**Name of Journal:** *World Journal of Gastrointestinal Surgery*

**Manuscript NO:** 58574

**Manuscript Type:** ORIGINAL ARTICLE

***Basic Study***

**Identification of key genes controlling cancer stem cell characteristics in gastric cancer**

Huang C *et al*. Key genes controlling GC stem cell characteristics

Chao Huang, Ce-Gui Hu, Zhi-Kun Ning, Jun Huang, Zheng-Ming Zhu

**Chao Huang, Ce-Gui Hu, Zhi-Kun Ning, Jun Huang, Zheng-Ming Zhu,** Department of Gastrointestinal Surgery, The Second Affiliated Hospital of Nanchang University, Nanchang 330006, Jiangxi Province, China

**Author contributions:** Huang C and Hu CG designed the study and contributed equally to this work; Huang C, Hu CG, and Ning ZK collected the data; Huang C and Huang J analysed the data; Huang C wrote the manuscript with contribution from all authors; and all authors read and approved the final version of the paper.

**Supported by** the National Natural Science Foundation of China, No. 81560389; and Key Research and Development Program of Jiangxi Province, No. 20181BBG70015.

**Corresponding author: Zheng-Ming Zhu, MD, Chairman, Surgical Oncologist,** Department of Gastrointestinal Surgery, The Second Affiliated Hospital of Nanchang University, No. 1 Minde Road, Nanchang 330006, Jiangxi Province, China. zzm8654@163.com

**Received:** July 28, 2020

**Revised:** August 13, 2020

**Accepted:** October 12, 2020

**Published online:** November 27, 2020

**Abstract**

BACKGROUND

Self-renewal of gastric cancer stem cells (GCSCs) is considered to be the underlying cause of the metastasis, drug resistance, and recurrence of gastric cancer (GC).

AIM

To characterize the expression of stem cell-related genes in GC.

METHODS

RNA sequencing results and clinical data for gastric adenoma and adenocarcinoma samples were obtained from The Cancer Genome Atlas database, and the results of the GC mRNA expression-based stemness index (mRNAsi) were analyzed. Weighted gene coexpression network analysis was then used to find modules of interest and their key genes. Survival analysis of key genes was performed using the online tool Kaplan-Meier Plotter, and the online database Oncomine was used to assess the expression of key genes in GC.

RESULTS

mRNAsi was significantly upregulated in GC tissues compared to normal gastric tissues (*P* < 0.0001). A total of 16 modules were obtained from the gene coexpression network; the brown module was most positively correlated with mRNAsi. Sixteen key genes (*BUB1*, *BUB1B*, *NCAPH*, *KIF14*, *RACGAP1*, *RAD54L*, *TPX2*, *KIF15*, *KIF18B*, *CENPF*, *TTK*, *KIF4A*, *SGOL2*, *PLK4*, *XRCC2*, and *C1orf112*) were identified in the brown module. The functional and pathway enrichment analyses showed that the key genes were significantly enriched in the spindle cellular component, the sister chromatid segregation biological process, the motor activity molecular function, and the cell cycle and homologous recombination pathways. Survival analysis and Oncomine analysis revealed that the prognosis of patients with GC and the expression of three genes (*RAD54L*, *TPX2,* and *XRCC2*) were consistently related.

CONCLUSION

Sixteen key genes are primarily associated with stem cell self-renewal and cell proliferation characteristics. *RAD54L*, *TPX2*, and *XRCC2* are the most likely therapeutic targets for inhibiting the stemness characteristics of GC cells.

**Key Words:** Gastric cancer; Cancer stem cell; Key gene; The Cancer Genome Atlas database; Weighted gene coexpression network analysis; mRNA expression-based stemness index

**Citation:** Huang C, Hu CG, Ning ZK, Huang J, Zhu ZM. Identification of key genes controlling cancer stem cell characteristics in gastric cancer. *World J Gastrointest Surg* 2020; 12(11): 442-459

URL: https://www.wjgnet.com/1948-9366/full/v12/i11/442.htm

DOI: https://dx.doi.org/10.4240/wjgs.v12.i11.442

**Core Tip:** This study found that the 16 key genes identified play an important role in the maintenance of gastric cancer stem cells. In addition, *RAD54L*, *TPX2*, and *XRCC2* are most likely to be therapeutic targets for inhibiting the stemness characteristics of gastric cancer, providing a new idea for the treatment of gastric cancer.

**INTRODUCTION**

Gastric cancer (GC) is a common malignant tumor, and its mortality rate ranks third among cancers globally[1]. In China, GC is also the major cause of cancer-related death, and patients with advanced disease have lower survival rates and higher recurrence rates[2]. Cancer stem cells (CSCs), defined functionally rather than by cellular origin, have superior tumour initiation, growth, and metastatic potential than other tumor cells[3]. CSCs may be derived from normal stem cells that have acquired malignant mutations and have lost the ability to self-regulate cell proliferation[4]. CSCs have a strong ability of self-renewal and expand in a symmetrical splitting manner to excessively increase cell growth, ultimately leading to tumor formation[5,6]. CSCs can lead to cancer recurrence, metastasis, multidrug resistance, and radiation resistance by blocking G0 phase, resulting in new tumors[7]. CSCs can effectively protect cancer cells from apoptosis by activating DNA repair capabilities[8]. The CSC model proposes that tumor growth is driven by a small number of self-sustaining cells that have longevity, infinite proliferation, and an ability to differentiate into the entire heterogeneous population of the tumour[9].

Self-renewal of GC stem cells (GCSCs) is considered to be the underlying cause of the metastasis, drug resistance, and recurrence of GC[10,11]. *In vitro* studies with cultured GCSCs have found that these cells are more resistant to chemotherapy and radiotherapy[12,13], which may be due to the high expression of anti-apoptotic proteins, the improvement of DNA repair efficiency, and the alterations in cell cycle kinetics[9, 14]. The identification and targeted therapy of cancer stem cells are of great significance in the treatment of GC[15]. The mRNA expression-based stemness index (mRNAsi), which can be used as a quantitative representation of cancer cell stemness, is an indicator describing the degree of similarity between tumor cells and stem cells. Moreover, inhibiting DNA methyltransferase can reduce the accumulation of tumorigenic ability of GCSCs[16]. The transcriptomes and methylomes were analyzed on multiple platforms to quantify stemness, and the DNA methylation based stemness index (mDNAsi) and mRNAsi were obtained; this analysis was also applied to The Cancer Genome Atlas (TCGA) dataset to calculate their scores of those samples[17]. Weighted gene coexpression network analysis (WGCNA) determines the correlation between genes based on systematic biological methods; it is used not only to build gene networks and detect modules but also to identify key genes[18,19]. WGCNA has been widely used to screen key genes related to the clinical characteristics of many tumour types, such as pancreatic cancer and kidney cancer[20,21]. In addition, the genes that regulate the maintenance and proliferation of GCSCs remain unknown. This study aimed to identify key genes related to stemness by combining WGCNA with GC mRNAsi in TCGA, thus providing new ideas for the treatment of GC.

