonds represent miRNAs, and the rectangles represent mRNAs. (A higher resolution / colour version of this figure is available in the electronic copy of the article)

significant cluster networks with an MCODE score exceeding 5. The analysis revealed a single module with an MCODE score of 10.2, comprising 11 nodes interconnected by 102 edges. Consequently, a total of 11 genes were identified as candidate hub genes in GC.

We performed survival analyses on these candidate hub genes, identifying five hub genes (COL5A2, PDGFRB, SPARC, COL1A2, and COL4A1) in GC, all of which consistently play prognostic roles in clinical outcomes. The regulatory axes of the circRNA-miRNA-hub gene interactions are depicted to elucidate the molecular mechanisms of these hub genes (Fig. 7B). In our survival analyses, we utilized Kaplan-Meier cumulative survival probability over time plots and compared them using the log-rank test. Our findings revealed that high expressions of COL5A2, PDGFRB, SPARC, COL1A2, and COL4A1 were correlated with poor overall survival among GC patients ($p < 0.05$). This suggests that these hub genes hold significant roles in the onset and progression of GC and may serve as potential prognostic markers. Fig. (7C) illustrates the significant correlations of these five hub genes with GC prognosis.

We proceeded to investigate the expressions of the five hub genes across different clinical T stages of GC. Based on the pathological severity of the tumor, GC samples were categorized into the T1 group ($n = 19$), T2 group ($n = 80$), T3 group ($n = 168$), and T4 group ($n = 100$). Our analysis revealed that the expressions of the five genes in stages T2, T3, and T4 were significantly higher compared to those in stage T1. This observation further underscores their significant impact on the progression of GC tumors (Fig. 7D).

4. DISCUSSION

Gastric cancer is a prevalent, aggressive, malignant tumor with a poor prognosis and high mortality. Despite recent advances in diagnosis and treatment, the molecular mechanisms underlying GC remain unclear, and the prognosis is still poor. Therefore, further studies are urgently needed to elucidate the precise molecular mechanisms of GC, identify and validate useful biomarkers, and discover effective therapeutic targets to improve the prognosis of GC patients. Within the realm of RNA research, circRNA has emerged as a promising potential biomarker for GC diagnosis, treatment, and prognosis evaluation, offering new hope for improved patient outcomes.

In this study, we constructed circRNA expression profiles using RNA-seq datasets with RNase R treatment from human GC tissues and paired adjacent normal tissues. Leveraging the CircAST tool developed by our group, we successfully reconstructed the full-length sequences of circRNAs and evaluated these circular transcripts at the isoform level, providing a foundation for subsequent accurate analyses. A total of 528 differentially expressed circRNA transcripts in GC were identified, with 351 upregulated and 177 downregulated. Using these dysregulated circRNA transcripts, we constructed a circRNA-miRNA-mRNA network to elucidate the circRNA-mediated ceRNA regulatory mechanism in GC. The ceRNA hypothesis, which describes an important molecular regulatory mechanism at the posttran-

scriptional level, was first proposed by Salmena *et al.* in 2011 and has since garnered extensive attention in tumorigenesis [56]. While the circRNA transcripts in the network are not yet fully functionally annotated, functional enrichment analysis of target genes with GO and KEGG revealed close relationships with GC, suggesting that circRNAs can act as miRNA sponges to regulate gene expression and subsequently affect GC progression. Subsequently, we constructed a PPI network and identified 5 hub genes with the highest degrees of connectivity and significant association with GC patient outcomes, including COL5A2, PDGFRB, SPARC, COL1A2, and COL4A1. All of these genes have been previously reported to be involved in the tumorigenesis and metastasis of cancer, and they may aid researchers in better understanding the progression of GC.

The genes COL5A2, COL1A2, and COL4A1 belong to the collagen family. As the most crucial component of the extracellular matrix (ECM), collagen plays a pivotal role in creating the network within the tumor microenvironment and can be classified into types I–V [57]. COL5A2, a type V collagen, is a key component of the ECM and can interact with type I collagen to regulate fiber diameter during fiber development [58]. Up-regulation of COL5A2 has been observed in various cancer types, including tongue cancer, colon cancer, and pancreatic cancer [59–61], indicating a potential close association between COL5A2 and malignancy.

