

Targeted regulation of miR-17-5p on TMOD1 promotes the development of cardia cancer

M.-Y. WANG¹, P. CHEN², C.-P. WANG³

¹Department of Gastroenterology, Affiliated Hospital of Weifang Medical University, Weifang, P.R. China

²Department of Traditional Chinese Medicine, Affiliated Hospital of Weifang Medical University, Weifang, P.R. China

³Department of Neurology, Affiliated Hospital of Weifang Medical University, Weifang, P.R. China

Abstract. – **OBJECTIVE:** Cardia cancer is a common type of gastric cancer. Most clinical prevention and prognosis focus on surgical resection, but the efficacy is not satisfactory. Studying the molecular mechanism of pathogenesis of cardia cancer helps us intervene in prognosis and treatment.

MATERIALS AND METHODS: a total of 134 normal cases related to cardia cancer and 62 cases of cardia cancer samples from the Gene Expression Omnibus (GEO) database were collected. A series of bioinformatics analyses, including differential gene analysis, co-expression analysis, enrichment analysis, regulator prediction, and (Protein-protein interaction) PPI analysis validation were performed.

RESULTS: Differential analysis highlighted 10882 differential genes ($p < 0.05$). Weighted gene co-expression network analysis indicated 6 functional disorder modules. TMOD1, JAM2, SPARC, ST18, NOS1 were key genes of each module. Enrichment analysis showed the dysfunctional module genes were mainly related to the proteinaceous extracellular matrix and neuroactive ligand-receptor interaction. Pivotal analysis of ncRNA demonstrated miR-17-5p significantly regulates modular genes including m1, m3, and m5. Target genes were backtracked according to the key regulators. Then, the Module_target gene_ncRNA interaction network diagram was constructed. The network shows m1 has the strongest regulation effect in the network. PPI showed that the core gene TMOD1 (Tropomodulin1) of m1 was at TOP10 in the algorithm. In other words, PPI indicated the importance of TMOD1 in the interaction network.

CONCLUSIONS: We believe that targeted regulation of miR-17-5p on TMOD1 gene affects the neuroactive ligand-receptor interaction pathway, and it promotes proliferation and apoptosis of cardia cancer cells.

Key Words:

Cardia cancer, Co-expression analysis, WGCNA, Enrichment analysis, PPI, ncRNA.

Abbreviations

EUS = endoscopic ultrasonography; PASG = phago-quosonographies; GEO = Gene Expression Omnibus; miRNAs = microRNAs.

Introduction

Gastric cancer consists of cardia cancer and non-cardia cancer. The diagnosis of gastric cardia tumour is difficult due to its poor clinicopathological features. The survival rate of cardia cancer in the last 5 years is 79.7%¹. Cardia cancer is becoming common in Western countries with increased incidence in recent years. Japanese national records show that the incidence of cardia cancer accounted for 12.2% of all registered cases in 1963, and 17.3% in 1990². Although the non-cardia tumour is often diagnosed early, cardia cancer is often found in advanced stage. Pathologically, cardia cancer is invasive. It is associated with more lymphatic invasion and lymph node metastasis³. Currently, the best surgical procedure for cardia cancer remains controversial. It is because surgery exhibit in advanced cardia cancer with unknown factors. What's worse, invisible cancer metastasis tends to appear after surgery⁴. According to clinical data, the possibility of recurrence of cardia cancer is not very high. Therefore, the prognosis of cancer is relatively difficult to control⁵. Current surgery can not cure cardia cancer; therefore, preoperative analysis is crucial. The resectability should be evaluated according to the influencing factors combined with imaging analysis⁶. To define the optimal degree of resection of cardia cancer, transabdominal total gastrectomy, and right thoracic esophagectomy achieved diffu-