**MATERIALS AND METHODS**

***Data download and processing***

RNA sequencing results of 373 tissues and 348 human gastric adenomas and adenocarcinoma samples were obtained from TCGA database (https://portal.gdc.cancer.gov). The RNA-seq results of 30 normal samples and 343 cancer samples were merged into a matrix file using a script in the Perl language (http://www.perl.org/). We then converted the Ensembl ID in the matrix file to the gene name using the Ensembl database (http://www.ensembl.org/index.html) and the Perl language script. In addition, 406 pieces of clinical data were downloaded, and the relevant clinical data were collated and extracted using scripts in the Perl language. The calculation of mRNAsi was performed from the molecular spectrum of normal cells with different degrees of stemness[17].

***Screening of differentially expressed genes***

The statistical software R (version 3.6.0, https://www.r-project.org/), limma package, and Wilcoxon test were used to screen differentially expressed gene (DEG) expression data between normal gastric tissue and GC samples. The values of the genes with the same name were averaged and the genes with expression levels < 0.2 were deleted. The false discovery rate (FDR) < 0.05 and |log2 fold change| > 1 were used as screening criteria.

***Construction of WGCNA***

The WGCNA R package was used to perform the WGCNA. The normal sample was deleted first, and the data were checked next for missing values. If a value was missing, the data was deleted. First, the similarity matrix was constructed by calculating the correlation of all genes. Second, the WGCNA software package was used to select appropriate soft-thresholding power β to improve coexpression similarity and achieve a scale-free topology. Third, the adjacency of genes was transformed into a topological overlap matrix (TOM), and then the corresponding dissimilarity was calculated. Average linkage hierarchical clustering was conducted with TOM-based dissimilarity measurements, and the minimum size of the gene dendrogram was 30. Finally, the dissimilarity of the genes was calculated and module dendrograms were built.

***Identification of important modules and key genes***

Gene significance (GS) refers to the correlation between genes and sample traits. Statistical significance was determined using the associated *P* values. To reduce the number of modules, we set a threshold (< 0.25) to merge some modules that were highly similar. The gene modules were then analyzed in combination with the clinical phenotype. Module membership (MM) refers to the correlation between a module’s own genes and the gene expression profile. After selecting the module of interest, GS and MM for each key gene were calculated. We defined cor. gene MM > 0.8 and cor. gene GS > 0.5 as the threshold for the key genes of the module.

***Differential expression and correlation of key genes***

We used the ggpubr package of the R language to draw a box plot of differential gene expression to understand the differential expression of key genes. The pheatmap package was used to create heat maps of key genes. Correlation analysis of key genes was performed using the corrplot package.

***Protein-protein interaction network analysis of key genes***

To explore the interaction between key genes, we imported the key genes into the Search Tool for the Retrieval of Interacting Genes (STRING, http://string.embl.de/) database[22]. The protein-protein interaction (PPI) network of key genes, which included genes with a combined score of > 0.4, was constructed using STRING.

***Functional and pathway enrichment analyses of key genes***

To study the biological functions of the module genes and key genes, we used the clusterProfiler package for gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses[23]. *P* < 0.05 and FDR < 0.05 were used as thresholds for identifying significant GO terms and pathways. A histogram of GO enrichment results for the key genes was plotted using the ggplot2 package. GO includes cellular component (CC), biological process (BP), and molecular function (MF).

***Prognostic analysis of key genes***

To understand the relationship between the expression of key genes and prognosis, we used the online tool Kaplan–Meier Plotter (http://kmplot.com/) for survival analysis. A log-rank *P* value < 0.05 was considered statistically significant. In addition, we used the online database Oncomine (http://www.oncomine.com) to assess the expression of key genes in GC.

**RESULTS**

***Differential analysis of mRNAsi and screening of DEG***

The Wilcoxon test showed a significant difference in the mRNAsi of normal gastric tissue and that of gastric tumor tissue (*P* < 0.0001, Figure 1A). A total of 6685 DEGs were identified by Wilcoxon test analysis, among which 5479 were upregulated and 1206 downregulated. The 20 with the most significant upregulation and the 20 with the most significant downregulated were mapped using the pheatmap package (Figure 1B).

***WGCNA: Identification of the modules of interest***

To identify biologically important gene modules and better understand genes associated with gastric cancer stemness, we used the WGCNA package to construct a gene coexpression network. Ultimately, we obtained 16 modules (Figure 2A); the brown module was most positively correlated with mRNAsi with a correlation coefficient of 0.75. In addition, the turquoise module exhibited a high negative correlation with mRNAsi with a correlation coefficient of -0.77 (Figure 2B). Therefore, the brown module was used as the most interesting module, so we further analyzed this module.

***Identification of key genes***

According to the threshold of key gene screening, we screened 16 key genes in the mRNAsi group, namely, *BUB1*, *BUB1B*, *NCAPH*, *KIF14*, *RACGAP1*, *RAD54L*, *TPX2*, *KIF15*, *KIF18B*, *CENPF*, *TTK*, *KIF4A*, *SGOL2*, *PLK4*, *XRCC2*, and *C1orf112* (Figure 2C).

***Differential expression and correlation analysis of key genes***

The expression of key genes differed significantly between normal gastric tissue and tumor tissue. The expression of key genes in gastric tumors was significantly higher than that in normal gastric tissues (*P* < 0.001 for all, Figure 3A). There was a clear correlation between the expressions of key genes, and *CENPF* had the highest correlation with *KIF14* (Figure 3B).

***PPI network analysis of key genes***

The PPI network of key genes consisted of 16 nodes and 94 edges, including 16 upregulated genes (Figure 3C). In the PPI network, *TTK*, *TPX2*, *NCAPH*, *KIF15*, *CENPF*, and *BUB1* all had 14 connection nodes (Figure 3D).

***GO and KEGG analyses of key genes***

GO and KEGG pathway enrichment analyses were performed using the clusterProfiler package. The most significantly enriched GO terms were spindle (ontology: CC), sister chromatid segregation (ontology: BP), and motor activity (ontology: MF) (Figure 4). The significantly enriched KEGG pathways have cell cycle and homologous recombination (Table 1).

***Prognostic analysis of key genes***

Survival analysis of key genes was performed using Kaplan-Meier curves, which showed that all key genes except *KIF14* and *KIF18B* were related to prognosis. GC patients with high expression of *RAD54L*, *TPX2*, and *XRCC2* had shorter survival times, while GC patients with high expression of other key genes had longer survival times (Figure 5). However, differential analysis showed that 16 key genes were highly expressed in GC tissues. We further analyzed 16 key genes using the online database Oncomine, which showed that 16 key genes were highly expressed in GC tissues (Figure 6).

**DISCUSSION**

Metastasis, drug resistance, and recurrence are the main causes of the low survival rate of patients with GC. Self-renewal of GCSCs is considered to be the underlying cause of the metastasis, drug resistance, and recurrence of GC[10,11]. After conventional surgical treatment or adjuvant chemotherapy, the number of GCSCs does not decrease but stemness is enriched, leading to metastasis, drug resistance, and recurrence of GC[10]. Moreover, Bekaii-Saab *et al*[15] indicated that identification and targeted therapy of cancer stem cells had important significance in the treatment of GC. Therefore, targeted therapy of GCSCs is crucial. This study used WGCNA based on the mRNAsi score to identify key genes related to GCSC characteristics. This study found that the mRNAsi of gastric tumor tissue was significantly higher than that of normal gastric tissue. In addition, we used the WGCNA package to construct a gene coexpression network. Finally, we obtained a total of 16 modules; the brown module was most positively correlated with mRNAsi, and the turquoise module showed a high negative correlation with mRNAsi. Therefore, we selected the brown module as the module of most interest and identified 16 key genes that are associated with the characteristics of GCSCs: *BUB1*, *BUB1B*, *NCAPH*, *KIF14*, *RACGAP1*, *RAD54L*, *TPX2*, *KIF15*, *KIF18B*, *CENPF*, *TTK*, *KIF4A*, *SGOL2*, *PLK4*, *XRCC2,* and *C1orf112*. These genes were upregulated in GC and had a certain correlation. The most significantly enriched GO terms of these key genes were the spindle cellular component, the sister chromatid segregation biological process, and the motor activity molecular function. The significantly enriched KEGG pathways were cell cycle and homologous recombination. Except for *KIF14* and *KIF18B*, the expression of all key genes was related to the prognosis of patients with GC.