COL1A2, also known as collagen type I alpha 2 chain, exhibits different expression patterns, leading to distinct collagen-mediated effects in various human cancers [62]. Previous studies have demonstrated that COL1A2 is downregulated in several cancers, such as bladder cancer, melanoma, and head and neck cancer [63–65]. However, in other malignancies, such as hepatoma, pancreatic cancer, and ovarian cancer, COL1A2 is upregulated and may serve as a prognostic biomarker and therapeutic target for these cancers [66–68]. COL1A2 was also detected to be overexpressed in GC and directly associated with the prognosis of GC patients [69].

As a type IV collagen, COL4A1 is the most abundant component of the ECM and is an essential constituent of the basement membrane in human tissues and cell types [70]. It has been revealed that COL4A1 functions as a new oncogene in various tumor types and is involved in numerous malignancies. Recent findings have shown that COL4A1 is highly expressed in GC tissues and cells, and its knockdown has been demonstrated to exhibit tumor-suppressive effects on GC cells, suggesting that COL4A1 could serve as a valuable molecular marker for GC treatment [71].

The PDGFRB gene encodes a protein called platelet-derived growth factor receptor beta (PDGFR β), a member of the receptor tyrosine kinase (RTK) family [72]. PDGFRB regulates a wide range of crucial biological activities, including growth, proliferation, motility, and survival. Dysregulation of PDGFRB is closely associated with carcinogenesis [73]. It has been reported that PDGFRB is highly expressed in the tumor stroma in gliomas and can enhance the migratory capacity of glioma cells [74]. Additionally, it is associated with short prostate cancer-specific survival [75].