sion of free margins and controlled lymph node metastasis. Transabdominal total gastrectomy and right thoracotomy of the esophagectomy at the azygous venous level radically resect the tumour, especially in poorly differentiated tumours. Therefore, surgical resection alone cannot cure cardia cancer^{7,8}. In the early imaging analysis, ultrasound endoscopic ultrasonography (EUS) was generally used to further evaluate cardia cancer because it can accurately locate the tumour. However, EUS still has certain limitations. Thus, we should carefully use it for imaging analysis⁹. At present, there are new ways to check for cardia cancer. For example, Phagaquosonographies (PASG) analysis is performed on healthy subjects and patients with cardia cancer. This method can be used for screening, diagnosis, and differential diagnosis and prognosis¹⁰. Therapies bring novel advancement. Laser photocoagulation and intubation are effective measures to alleviate unresectable cardia cancer. Laser treatment is especially useful for short stenosis and non-circular and proximal lesions. Intubation usually relieves swallowing difficulties quickly¹¹. Despite no effective way to predict the pathogenesis of cardia cancer, with the advancement of biotechnology, more and more research methods are available. Due to the development of high-throughput sequencing, we have a better understanding of the molecular mechanisms of disease. This study aims to explore the molecular mechanism of cardia cancer by analysing the data from the GEO database through bioinformatics. Scientific predictions discover related regulators. TF (Transcription Factor) and ncRNA are generally believed to be the key regulators. TF affects transcription process, and ncRNA affects the translation process. In the follow-up clinical treatment study, these two regulators can be used as reference for our research. It also provides reference for the regulation of occurrence and development of cardia cancer.

Materials and Methods

Data Resource

Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) is an international public data platform for researchers to submit data like high-throughput sequencing. It records experimental data from the world. The data for cardia cancer in this study were derived from the GEO database. We selected the microarray data for

the mRNA of the GSE29272 database, and the platform of the probe was GPL96. The resource supports archiving raw data, processed data, and indexes. People can search for source data from it, and all data can be downloaded for free¹². We downloaded chip data from 62 patients with cardia cancer and 134 normal samples at GEO for gene differential analysis and co-expression analysis, regulator prediction, etc^{13,14}.

Variance Analysis

In this study, the R language was used for data processing and analysis. We first standardized pre-processing the mRNA microarray data of the GEO samples collected above. After standardization, the data were used for difference analysis. In this study, the limma package of R language was mainly used to construct the expression data of differential genes, including preliminary screening and correction of data¹⁵⁻¹⁷. First, the background Correct function was used to correct and standardize the data with the setting threshold $p < 0.05$. Then, the quantile normalization in the normalize Between Arrays function was used to filter out the control probe and the low-expressed probe in the probe data. Finally, the eBayes and lmFit functions were applied to identify the differentially expressed genes in the dataset with the default parameters.

Co-Expression Analysis

Co-expression analysis mainly uses WGCNA in this study. Due to numerous differential genes obtained, it was not convenient to find the key genes therein, thus we needed to cluster the similar parts of the table to form a module. This method of mining useful information from high-throughput data was referred as weighted co-expression network analysis. WGCNA tool can detect clusters of highly related genes^{18,19}. To explore the molecular mechanisms of gene expression during progression of cardia cancer, the expression profiles for differentially expressed genes in cardia cancer was established. Similarly, WGCNA was applied to explore the synergistic expression of these differential genes. First, before clustering the results of correlation coefficients, the links between genes in the network conform to scaled-free networks with more biological meaning. We used the correlation coefficient weighting method to take the N-th power of the correlation coefficient between genes. We then got the correlation coefficient between any two genes (Pearson Coefficient). After obtaining

the weighted correlation coefficients, we clustered them and constructed the clustering tree through the correlation coefficients between different genes. Various branches of the clustering tree stand for distinct functional obstacle modules. Colours symbolize modules. Based on the ability of the gene to regulate in each dysfunction module, we unearthed the key genes that cause the dysfunction of the functional modules. The key genes of the module imbalance were called the core gene. Imbalance of the expression will eventually lead to the disorder of the module, causing disease.

Enrichment Analysis

After obtaining the functional disorder module, understanding the functions and signal pathways involved in the key genes of the module remains essential. This is for better comprehension of the molecular mechanism of development in cardia cancer. The function and pathway enrichment analysis of the genes of the dysfunctional module can effectively explore the potential mechanism of gene expression in the pathogenesis of cardia cancer. In this study, 6 functional barrier modules were analysed. The clustering package of Bioconductor²⁰ in R language was used for the enrichment analysis of functions and pathways. Besides, the ClusterProfiler package can perform statistical analysis and visualization of functional clustering of genes or gene clusters. We enriched the function and path of each module using GlueGO of Cytoscape, built corresponding function, accessed network, determined the proportion of module participating corresponding functions and paths, and then found out the relevant channels.

Regulator Analysis

Gene and post-transcriptional regulation are often regulated by non-coding RNA (ncRNA). Therefore, we believe ncRNA is a regulator of the entire molecular mechanism. We need to analyse the ncRNA to better understand the molecular mechanism of cardia cancer. We scientifically predict and test the role of dysfunction modules of cardia cancer through bioinformatics, thereby improving the accuracy and efficiency of the experiment. During ncRNA pivot analysis, we screened for ncRNA that regulates the core genes to make module ncRNA network interaction maps. We entered the interactive network into Cytoscape and then filter the connectivity, preserving the node with degree >1.