*BUB1* encodes a protein that is essential for the function of the mitotic spindle check-point[24]. Grabsch *et al*[25] showed that *BUB1* was overexpressed in up to 80% of GCs and was associated with tumor cell proliferation. Meanwhile, related studies indicated that high expression of *BUB1B* was related to the invasion, lymph node metastasis, liver metastasis, and recurrence of GC[26]. Kinesins are a family of molecular motors that play important roles in intracellular transport and cell division[27]. Previous studies demonstrated that *KIF14* can enhance tumor adhesion, invasion, and chemical resistance during tumor development[28]. Tong *et al*[29] also revealed that *KIF14* was overexpressed in GC and that this was related to tumor progression, invasion, and metastasis. Spheroid colony formation is an effective model for the characterization of cancer stem cells[11]. Oue *et al*[30] showed that *KIF15* and *KIF4A* were upregulated in GC spheroid cells. A previous study also demonstrated that patients with high expression of *KIF14* mRNA had a higher risk of malignancy, recurrence, and metastasis than those without high expression of *KIF14* mRNA[31]. Overexpression of *KIF18B* increases the proliferation of hepatocellular carcinoma cells, and downregulation of *KIF18B* *in vitro* can inhibit the migration and invasion of cervical cancer cells[32]. *KIF4A* is considered to be an oncogene, indicating its involvement in malignancy, and its expression was indeed related to the occurrence, development, and metastasis of GC[33]. Rac GTPase accelerating protein1 (*RacGAP1*), which belongs to the GTPase activation protein family, not only induces cytokinesis but also interferes with mitotic spindle assembly, thus participating in the regulation of cell proliferation[34]. Studies have indicated that *RacGAP1* is involved in cell transformation, movement, migration, and metastasis[35] and is positively correlated with the proliferation marker Ki67[36]. In addition, *RacGAP1* is highly expressed in GC and significantly associated with tumor progression and poor prognosis[37]. *CENPF* is a member of the mitochondrial family and regulates the proliferation of various tumor cells[38]. Chen *et al*[39] showed that *CENPF* was overexpressed in GC and promoted its proliferation and metastasis. *TTK* is also known as a mitotic kinase[40], which is highly expressed in many types of tumors and promotes the proliferation of cancer cells[41,42]. Park *et al*[43] found that *TTK* plays an indispensable role in the development and maintenance of tumor stem cells. *PLK4* is a major regulator of centromeric repeats, and its overexpression in somatic cells can disrupt spindle formation, while its depletion can inhibit cell proliferation[44]. Shinmura *et al*[45] revealed that *PLK4* is overexpressed in GC and plays an important role in cell proliferation, tumorigenesis, invasion, and drug resistance. *C1ORF112* is overexpressed in various tumors, including GC, and plays an important role in the growth of cancer cells, but it is still unclear whether *C1ORF112* is particularly necessary for the proliferation of cancer cells. Condensins are multiprotein complexes that play a central role in chromosome assembly and segregation during mitosis and meiosis[46]. *NCAPH* is one of the three non-SMC subunits of condensation I. Studies have indicated that non-SMC condensin I complex subunit G is overexpressed in colon cancer and hepatocellular carcinoma and promotes the proliferation and migration of tumor cells[47].

In addition, survival analysis and Oncomine analysis revealed that the prognosis of patients with GC and the expression of three genes (*RAD54L*, *TPX2*, and *XRCC2*) were consistently related. *XRCC2* plays a crucial role in DNA repair and chromosome arrangement, and its dysfunction may lead to tumor development and progression[48,49]. Wang *et al*[50] found that *XRCC2* is overexpressed in GC. *RAD54L* and *XRCC2* are homologous recombination factors that are overexpressed in lung cancer and GC, and there is a significant correlation between the expression levels of these two proteins[51]. The inactivation of these factors leads to an obvious deficiency in homologous recombination and DNA repair in all eukaryotes[52]. In addition, related studies indicated that abnormal function of homologous recombination leads to genomic instability, resulting in the occurrence of cancer[53,54]. In this study, *KEGG* enrichment analysis indicated that *RAD54L* and *XRCC2* were significantly enriched in homologous recombination. *TPX2* is a microtubule-associated protein that activates the cell cycle kinase protein Aurora-A and then plays an important role in the formation of spindles in mitosis[55]. Liang *et al*[56] showed that *TPX2* was overexpressed in GC and was related to poor progression and prognosis of GC.

**CONCLUSION**

In conclusion, this study found that the 16 key genes identified play an important role in the maintenance of GCSCs. In addition, *RAD54L*, *TPX2*, and *XRCC2* are most likely to be therapeutic targets for inhibiting the stemness characteristics of GC, providing a new idea for the treatment of GC. However, further research is needed to verify this hypothesis.

**ARTICLE HIGHLIGHTS**

***Research background***

Gastric cancer (GC) stem cells are the primary cause of GC metastasis and drug resistance.

***Research motivation***

The purpose of this study was to characterize the expression of stem cell-related genes in GC.

***Research objectives***

Targeting for inhibiting the stemness characteristics of GC cells will become a new treatment for tumors

***Research methods***

RNA sequencing results and clinical data for gastric adenoma and adenocarcinoma samples were obtained from The Cancer Genome Atlas (TCGA) database, and the results of the GC mRNA expression-based stemness index (mRNAsi) were analyzed. Weighted gene coexpression network analysis (WGCNA) was then used to find modules of interest and their key genes. Survival analysis of key genes was performed using the online tool Kaplan-Meier Plotter, and the online database Oncomine was used to assess the expression of key genes in GC.

***Research results***

mRNAsi was significantly upregulated in GC tissues compared to normal gastric tissues (*P* < 0.0001). A total of 16 modules were obtained from the gene coexpression network; the brown module was most positively correlated with mRNAsi. Sixteen key genes (*BUB1*, *BUB1B*, *NCAPH*, *KIF14*, *RACGAP1*, *RAD54L*, *TPX2*, *KIF15*, *KIF18B*, *CENPF*, *TTK*, *KIF4A*, *SGOL2*, *PLK4*, *XRCC2*, and *C1orf112*) were identified in the brown module.

***Research conclusions***

*RAD54L*, *TPX2*, and *XRCC2* are the most positively correlated with mRNAsi and are the most likely therapeutic targets for inhibiting the stemness characteristics of gastric cancer cells.

***Research perspectives***

This study aimed to identify key genes related to stemness by combining WGCNA with GC mRNAsi in TCGA, thus providing new ideas for the treatment of GC.