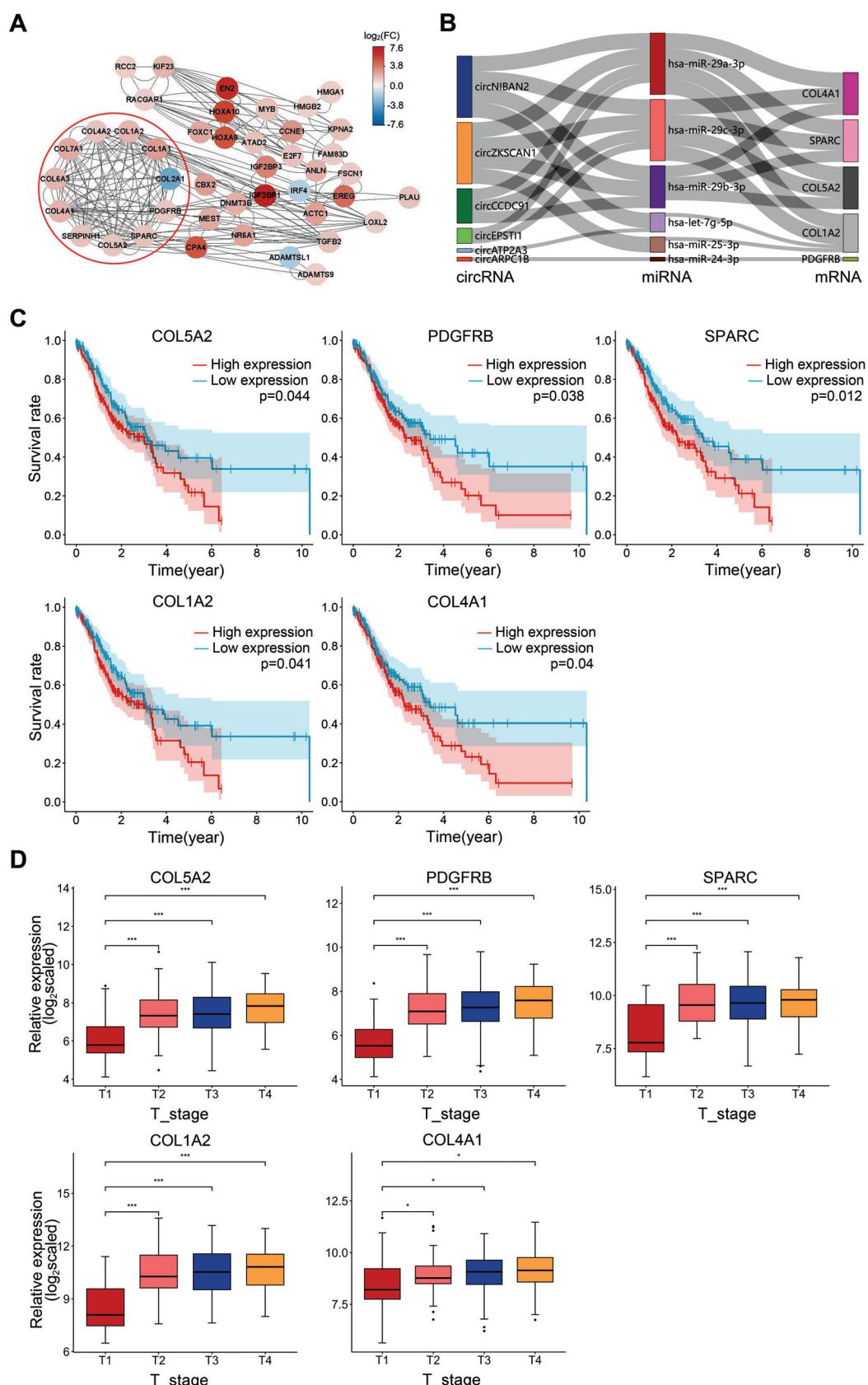


Fig. (7). Identification and analysis of hub genes in GC. **(A)** The protein-protein interaction (PPI) network. **(B)** circRNA-miRNA-hub gene regulatory axes. **(C)** Kaplan-Meier curves for survival analyses of five hub genes in GC patients (using TCGA dataset). **(D)** The different expression of five hub genes in T staging of GC patients (using TCGA dataset). * $p < 0.05$, *** $p < 0.001$, Wilcoxon rank-sum test. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

SPARC provides instructions for producing a cysteine-rich acidic matrix-associated protein. This protein is essential for the calcification of collagen in bone and is involved in modulating the cross-talk between cells and their surrounding ECM [76]. SPARC has been found to play a dual role in tumors. In cancers, such as gliomas and melanoma, high expression of SPARC promotes the survival, proliferation, and invasion of tumor cells [77, 78]. However, in certain types of cancer like medulloblastoma, prostate cancer, and ovarian cancer, overexpression of SPARC may inhibit tumor cell proliferation, invasion, and angiogenesis, suggesting it could act as a tumor suppressor gene [79-81].

Survival analysis revealed a negative correlation between the expression of the 5 hub genes and the overall survival of GC patients, with significantly higher expression in the T2-T4 stages compared to the T1 stage. This suggests that these hub genes may act as oncogenes and play critical roles in the occurrence and progression of GC. Additionally, we constructed a circRNA-miRNA-hub gene regulatory subnetwork based on the relationships identified in the circRNA-miRNA-mRNA network. This subnetwork unveiled six important circRNA transcripts that may be involved in GC carcinogenesis, offering new insights for further exploration of the disease's pathogenesis and presenting potential prognostic biomarkers or therapeutic targets that warrant future in-depth investigation. However, it is important to note that the results of the present study are solely based on bioinformatics analysis. The physiological mechanisms of the differential molecules identified in this study and their potential as GC biomarkers require further extensive investigation.

CONCLUSION

In conclusion, this study has unveiled a comprehensive landscape of full-length circRNA transcripts in GC, offering new perspectives on the pathogenesis and potential therapeutic strategies for this cancer. The novelty of this study lies in the reconstruction of circRNAs in GC samples into full-length circular transcripts and their quantification at the isoform level, laying a foundation for subsequent computational analysis. Our findings are expected to enhance our understanding of cancerogenesis and may prove valuable for further investigations into the biological functions of circRNAs in GC.

AUTHORS' CONTRIBUTION

QY carried out data analysis and interpretation of results. JL drafted the manuscript. RC was involved in visualization. JZ analysed the data. MHZ and YW collected the data. JC and MZ took part in data curation. TX contributed to the conceptualization. XS proposed the methodology. JW conceived and designed the study. All authors reviewed the results and approved the final version of the manuscript.

LIST OF ABBREVIATIONS

AS	=	Alternative Splicing
BP	=	Biological Process

BSJ	=	Back-Spliced Junction
CC	=	Cellular Component
ceRNA	=	Competitive Endogenous RNA
circRNA	=	Circular RNA
DAVID	=	Database for Annotation, Visualization and Integrated Discovery
FC	=	Fold Change
GC	=	Gastric Cancer
MF	=	Molecular Function
miRNA	=	microRNA
PPI	=	Protein-Protein Interaction

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Public data used in this work are openly available in the GEO and TCGA databases. The analysis codes supporting the conclusions of this article are available from the corresponding author.

FUNDING

Research reported in this publication was supported by the National Natural Science Foundation of China (Grant Nos. 61901225, 62273175, 61973155, and 62003164) and Jiangsu Provincial Key Research and Development Program (Grant No. BE2022843).