Establishment of Module Target Gene_ncRNA Interaction Network Diagram

With above-obtained ncRNA, its regulated target genes were backtracked. The obtained data list into Cytoscape were inserted, and nodes with low connectivity in the network were deleted. Finally, the interaction network diagram was obtained, and the size of the node was related to the connectivity. The greater the connectivity, the larger the node.

12 Algorithms to Verify Hubgene

This study validated the core genes in the clustering module obtained by WGCNA, and analysed the network using twelve different algorithms. Then, it checked whether there was a coincidence. We used the R language to analyse the differential genes, get the PPI network to display the regulatory relationship of genes. With screening score greater than 900 points, the data were input into Cytoscape, and cytoHubba plugin was applied to analyse the genes of the PPI network. TOP10 was collected, and the outcomes of 12 algorithms were output. Finally, we compared the TOP10 of the twelve algorithms with the core genes of the module to find the coincident core genes.

Results

Identification of Expression Disorder Molecules in Cardia Cancer

This study compared the mRNA microarray data of patients with cardia cancer and normal samples. It aimed to explore the regulation of non-coding RNA on the development of cardia cancer. A differential analysis of the two sets of mRNA chips highlighted 10882 differential genes ($p < 0.05$) (**Supplementary Table I**). Therefore, we recognize these 10,882 differential genes as expression-deficient molecules in cardia cancer. Concerning these molecules, we need to cluster differential genes and search for core genes.

Identification of Relevant Functional Disorder Modules in Cardia Cancer

The WGCNA tool can detect clusters of highly related genes. Therefore, it can characterize the underlying pathogenesis of the biological disease. The weighted analysed genes are clustered to form a module. Each module characterizes a potential mechanism of action. Each module contains a core gene, and a core gene disorder can cause an abnormality in a global function,

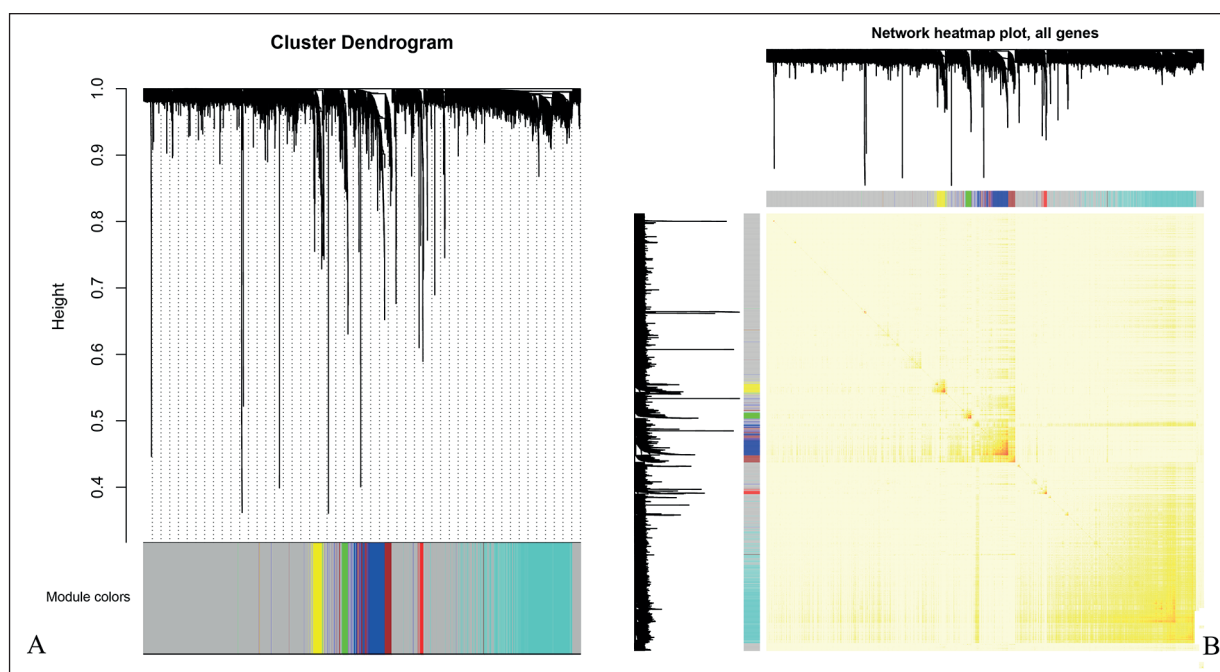


Figure 1. Synergistic expression of cardia cancer. *A*, 6 co-expression groups via clustering were identified as modules, and 6 colours represent 6 co-expression modules. *B*, Heat maps of all gene's expression in the sample, whose expression behaviour is clustered into 6 co-expression modules.