**REFERENCES**

1 **Bray F**, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68**: 394-424 [PMID: 30207593 DOI: 10.3322/caac.21492]

2 **Zong L**, Abe M, Seto Y, Ji J. The challenge of screening for early gastric cancer in China. *Lancet* 2016; **388**: 2606 [PMID: 27894662 DOI: 10.1016/S0140-6736(16)32226-7]

3 **Shackleton M**, Quintana E, Fearon ER, Morrison SJ. Heterogeneity in cancer: cancer stem cells versusclonal evolution. *Cell* 2009; **138**: 822-829 [PMID: 19737509 DOI: 10.1016/j.cell.2009.08.017]

4 **Smalley M**, Ashworth A. Stem cells and breast cancer: A field in transit. *Nat Rev Cancer* 2003; **3**: 832-844 [PMID: 14668814 DOI: 10.1038/nrc1212]

5 **Bjerkvig R**, Tysnes BB, Aboody KS, Najbauer J, Terzis AJ. Opinion: the origin of the cancer stem cell: current controversies and new insights. *Nat Rev Cancer* 2005; **5**: 899-904 [PMID: 16327766 DOI: 10.1038/nrc1740]

6 **Hill RP**. Identifying cancer stem cells in solid tumors: case not proven. *Cancer Res* 2006; **66**: 1891-5; discussion 1890 [PMID: 16488984 DOI: 10.1158/0008-5472.CAN-05-3450]

7 **Chen W**, Dong J, Haiech J, Kilhoffer MC, Zeniou M. Cancer Stem Cell Quiescence and Plasticity as Major Challenges in Cancer Therapy. *Stem Cells Int* 2016; **2016**: 1740936 [PMID: 27418931 DOI: 10.1155/2016/1740936]

8 **Diehn M**, Cho RW, Lobo NA, Kalisky T, Dorie MJ, Kulp AN, Qian D, Lam JS, Ailles LE, Wong M, Joshua B, Kaplan MJ, Wapnir I, Dirbas FM, Somlo G, Garberoglio C, Paz B, Shen J, Lau SK, Quake SR, Brown JM, Weissman IL, Clarke MF. Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature* 2009; **458**: 780-783 [PMID: 19194462 DOI: 10.1038/nature07733]

9 **O'Connor ML**, Xiang D, Shigdar S, Macdonald J, Li Y, Wang T, Pu C, Wang Z, Qiao L, Duan W. Cancer stem cells: A contentious hypothesis now moving forward. *Cancer Lett* 2014; **344**: 180-187 [PMID: 24333726 DOI: 10.1016/j.canlet.2013.11.012]

10 **Hayakawa Y**, Ariyama H, Stancikova J, Sakitani K, Asfaha S, Renz BW, Dubeykovskaya ZA, Shibata W, Wang H, Westphalen CB, Chen X, Takemoto Y, Kim W, Khurana SS, Tailor Y, Nagar K, Tomita H, Hara A, Sepulveda AR, Setlik W, Gershon MD, Saha S, Ding L, Shen Z, Fox JG, Friedman RA, Konieczny SF, Worthley DL, Korinek V, Wang TC. Mist1 Expressing Gastric Stem Cells Maintain the Normal and Neoplastic Gastric Epithelium and Are Supported by a Perivascular Stem Cell Niche. *Cancer Cell* 2015; **28**: 800-814 [PMID: 26585400 DOI: 10.1016/j.ccell.2015.10.003]

11 **Takaishi S**, Okumura T, Wang TC. Gastric cancer stem cells. *J Clin Oncol* 2008; **26**: 2876-2882 [PMID: 18539967 DOI: 10.1200/JCO.2007.15.2603]

12 **Takaishi S**, Okumura T, Tu S, Wang SS, Shibata W, Vigneshwaran R, Gordon SA, Shimada Y, Wang TC. Identification of gastric cancer stem cells using the cell surface marker CD44. *Stem Cells* 2009; **27**: 1006-1020 [PMID: 19415765 DOI: 10.1002/stem.30]

13 **Liu J**, Ma L, Xu J, Liu C, Zhang J, Liu J, Chen R, Zhou Y. Spheroid body-forming cells in the human gastric cancer cell line MKN-45 possess cancer stem cell properties. *Int J Oncol* 2013; **42**: 453-459 [PMID: 23229446 DOI: 10.3892/ijo.2012.1720]

14 **Reya T**, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature* 2001; **414**: 105-111 [PMID: 11689955 DOI: 10.1038/35102167]

15 **Bekaii-Saab T**, El-Rayes B. Identifying and targeting cancer stem cells in the treatment of gastric cancer. *Cancer* 2017; **123**: 1303-1312 [PMID: 28117883 DOI: 10.1002/cncr.30538]

16 **Norollahi SE**, Mansour-Ghanaei F, Joukar F, Ghadarjani S, Mojtahedi K, Gharaei Nejad K, Hemmati H, Gharibpoor F, Khaksar R, Samadani AA. Therapeutic approach of Cancer stem cells (CSCs) in gastric adenocarcinoma; DNA methyltransferases enzymes in cancer targeted therapy. *Biomed Pharmacother* 2019; **115**: 108958 [PMID: 31075731 DOI: 10.1016/j.biopha.2019.108958]

17 **Malta TM**, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O, Colaprico A, Czerwińska P, Mazurek S, Mishra L, Heyn H, Krasnitz A, Godwin AK, Lazar AJ; Cancer Genome Atlas Research Network, Stuart JM, Hoadley KA, Laird PW, Noushmehr H, Wiznerowicz M. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 2018; **173**: 338-354.e15 [PMID: 29625051 DOI: 10.1016/j.cell.2018.03.034]

18 **Langfelder P**, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**: 559 [PMID: 19114008 DOI: 10.1186/1471-2105-9-559]

19 **Zhang B**, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; **4**: Article17 [PMID: 16646834 DOI: 10.2202/1544-6115.1128]

20 **Yuan L**, Zeng G, Chen L, Wang G, Wang X, Cao X, Lu M, Liu X, Qian G, Xiao Y, Wang X. Identification of key genes and pathways in human clear cell renal cell carcinoma (ccRCC) by co-expression analysis. *Int J Biol Sci* 2018; **14**: 266-279 [PMID: 29559845 DOI: 10.7150/ijbs.23574]

21 **Zhou Z**, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, Zhao Q. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *Int J Biol Sci* 2018; **14**: 124-136 [PMID: 29483831 DOI: 10.7150/ijbs.22619]

22 **Szklarczyk D**, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015; **43**: D447-D452 [PMID: 25352553 DOI: 10.1093/nar/gku1003]

23 **Yu G**, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; **16**: 284-287 [PMID: 22455463 DOI: 10.1089/omi.2011.0118]

24 **Sharp-Baker H**, Chen RH. Spindle checkpoint protein Bub1 is required for kinetochore localization of Mad1, Mad2, Bub3, and CENP-E, independently of its kinase activity. *J Cell Biol* 2001; **153**: 1239-1250 [PMID: 11402067 DOI: 10.1083/jcb.153.6.1239]

25 **Grabsch H**, Takeno S, Parsons WJ, Pomjanski N, Boecking A, Gabbert HE, Mueller W. Overexpression of the mitotic checkpoint genes BUB1, BUBR1, and BUB3 in gastric cancer--association with tumour cell proliferation. *J Pathol* 2003; **200**: 16-22 [PMID: 12692836 DOI: 10.1002/path.1324]