CONFLICT OF INTEREST

Dr. Xiaofeng Song is the editorial advisory board member of the journal CBIO.

ACKNOWLEDGEMENTS

The English language of the article was improved with ChatGPT.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

- [1] Sung H, Ferlay J, Siegel RL, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; 71(3): 209-49. <http://dx.doi.org/10.3322/caac.21660> PMID: 33538338
- [2] Thrift AP, El-Serag HB. Burden of gastric cancer. *Clin Gastroenterol Hepatol* 2020; 18(3): 534-42. <http://dx.doi.org/10.1016/j.cgh.2019.07.045> PMID: 31362118
- [3] Tang S, Daram SR, Wu R, Bhajee F. Pathogenesis, diagnosis, and management of gastric ischemia. *Clin Gastroenterol Hepatol* 2014; 12(2): 246-252.e1. <http://dx.doi.org/10.1016/j.cgh.2013.07.025> PMID: 23920033
- [4] Tan P, Yeoh KG. Genetics and molecular pathogenesis of gastric adenocarcinoma. *Gastroenterology* 2015; 149(5): 1153-1162.e3. <http://dx.doi.org/10.1053/j.gastro.2015.05.059> PMID: 26073375
- [5] Uemura N, Okamoto S, Yamamoto S, *et al.* *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med* 2001; 345(11): 784-9. <http://dx.doi.org/10.1056/NEJMoa001999> PMID: 11556297
- [6] Sitarz R, Skierucha M, Mielko J, Offerhaus J, Maciejewski R, Polkowski W. Gastric cancer: Epidemiology, prevention, classification, and treatment. *Cancer Manag Res* 2018; 10: 239-48. <http://dx.doi.org/10.2147/CMAR.S149619> PMID: 29445300
- [7] Ajani JA, Lee J, Sano T, Janjigian YY, Fan D, Song S. Gastric adenocarcinoma. *Nat Rev Dis Primers* 2017; 3(1): 17036. <http://dx.doi.org/10.1038/nrdp.2017.36> PMID: 28569272
- [8] Seenevassen L, Bessède E, Mégraud F, Lehours P, Dubus P, Varon C. Gastric cancer: Advances in carcinogenesis research and new therapeutic strategies. *Int J Mol Sci* 2021; 22(7): 3418. <http://dx.doi.org/10.3390/ijms22073418> PMID: 33810350
- [9] Kung CH, Jestin Hannan C, Linder G, *et al.* Impact of surgical resection rate on survival in gastric cancer: Nationwide study. *BJS Open* 2021; 5(2): zraa017. <http://dx.doi.org/10.1093/bjsopen/zraa017> PMID: 33688944
- [10] Guan WL, He Y, Xu RH. Gastric cancer treatment: Recent progress and future perspectives. *J Hematol Oncol* 2023; 16(1): 57. <http://dx.doi.org/10.1186/s13045-023-01451-3> PMID: 37245017
- [11] Qiu H, Cao S, Xu R. Cancer incidence, mortality, and burden in China: a time-trend analysis and comparison with the United States and United Kingdom based on the global epidemiological data released in 2020. *Cancer Commun* 2021; 41(10): 1037-48. <http://dx.doi.org/10.1002/cac2.12197> PMID: 34288593
- [12] Chan WL, Yun HKB, Cheung EE, *et al.* Association of sarcopenia with severe chemotherapy toxicities and survival in patients with advanced gastric cancer. *Oncologist* 2024; oyae123. <http://dx.doi.org/10.1093/onco/oya123> PMID: 38885304
- [13] Janiczek-Polewska M, Szylberg Ł, Malicki J, Marszałek A. Role of interleukins and new perspectives in mechanisms of resistance to chemotherapy in gastric cancer. *Biomedicines* 2022; 10(7): 1600. <http://dx.doi.org/10.3390/biomedicines10071600> PMID: 35884907
- [14] Janjigian YY, Kawazoe A, Yañez P, *et al.* The KEYNOTE-811 trial of dual PD-1 and HER2 blockade in HER2-positive gastric cancer. *Nature* 2021; 600(7890): 727-30. <http://dx.doi.org/10.1038/s41586-021-04161-3> PMID: 34912120
- [15] Huang SC, Ng KF, Yeh TS, *et al.* Subtraction of Epstein-Barr virus and microsatellite instability genotypes from the Lauren histotypes: Combined molecular and histologic subtyping with clinicopathological and prognostic significance validated in a cohort of 1,248 cases. *Int J Cancer* 2019; 145(12): 3218-30. <http://dx.doi.org/10.1002/ijc.32215> PMID: 30771224
- [16] Kim ST, Cristescu R, Bass AJ, *et al.* Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med* 2018; 24(9): 1449-58. <http://dx.doi.org/10.1038/s41591-018-0101-z> PMID: 30013197
- [17] Shang Q, Yang Z, Jia R, Ge S. The novel roles of circRNAs in human cancer. *Mol Cancer* 2019; 18(1): 6. <http://dx.doi.org/10.1186/s12943-018-0934-6> PMID: 30626395