thus leading to disease. To study the functional disorder module associated with progression in cardia cancer, an expression profile matrix needs to be constructed in the sample for the 10882 differential genes obtained above. Based on the WGCNA, we observed these genes showed significant grouping expression in the sample. Besides, genes expressing similar behaviour formed a module for cluster. In this study, we identified six modules with agminated gene expression behaviour as functional disorder modules (Figure 1A, 1B). This helped us observe the complex collaborative relationships between these genes from the perspective of expression behaviour. The key genes of each module were identified on the basis of the functional disorder module. They were JAM2, ST18, TNXB, and DCTN2 and RBM23 (Table I).

Functions and Pathways Involved in the Gene of Interest

Function and pathway are important mediators of disease physiological response. Therefore, the functional and pathway enrichment analysis on the six functional barrier modules was carried out. A total of 3743 biological processes were obtained including 402 cells, 675 molecular func-

tions, and 133 KEGG pathways (**Supplementary Table II**) with a broad threshold ($p < 0.05$). We performed a functional interaction network analysis on functional disorder module (Figure 2A) using BinGO. We selected some functions to make a bubble map (Figure 2B) and found that the functional disorder module gene is mainly related to the proteinaceous extracellular matrix function. ClueGO was used to perform an interactive network analysis of the relevant pathways (Figure 2C), and some of the pathways for bubble mapping were selected (Figure 2D). The results showed that the functional barrier module genes were mainly involved in the neuroactive ligand-receptor interaction pathway. The count value is relatively large, and the correlation is

Table I. Hub gene of modules.

Colour	HubGenes	Module
Blue	TMOD1	m1
Brown	JAM2	m3
Green	SPARC	m6
Red	ST18	m4
Turquoise	NOS1	m2
Yellow	CD48	m5