26 **Ando K**, Kakeji Y, Kitao H, Iimori M, Zhao Y, Yoshida R, Oki E, Yoshinaga K, Matumoto T, Morita M, Sakaguchi Y, Maehara Y. High expression of BUBR1 is one of the factors for inducing DNA aneuploidy and progression in gastric cancer. *Cancer Sci* 2010; **101**: 639-645 [PMID: 20132214 DOI: 10.1111/j.1349-7006.2009.01457.x]

27 **Hirokawa N**. Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* 1998; **279**: 519-526 [PMID: 9438838 DOI: 10.1126/science.279.5350.519]

28 **Ahmed SM**, Thériault BL, Uppalapati M, Chiu CW, Gallie BL, Sidhu SS, Angers S. KIF14 negatively regulates Rap1a-Radil signaling during breast cancer progression. *J Cell Biol* 2012; **199**: 951-967 [PMID: 23209302 DOI: 10.1083/jcb.201206051]

29 **Tong H**, Wang J, Chen H, Wang Z, Fan H, Ni Z. Transcriptomic analysis of gene expression profiles of stomach carcinoma reveal abnormal expression of mitotic components. *Life Sci* 2017; **170**: 41-49 [PMID: 27923621 DOI: 10.1016/j.lfs.2016.12.001]

30 **Oue N**, Mukai S, Imai T, Pham TT, Oshima T, Sentani K, Sakamoto N, Yoshida K, Yasui W. Induction of KIFC1 expression in gastric cancer spheroids. *Oncol Rep* 2016; **36**: 349-355 [PMID: 27176706 DOI: 10.3892/or.2016.4781]

31 **Nakamura Y**, Tanaka F, Haraguchi N, Mimori K, Matsumoto T, Inoue H, Yanaga K, Mori M. Clinicopathological and biological significance of mitotic centromere-associated kinesin overexpression in human gastric cancer. *Br J Cancer* 2007; **97**: 543-549 [PMID: 17653072 DOI: 10.1038/sj.bjc.6603905]

32 **Wu Y**, Wang A, Zhu B, Huang J, Lu E, Xu H, Xia W, Dong G, Jiang F, Xu L. KIF18B promotes tumor progression through activating the Wnt/β-catenin pathway in cervical cancer. *Onco Targets Ther* 2018; **11**: 1707-1720 [PMID: 29636620 DOI: 10.2147/OTT.S157440]

33 **Taniwaki M**, Takano A, Ishikawa N, Yasui W, Inai K, Nishimura H, Tsuchiya E, Kohno N, Nakamura Y, Daigo Y. Activation of KIF4A as a prognostic biomarker and therapeutic target for lung cancer. *Clin Cancer Res* 2007; **13**: 6624-6631 [PMID: 18006763 DOI: 10.1158/1078-0432.CCR-07-1328]

34 **Zhao WM**, Fang G. MgcRacGAP controls the assembly of the contractile ring and the initiation of cytokinesis. *Proc Natl Acad Sci U S A* 2005; **102**: 13158-13163 [PMID: 16129829 DOI: 10.1073/pnas.0504145102]

35 **Sanz-Moreno V**, Gadea G, Ahn J, Paterson H, Marra P, Pinner S, Sahai E, Marshall CJ. Rac activation and inactivation control plasticity of tumor cell movement. *Cell* 2008; **135**: 510-523 [PMID: 18984162 DOI: 10.1016/j.cell.2008.09.043]

36 **Ke HL**, Ke RH, Li ST, Li B, Lu HT, Wang XQ. Expression of RACGAP1 in high grade meningiomas: a potential role in cancer progression. *J Neurooncol* 2013; **113**: 327-332 [PMID: 23525949 DOI: 10.1007/s11060-013-1121-7]

37 **Saigusa S**, Tanaka K, Mohri Y, Ohi M, Shimura T, Kitajima T, Kondo S, Okugawa Y, Toiyama Y, Inoue Y, Kusunoki M. Clinical significance of RacGAP1 expression at the invasive front of gastric cancer. *Gastric Cancer* 2015; **18**: 84-92 [PMID: 24615626 DOI: 10.1007/s10120-014-0355-1]

38 **Aytes A**, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, Eastham JA, Gopalan A, Pienta KJ, Shen MM, Califano A, Abate-Shen C. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 2014; **25**: 638-651 [PMID: 24823640 DOI: 10.1016/j.ccr.2014.03.017]

39 **Chen EB**, Qin X, Peng K, Li Q, Tang C, Wei YC, Yu S, Gan L, Liu TS. HnRNPR-CCNB1/CENPF axis contributes to gastric cancer proliferation and metastasis. *Aging (Albany NY)* 2019; **11**: 7473-7491 [PMID: 31527303 DOI: 10.18632/aging.102254]

40 **Dominguez-Brauer C**, Thu KL, Mason JM, Blaser H, Bray MR, Mak TW. Targeting Mitosis in Cancer: Emerging Strategies. *Mol Cell* 2015; **60**: 524-536 [PMID: 26590712 DOI: 10.1016/j.molcel.2015.11.006]

41 **Jemaà M**, Galluzzi L, Kepp O, Senovilla L, Brands M, Boemer U, Koppitz M, Lienau P, Prechtl S, Schulze V, Siemeister G, Wengner AM, Mumberg D, Ziegelbauer K, Abrieu A, Castedo M, Vitale I, Kroemer G. Characterization of novel MPS1 inhibitors with preclinical anticancer activity. *Cell Death Differ* 2013; **20**: 1532-1545 [PMID: 23933817 DOI: 10.1038/cdd.2013.105]

42 **Maia AR**, de Man J, Boon U, Janssen A, Song JY, Omerzu M, Sterrenburg JG, Prinsen MB, Willemsen-Seegers N, de Roos JA, van Doornmalen AM, Uitdehaag JC, Kops GJ, Jonkers J, Buijsman RC, Zaman GJ, Medema RH. Inhibition of the spindle assembly checkpoint kinase TTK enhances the efficacy of docetaxel in a triple-negative breast cancer model. *Ann Oncol* 2015; **26**: 2180-2192 [PMID: 26153498 DOI: 10.1093/annonc/mdv293]

43 **Park JH**, Chung S, Matsuo Y, Nakamura Y. Development of small molecular compounds targeting cancer stem cells. *Medchemcomm* 2017; **8**: 73-80 [PMID: 30108692 DOI: 10.1039/c6md00385k]

44 **Coelho PA**, Bury L, Sharif B, Riparbelli MG, Fu J, Callaini G, Glover DM, Zernicka-Goetz M. Spindle formation in the mouse embryo requires Plk4 in the absence of centrioles. *Dev Cell* 2013; **27**: 586-597 [PMID: 24268700 DOI: 10.1016/j.devcel.2013.09.029]

45 **Shinmura K**, Kurabe N, Goto M, Yamada H, Natsume H, Konno H, Sugimura H. PLK4 overexpression and its effect on centrosome regulation and chromosome stability in human gastric cancer. *Mol Biol Rep* 2014; **41**: 6635-6644 [PMID: 24981932 DOI: 10.1007/s11033-014-3546-2]

46 **Wood AJ**, Severson AF, Meyer BJ. Condensin and cohesin complexity: the expanding repertoire of functions. *Nat Rev Genet* 2010; **11**: 391-404 [PMID: 20442714 DOI: 10.1038/nrg2794]