- [18] Cocquerelle C, Mascréz B, Hétuin D, Bailleul B. Mis-splicing yields circular RNA molecules. *FASEB J* 1993; 7(1): 155-60. <http://dx.doi.org/10.1096/fasebj.7.1.7678599> PMID: 7678599
- [19] Memczak S, Jens M, Elefsinioti A, *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013; 495(7441): 333-8. <http://dx.doi.org/10.1038/nature1928> PMID: 23446348
- [20] Qu S, Yang X, Li X, *et al.* Circular RNA: A new star of noncoding RNAs. *Cancer Lett* 2015; 365(2): 141-8. <http://dx.doi.org/10.1016/j.canlet.2015.06.003> PMID: 26052092
- [21] Hansen TB, Jensen TI, Clausen BH, *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* 2013; 495(7441): 384-8. <http://dx.doi.org/10.1038/nature11993> PMID: 23446346
- [22] Ashwal-Fluss R, Meyer M, Pamudurti NR, *et al.* circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 2014; 56(1): 55-66. <http://dx.doi.org/10.1016/j.molcel.2014.08.019> PMID: 25242144
- [23] Li Z, Huang C, Bao C, *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* 2015; 22(3): 256-64. <http://dx.doi.org/10.1038/nsmb.2959> PMID: 25664725
- [24] Meyer KD, Patil DP, Zhou J, *et al.* 5' UTR m(6)A promotes cap-independent translation. *Cell* 2015; 163(4): 999-1010. <http://dx.doi.org/10.1016/j.cell.2015.10.012> PMID: 26593424
- [25] Zhou J, Wan J, Gao X, Zhang X, Jaffrey SR, Qian SB. Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature* 2015; 526(7574): 591-4. <http://dx.doi.org/10.1038/nature15377> PMID: 26458103
- [26] Kristensen LS, Andersen MS, Stagsted LVW, Ebbesen KK, Hansen TB, Kjems J. The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 2019; 20(11): 675-91. <http://dx.doi.org/10.1038/s41576-019-0158-7> PMID: 31395983
- [27] Barrett SP, Salzman J. Circular RNAs: Analysis, expression and potential functions. *Development* 2016; 143(11): 1838-47. <http://dx.doi.org/10.1242/dev.128074> PMID: 27246710
- [28] Rybak-Wolf A, Stottmeister C, Glažar P, *et al.* Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell* 2015; 58(5): 870-85. <http://dx.doi.org/10.1016/j.molcel.2015.03.027> PMID: 25921068
- [29] Zhang H, Jiang L, Sun D, Hou J, Ji Z. CircRNA: A novel type of biomarker for cancer. *Breast Cancer* 2018; 25(1): 1-7. <http://dx.doi.org/10.1007/s12282-017-0793-9> PMID: 28721656
- [30] Ding L, Zhao Y, Dang S, *et al.* Circular RNA circ-DONSON facilitates gastric cancer growth and invasion via NURF complex dependent activation of transcription factor SOX4. *Mol Cancer* 2019; 18(1): 45. <http://dx.doi.org/10.1186/s12943-019-1006-2> PMID: 30922402
- [31] Zhu Z, Rong Z, Luo Z, *et al.* Circular RNA circNHS1 promotes gastric cancer progression through the miR-1306-3p/SIX1/vimentin axis. *Mol Cancer* 2019; 18(1): 126. <http://dx.doi.org/10.1186/s12943-019-1054-7> PMID: 31438963
- [32] Zhang J, Liu H, Hou L, *et al.* Circular RNA_LARP4 inhibits cell proliferation and invasion of gastric cancer by sponging miR-424-5p and regulating LATS1 expression. *Mol Cancer* 2017; 16(1): 151. <http://dx.doi.org/10.1186/s12943-017-0719-3> PMID: 28893265
- [33] Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* 2014; 15(7): 409. <http://dx.doi.org/10.1186/s13059-014-0409-z> PMID: 25070500
- [34] Wu J, Li Y, Wang C, *et al.* CircAST: Full-length assembly and quantification of alternatively spliced isoforms in circular RNAs. *Genom Proteom Bioinform* 2019; 17(5): 522-34. <http://dx.doi.org/10.1016/j.gpb.2019.03.004> PMID: 32007626
- [35] Zheng Y, Ji P, Chen S, Hou L, Zhao F. Reconstruction of full-length circular RNAs enables isoform-level quantification. *Genome Med* 2019; 11(1): 2. <http://dx.doi.org/10.1186/s13073-019-0614-1> PMID: 30660194
- [36] Hossain MT, Zhang J, Reza MS, Peng Y, Feng S, Wei Y. Reconstruction of full-length circRNA sequences using chimeric alignment information. *Int J Mol Sci* 2022; 23(12): 6776. <http://dx.doi.org/10.3390/ijms23126776> PMID: 35743218
- [37] Zhang J, Zhang H, Ju Z, *et al.* JCcirc: CircRNA full-length sequence assembly through integrated junction contigs. *Brief Bioinform* 2023; 24(6): bbad363. <http://dx.doi.org/10.1093/bib/bbad363> PMID: 37833842
- [38] Zhang Y, Liu W, Feng W, *et al.* Identification of 14 differentially-expressed metabolism-related genes as potential targets of gastric

- cancer by integrated proteomics and transcriptomics. *Front Cell Dev Biol* 2022; 10: 816249.
<http://dx.doi.org/10.3389/fcell.2022.816249> PMID: 35265615
- [39] Gao Y, Zhang J, Zhao F. Circular RNA identification based on multiple seed matching. *Brief Bioinform* 2018; 19(5): 803-10.
<http://dx.doi.org/10.1093/bib/bbx014> PMID: 28334140
- [40] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4(1): 44-57.
<http://dx.doi.org/10.1038/nprot.2008.211> PMID: 19131956
- [41] Kozomara A, Griffiths-Jones S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014; 42(D1): D68-73.
<http://dx.doi.org/10.1093/nar/gkt1181> PMID: 24275495
- [42] Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in drosophila. *Genome Biol* 2003; 5(1): R1.
<http://dx.doi.org/10.1186/gb-2003-5-1-r1> PMID: 14709173
- [43] Huang HY, Lin YCD, Cui S, et al. miRTarBase update 2022: An informative resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res* 2022; 50(D1): D222-30.
<http://dx.doi.org/10.1093/nar/gkab1079> PMID: 34850920
- [44] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; 120(1): 15-20.
<http://dx.doi.org/10.1016/j.cell.2004.12.035> PMID: 15652477
- [45] Wong N, Wang X. miRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res* 2015; 43(D1): D146-52.
<http://dx.doi.org/10.1093/nar/gku1104> PMID: 25378301
- [46] Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13(11): 2498-504.
<http://dx.doi.org/10.1101/gr.1239303> PMID: 14597658
- [47] Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017; 45(D1): D362-8.
<http://dx.doi.org/10.1093/nar/gkw937> PMID: 27924014
- [48] Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016; 44(8): e71.
<http://dx.doi.org/10.1093/nar/gkv1507> PMID: 26704973
- [49] Samur MK. RTCGAToolbox: A new tool for exporting TCGA Firehose data. *PLoS One* 2014; 9(9): e106397.
<http://dx.doi.org/10.1371/journal.pone.0106397> PMID: 25181531
- [50] Ivanov A, Memczak S, Wyler E, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* 2015; 10(2): 170-7.
<http://dx.doi.org/10.1016/j.celrep.2014.12.019> PMID: 25558066
- [51] Yu Y, Elble R. Homeostatic signaling by cell-cell junctions and its dysregulation during cancer progression. *J Clin Med* 2016; 5(2): 26.
<http://dx.doi.org/10.3390/jcm5020026> PMID: 26901232
- [52] Liu L, Wang H, Yu S, et al. An update on the roles of circRNA-ZFR in human malignant tumors. *Front Cell Dev Biol* 2022; 9: 806181.
<http://dx.doi.org/10.3389/fcell.2021.806181> PMID: 35186956
- [53] Zhou ZH, Ji CD, Xiao HL, Zhao HB, Cui YH, Bian XW. Reorganized collagen in the tumor microenvironment of gastric cancer and its association with prognosis. *J Cancer* 2017; 8(8): 1466-76.
<http://dx.doi.org/10.7150/jca.18466> PMID: 28638462
- [54] Andreuzzi E, Capuano A, Poletti E, et al. Role of extracellular matrix in gastrointestinal cancer-associated angiogenesis. *Int J Mol Sci* 2020; 21(10): 3686.
<http://dx.doi.org/10.3390/ijms21103686> PMID: 32456248
- [55] Singh SS, Yap WN, Arfuso F, et al. Targeting the PI3K/Akt signalling pathway in gastric carcinoma: A reality for personalized medicine? *World J Gastroenterol* 2015; 21(43): 12261-73.
<http://dx.doi.org/10.3748/wjg.v21.i43.12261> PMID: 26604635
- [56] Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? *Cell* 2011; 146(3): 353-8.
<http://dx.doi.org/10.1016/j.cell.2011.07.014> PMID: 21802130
- [57] Walma DAC, Yamada KM. The extracellular matrix in development *Dev Camb Engl* 2020; 147(10):
- [58] http://dx.doi.org/10.1242/dev.175596
[59] Sun M, Chen S, Adams SM, et al. Collagen V is a dominant regulator of collagen fibrillogenesis: Dysfunctional regulation of structure and function in a corneal-stroma-specific *Col5a1* -null mouse model. *J Cell Sci* 2011; 124(23): 4096-105.