47 **Zhang Q**, Su R, Shan C, Gao C, Wu P. Non-SMC Condensin I Complex, Subunit G (NCAPG) is a Novel Mitotic Gene Required for Hepatocellular Cancer Cell Proliferation and Migration. *Oncol Res* 2018; **26**: 269-276 [PMID: 29046167 DOI: 10.3727/096504017X15075967560980]

48 **Curtin K**, Lin WY, George R, Katory M, Shorto J, Cannon-Albright LA, Smith G, Bishop DT, Cox A, Camp NJ; Colorectal Cancer Study Group. Genetic variants in XRCC2: new insights into colorectal cancer tumorigenesis. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 2476-2484 [PMID: 19690184 DOI: 10.1158/1055-9965.EPI-09-0187]

49 **Griffin CS**, Simpson PJ, Wilson CR, Thacker J. Mammalian recombination-repair genes XRCC2 and XRCC3 promote correct chromosome segregation. *Nat Cell Biol* 2000; **2**: 757-761 [PMID: 11025669 DOI: 10.1038/35036399]

50 **Wang N**, Dong XJ, Zhou RM, Guo W, Zhang XJ, Li Y. An investigation on the polymorphisms of two DNA repair genes and susceptibility to ESCC and GCA of high-incidence region in northern China. *Mol Biol Rep* 2009; **36**: 357-364 [PMID: 18046624 DOI: 10.1007/s11033-007-9187-y]

51 **Oji Y**, Tatsumi N, Kobayashi J, Fukuda M, Ueda T, Nakano E, Saito C, Shibata S, Sumikawa M, Fukushima H, Saito A, Hojo N, Suzuki M, Hoshikawa T, Shimura T, Morii E, Oka Y, Hosen N, Komatsu K, Sugiyama H. Wilms' tumor gene WT1 promotes homologous recombination-mediated DNA damage repair. *Mol Carcinog* 2015; **54**: 1758-1771 [PMID: 25418835 DOI: 10.1002/mc.22248]

52 **Essers J**, Hendriks RW, Swagemakers SM, Troelstra C, de Wit J, Bootsma D, Hoeijmakers JH, Kanaar R. Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous recombination. *Cell* 1997; **89**: 195-204 [PMID: 9108475 DOI: 10.1016/s0092-8674(00)80199-3]

53 **Jasin M**. Homologous repair of DNA damage and tumorigenesis: the BRCA connection. *Oncogene* 2002; **21**: 8981-8993 [PMID: 12483514 DOI: 10.1038/sj.onc.1206176]

54 **Venkitaraman AR**. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* 2002; **108**: 171-182 [PMID: 11832208 DOI: 10.1016/s0092-8674(02)00615-3]

55 **Ozlü N**, Srayko M, Kinoshita K, Habermann B, O'toole ET, Müller-Reichert T, Schmalz N, Desai A, Hyman AA. An essential function of the C. elegans ortholog of TPX2 is to localize activated aurora A kinase to mitotic spindles. *Dev Cell* 2005; **9**: 237-248 [PMID: 16054030 DOI: 10.1016/j.devcel.2005.07.002]

56 **Liang B**, Zheng W, Fang L, Wu L, Zhou F, Yin X, Yu X, Zou Z. Overexpressed targeting protein for Xklp2 (TPX2) serves as a promising prognostic marker and therapeutic target for gastric cancer. *Cancer Biol Ther* 2016; **17**: 824-832 [PMID: 27314162 DOI: 10.1080/15384047.2016.1195046]

**Footnotes**

**Institutional review board statement:** No human and/or animal subjects were involved in this study.

**Conflict-of-interest statement:** The authors declare no competing financial interests.

**Data sharing statement:** RNA sequencing results of 373 tissues and 348 human gastric adenomas and adenocarcinoma samples were obtained from TCGA database (https://portal.gdc.cancer.gov). The RNA-seq results of 30 normal samples and 343 cancer samples were merged into a matrix file using a script in the Perl language (http://www.perl.org/). We then converted the Ensembl ID in the matrix file to the gene name using the Ensembl database (http://www.ensembl.org/index.html) and the Perl language script. In addition, 406 pieces of clinical data were downloaded, and the relevant clinical data were collated and extracted using scripts in the Perl language. The calculation of mRNAsi was performed from the molecular spectrum of normal cells with different degrees of stemness.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/Licenses/by-nc/4.0/

**Manuscript source:** Unsolicited manuscript

**Peer-review started:** July 28, 2020

**First decision:** August 9, 2020

**Article in press:** October 12, 2020

**Specialty type:** Gastroenterology and hepatology

**Country/Territory of origin:** China

**Peer-review report’s scientific quality classification**

Grade A (Excellent): 0

Grade B (Very good): B

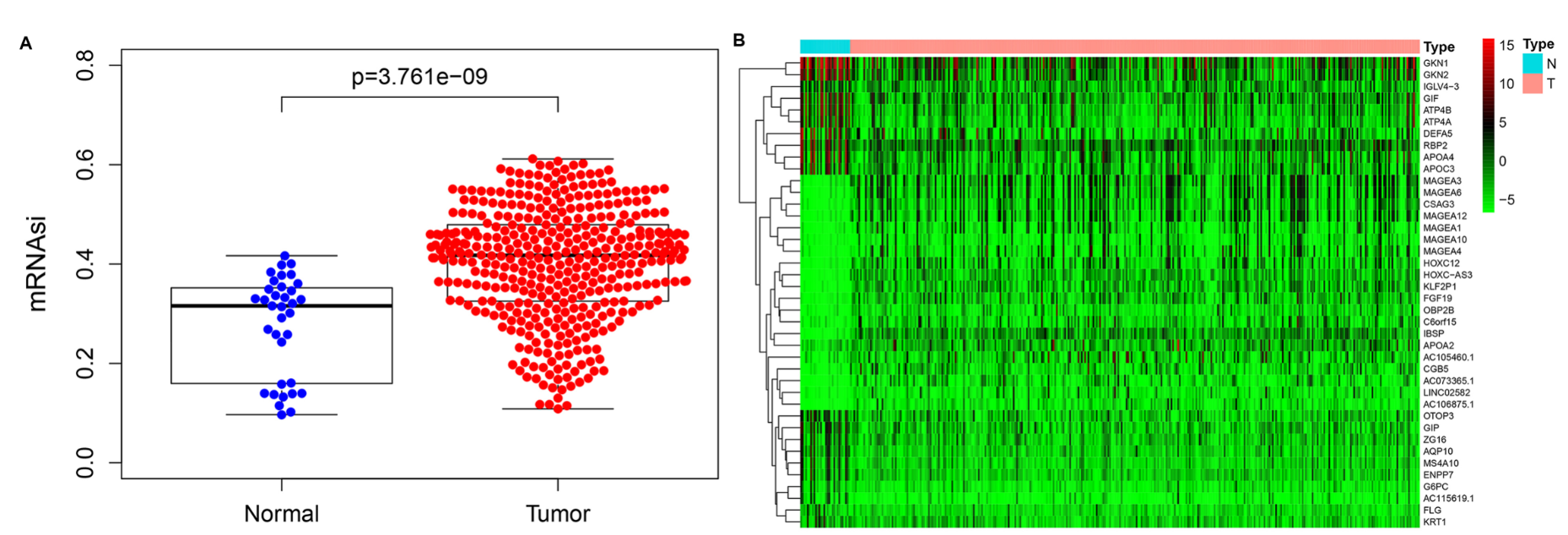
Grade C (Good): C

Grade D (Fair): 0

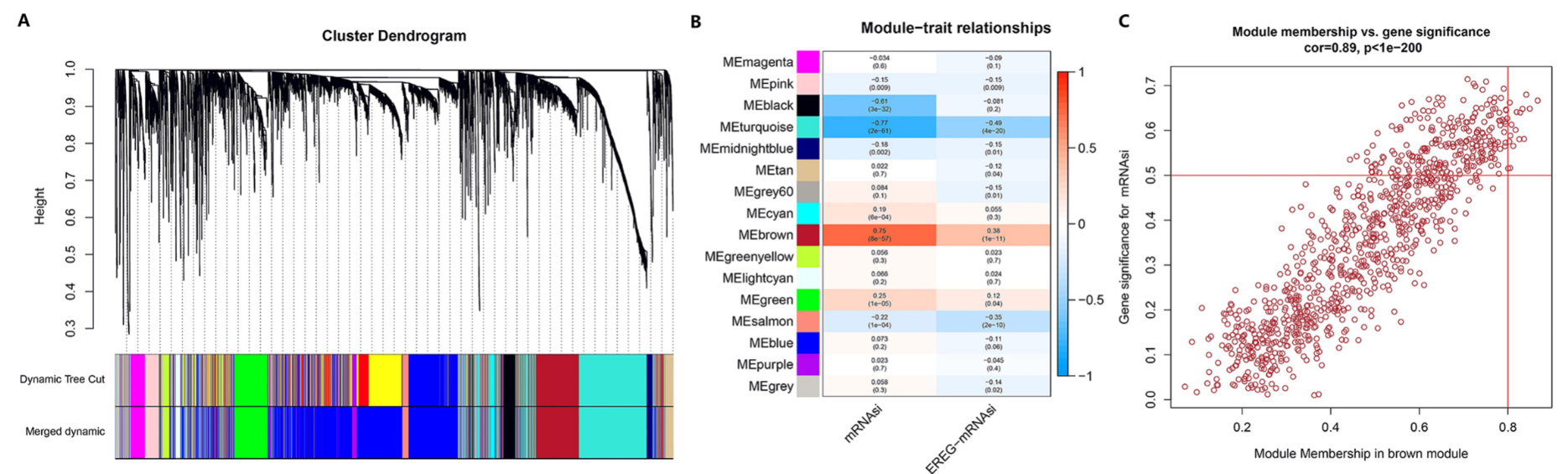
Grade E (Poor): 0

**P-Reviewer:** Tabarzad M **S-Editor:** Wang JL **L-Editor:** Wang TQ **P-Editor:** Li JH

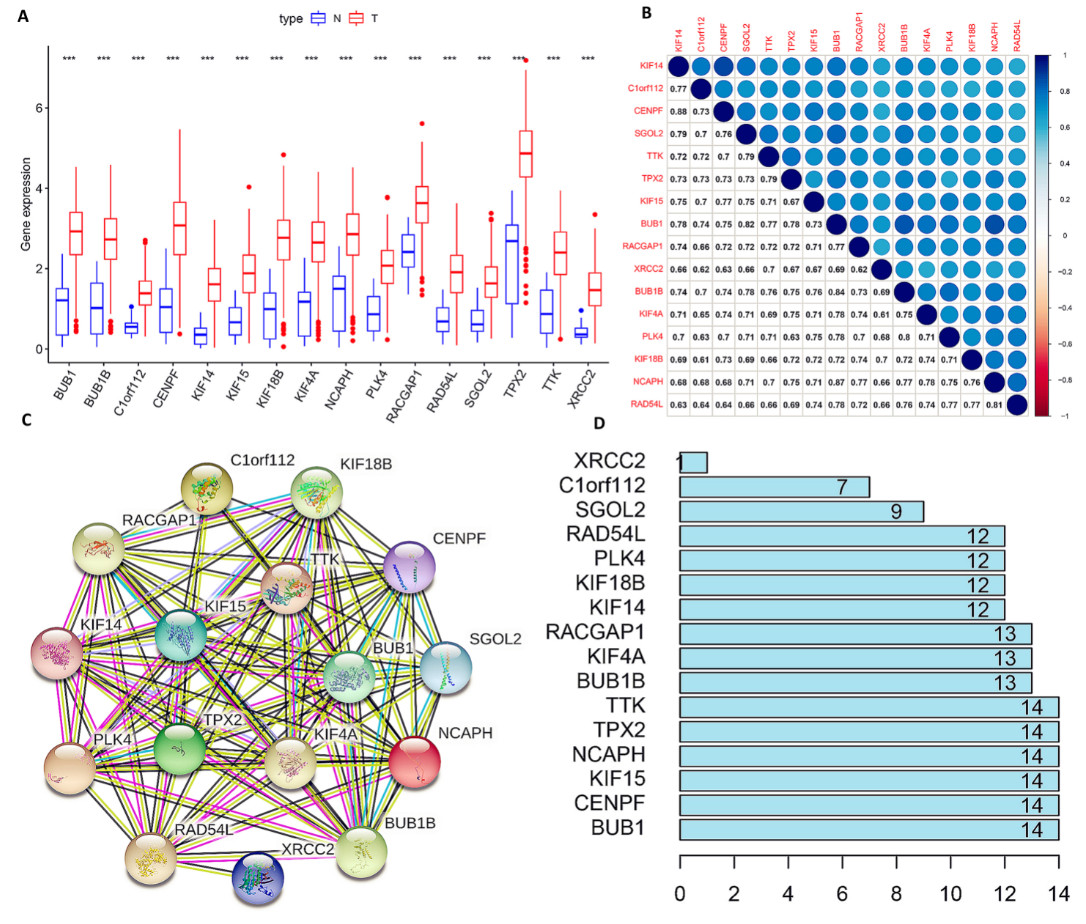
**Figure Legends**



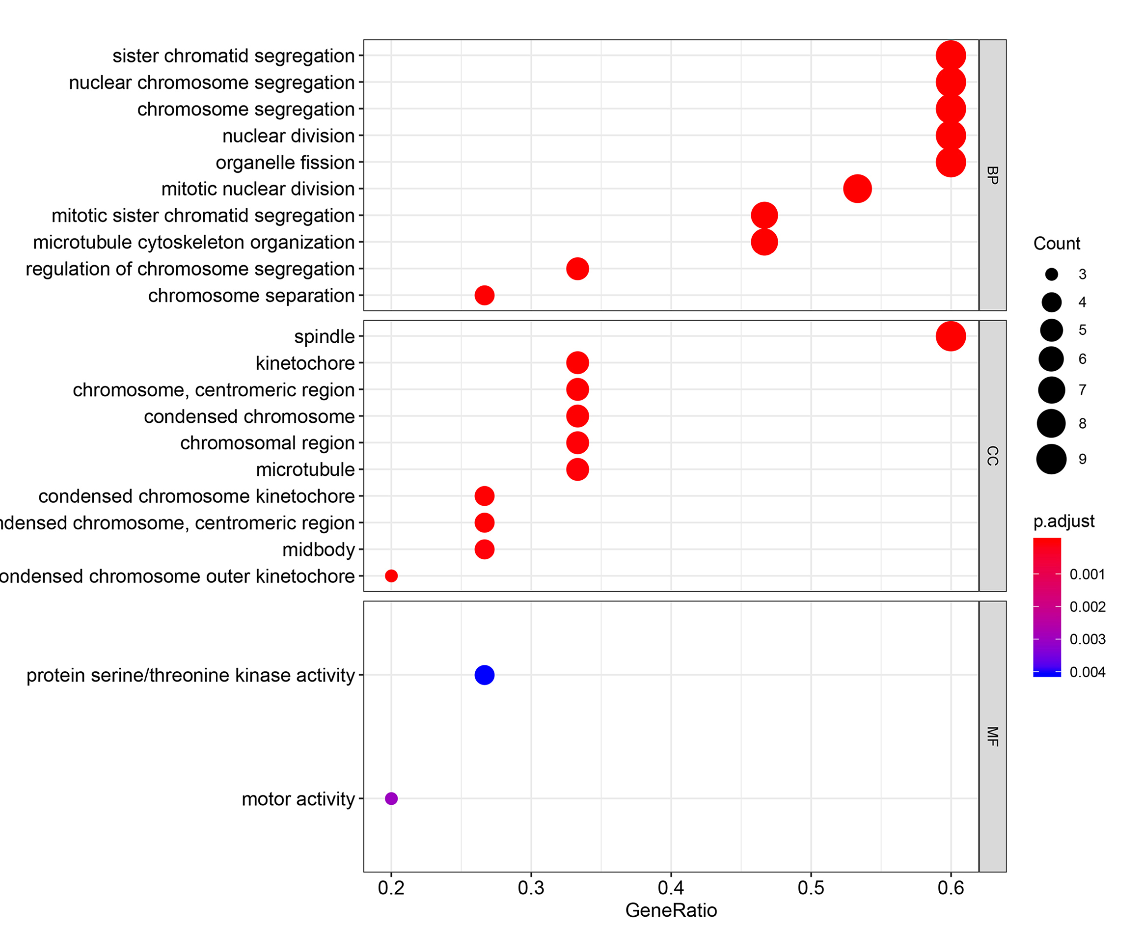
**Figure 1 Differential analysis of mRNA expression-based stemness index and screening of differentially expressed genes.** A: Expression of mRNA expression-based stemness index in gastric tumor tissues and normal gastric tissues; B:Significantly differentially expressed genes in gastric tumor tissues and normal gastric tissues.



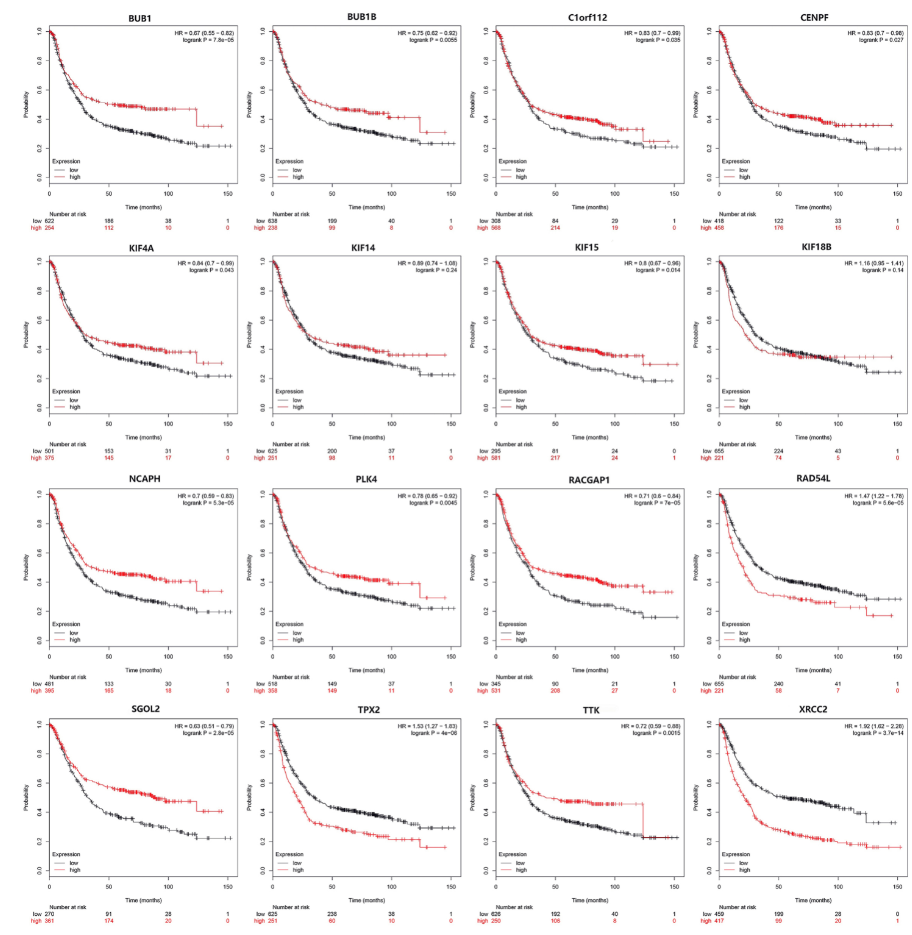
**Figure 2 Weighted gene coexpression network analysis.** A: Clustering dendrograms of differentially expressed genes. Each piece of the leaves on the cluster dendrogram corresponds to a gene. The first color band indicates the modules detected by dynamic tree cut methods. The second color band indicates the modules after merging similar modules; B: Correlation between the gene module and mRNA expression-based stemness index (mRNAsi). The correlation coefficient in each cell represents the correlation between the gene module and mRNAsi, which increases in size from blue to red. The corresponding *P* value is also annotated; C: Scatter plot of module eigengenes in brown. The top right corner is where the key genes are located.