<http://dx.doi.org/10.1242/jcs.091363> PMID: 22159420
- [60] Chen HC, Tseng YK, Shu CW, et al. Differential clinical significance of COL 5A1 and COL 5A2 in tongue squamous cell carcinoma. *J Oral Pathol Med* 2019; 48(6): 468-76.
<http://dx.doi.org/10.1111/jop.12861> PMID: 30972812
- [61] Chivu-Economescu M, Necula LG, Matei L, et al. Collagen family and other matrix remodeling proteins identified by bioinformatics analysis as hub genes involved in gastric cancer progression and prognosis. *Int J Mol Sci* 2022; 23(6): 3214.
<http://dx.doi.org/10.3390/ijms23063214> PMID: 35328635
- [62] Berchtold S, Grünwald B, Krüger A, et al. Collagen type V promotes the malignant phenotype of pancreatic ductal adenocarcinoma. *Cancer Lett* 2015; 356(2)(Pt B): 721-32.
<http://dx.doi.org/10.1016/j.canlet.2014.10.020> PMID: 25449434
- [63] Wu J, Liu J, Wei X, et al. A feature-based analysis identifies *COL1A2* as a regulator in pancreatic cancer. *J Enzyme Inhib Med Chem* 2019; 34(1): 420-8.
<http://dx.doi.org/10.1080/14756366.2018.1484734> PMID: 30734598
- [64] Mori K, Enokida H, Kagara I, et al. CpG hypermethylation of collagen type I alpha 2 contributes to proliferation and migration activity of human bladder cancer. *Int J Oncol* 2009; 34(6): 1593-602.
<http://dx.doi.org/10.1007/s10915-009-0945-7> PMID: 19424577
- [65] Bonazzi VF, Nancarrow DJ, Stark MS, et al. Cross-platform array screening identifies *COL1A2*, *THBS1*, *TNFRSF10D* and *UCHL1* as genes frequently silenced by methylation in melanoma. *PLoS One* 2011; 6(10): e26121.
<http://dx.doi.org/10.1371/journal.pone.0026121> PMID: 22028813
- [66] Misawa K, Kanazawa T, Misawa Y, et al. Hypermethylation of collagen α2(I) gene (*COL1A2*) is an independent predictor of survival in head and neck cancer. *Cancer Biomark* 2012-2012; 10(3-4): 135-44.
<http://dx.doi.org/10.3233/CBM-2012-0242> PMID: 22674299
- [67] Ji J, Zhao L, Budhu A, et al. Let-7g targets collagen type I α2 and inhibits cell migration in hepatocellular carcinoma. *J Hepatol* 2010; 52(5): 690-7.
<http://dx.doi.org/10.1016/j.jhep.2009.12.025> PMID: 20338660
- [68] Pan Z, Li L, Fang Q, et al. Analysis of dynamic molecular networks for pancreatic ductal adenocarcinoma progression. *Cancer Cell Int* 2018; 18(1): 214.
<http://dx.doi.org/10.1186/s12935-018-0718-5> PMID: 30598639
- [69] Wu Y-H, Chang T-H, Huang Y-F, Huang H-D, Chou C-Y. *COL11A1* promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene* 2014; 33(26): 3432-40.
<http://dx.doi.org/10.1038/onc.2013.307> PMID: 23934190
- [70] Li J, Ding Y, Li A. Identification of *COL1A1* and *COL1A2* as candidate prognostic factors in gastric cancer. *World J Surg Oncol* 2016; 14(1): 297.
<http://dx.doi.org/10.1186/s12957-016-1056-5> PMID: 27894325
- [71] Zagaglia S, Selch C, Nisevic JR, et al. Neurologic phenotypes associated with *COL4A1* / 2 mutations. *Neurology* 2018; 91(22): e2078-88.
<http://dx.doi.org/10.1212/WNL.000000000006567> PMID: 30413629
- [72] Cui X, Shan T, Qiao L. Collagen type IV alpha 1 (*COL4A1*) silence hampers the invasion, migration and epithelial-mesenchymal transition (EMT) of gastric cancer cells through blocking Hedgehog signaling pathway. *Bioengineered* 2022; 13(4): 8972-81.
<http://dx.doi.org/10.1080/21655979.2022.2053799> PMID: 35297303
- [73] Steller EJA, Raats DA, Koster J, et al. PDGFRB promotes liver metastasis formation of mesenchymal-like colorectal tumor cells. *Neoplasia* 2013; 15(2): 204-IN30.
<http://dx.doi.org/10.1593/neo.121726> PMID: 23441134
- [74] Kim Y, Kim E, Wu Q, et al. Platelet-derived growth factor receptors differentially inform intertumoral and intratumoral heterogeneity. *Genes Dev* 2012; 26(11): 1247-62.
<http://dx.doi.org/10.1101/gad.193565.112> PMID: 22661233

- [74] Wallmann T, Zhang XM, Wallerius M, *et al.* Microglia induce PDGFRB expression in glioma cells to enhance their migratory capacity. *iScience* 2018; 9: 71-83.
<http://dx.doi.org/10.1016/j.isci.2018.10.011> PMID: 30384135
- [75] Hägglöf C, Hammarsten P, Josefsson A, *et al.* Stromal PDGFRbeta expression in prostate tumors and non-malignant prostate tissue predicts prostate cancer survival. *PLoS One* 2010; 5(5): e10747.
<http://dx.doi.org/10.1371/journal.pone.0010747> PMID: 20505768
- [76] Ucaryilmaz Metin C, Ozcan G. Comprehensive bioinformatic analysis reveals a cancer-associated fibroblast gene signature as a poor prognostic factor and potential therapeutic target in gastric cancer. *BMC Cancer* 2022; 22(1): 692.
<http://dx.doi.org/10.1186/s12885-022-09736-5> PMID: 35739492
- [77] Li X, Oh JS, Lee Y, *et al.* Albumin-binding photosensitizer capable of targeting glioma via the SPARC pathway. *Biomater Res* 2023; 27(1): 23.
<http://dx.doi.org/10.1186/s40824-023-00360-3> PMID: 36945032
- [78] Ruiz EM, Alhassan SA, Errami Y, *et al.* A predictive model of adaptive resistance to BRAF/MEK inhibitors in melanoma. *Int J Mol Sci* 2023; 24(9): 8407.
<http://dx.doi.org/10.3390/ijms24098407> PMID: 37176114
- [79] Bhoopathi P, Gondi CS, Gujrati M, Dinh DH, Lakka SS. SPARC mediates Src-induced disruption of actin cytoskeleton via inactivation of small GTPases Rho–Rac–Cdc42. *Cell Signal* 2011; 23(12): 1978-87.
<http://dx.doi.org/10.1016/j.cellsig.2011.07.008> PMID: 21798346
- [80] Enriquez C, Cancila V, Ferri R, *et al.* Castration-induced downregulation of SPARC in stromal cells drives neuroendocrine differentiation of prostate cancer. *Cancer Res* 2021; 81(16): 4257-74.
<http://dx.doi.org/10.1158/0008-5472.CAN-21-0163> PMID: 34185677
- [81] John B, Naczki C, Patel C, *et al.* Regulation of the bi-directional cross-talk between ovarian cancer cells and adipocytes by SPARC. *Oncogene* 2019; 38(22): 4366-83.
<http://dx.doi.org/10.1038/s41388-019-0728-3> PMID: 30765860