**Figure 3 Analysis of key genes.** A: Expression of the key genes in gastric tumor tissues and normal gastric tissues. The red histogram indicates gastric tumor tissue. The blue histogram indicates normal gastric tissue. c*P* < 0.001; B:Correlation of key genes. The number in the box represents the size of the correlation, and the larger the number, the higher the correlation; C: Protein-protein interaction networks of key genes. The color of the lines represents the source of the evidence; D:Connection nodes of key genes. The number represents the number of connection nodes.



**Figure 4 Significantly enriched gene ontology biological process terms.** The larger the circle, the more number of enrichment, and the redder the color, the more significant the difference.BP: Biological process; CC: Cell component; MF: Molecular function.



**Figure 5 Survival curves of key genes using the online tool Kaplan-Meier Plotter.** The red indicates high expression and black indicates low expression. A log-rank *P* < 0.05 was considered statistically significant.



**Figure 6 Expression of key genes in gastric cancer based on the Oncomine online database.** The key genes were highly expressed in gastric cancer tissues.

**Table 1 Significantly enriched Kyoto Encyclopedia of Genes and Genomes pathways**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | **Term** | **Description** | **p.adjust** | **Gene** | **Gene counts** |
| hsa04110 | KEGG | Cell cycle | 0.0004 | *BUB1*/*BUB1B*/*TTK* | 3 |
| hsa03440 | KEGG | Homologous recombination | 0.001 | *RAD54L*/*XRCC2* | 2 |

KEGG: Kyoto Encyclopedia of Genes and Genomes.