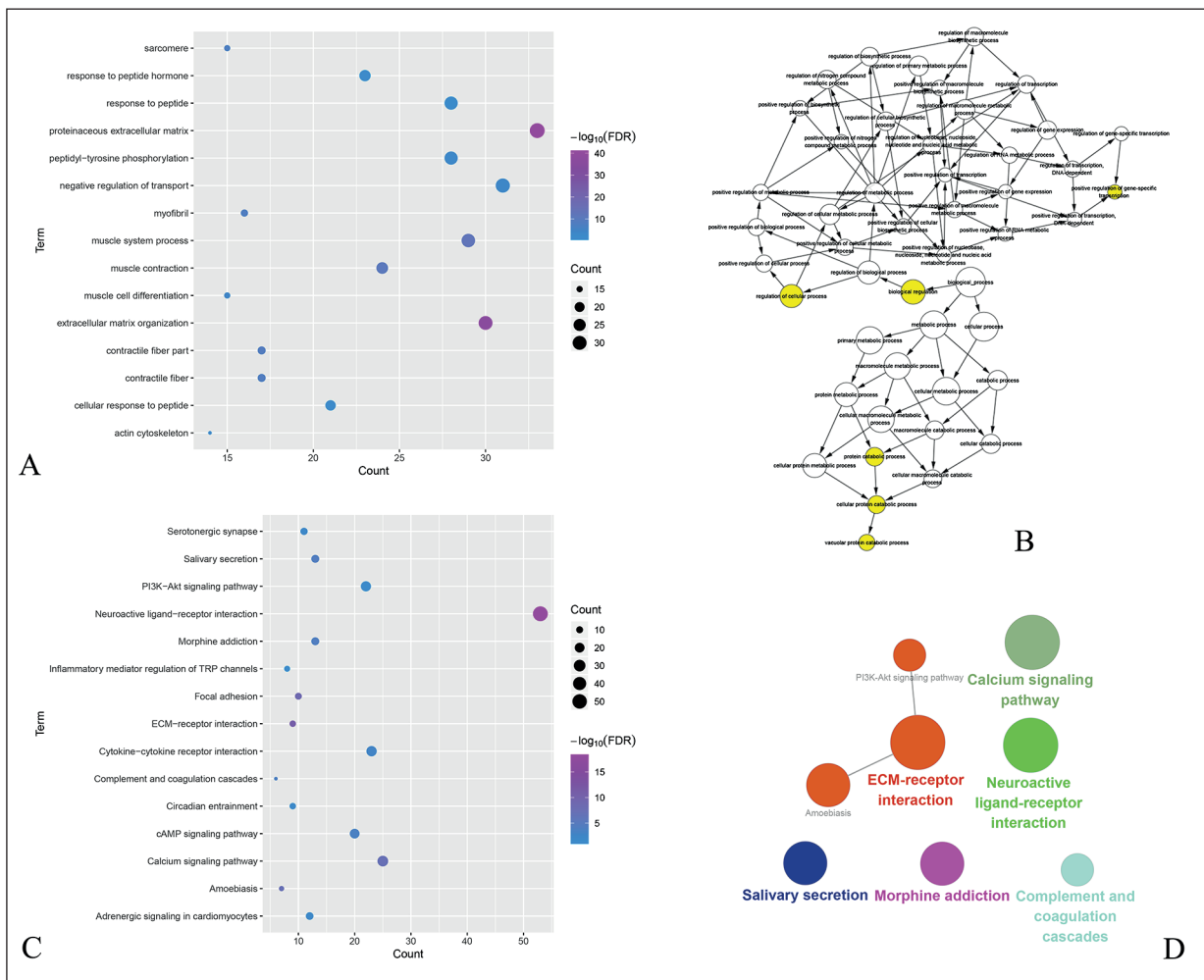


Figure 2. Functional and pathway enrichment analysis excerpts of the module gene. **A**, Functional enrichment analysis excerpt of module gene GO. From blue to purple, the enrichment increases significantly. The larger the circle, the greater the proportion of the module gene in GO function entry gene. **B**, Analysis of the functional enrichment of the module gene GO. Yellow represents the significance of the pathway involved in the module gene reaching 0.05. **C**, Module gene KEGG pathway enrichment analysis excerpt. From blue to purple, the enrichment increases significantly. The larger the circle, the greater the proportion of the module genes account for the entry gene of KEGG pathway. **D**, Analysis of modular gene KEGG pathway enrichment, different colours represent various pathways. Size of the circle refers to the count value (p -value<0.05).

relatively strong. We believe that the development of cardia cancer is related to the neuroactive ligand-receptor interaction pathway.

ncRNA and its Interaction Network Map for the Development of Cardia Cancer

Through the above analysis, we got functional pathways and key genes related to a progression in cardia cancer. However, concerning systems biology and systems genetics, post-transcriptional regulation of genes has been considered as a key regulator of disease. NcRNA is a common post-transcriptional regulator. Therefore, a pivotal analysis based on the targeted regulation of

ncRNA on the core genes of the module was conducted (**Supplementary Table III**, Figure 3). The result shows miR-17-5p significantly regulates the m1, m3, and m5. The core genes of each module were TMOD1, JAM2, and CD48, respectively. Therefore, we believe that miR-17-5p is a key regulator in this network.

Establishment of Module Target Gene_ncRNA Interaction Network Diagram

With above-obtained ncRNA, its regulated target genes were backtracked. We input the obtained data list into Cytoscape and then deleted the nodes with low connectivity in the network.

and pathways of the module have been identified and have failed to help the prognosis and treatment of cardia cancer. We need to analyse post-transcriptional regulators, while ncRNA is the most common regulator. Research based on ncRNAs provides a reference for clinical treatment. micro-RNAs (miRNAs) are post-transcriptional gene regulators that are involved in a wide range of biological processes like tumorigenesis. Disorders in the miRNA pathway are associated with cancer, but little is known about the contribution of genetic variability to the disease. Current data from studies have documented many ncRNAs that regulate cardia cancer. Data suggest dysregulation of lncRNAs makes difference to lymph node metastasis in cardia cancer²⁴. Gene ontology analysis indicated the up-regulated transcript is most abundant in SRP-dependent co-translated proteins, targeting membranes, structural components of cytoplasmic ribosomes and ribosomes. Down-regulated transcripts are highly enriched in carboxylic acid transport, focal adhesions, and cations. Therefore, its regulation can also affect the whole disease process. MEG3, miR-770, miR-135b-5p, miR-203a, miR-203b, MIR196A2 and C5orf66-AS1 may be used as potential biomarkers for predicting cardia cancer patients after prognosis²⁵⁻³⁰.

Through ipRNA pivot analysis, this study found miR-17-5p significantly regulates m1, m3, m5. miR-17-5p is believed to be a key regulator in the progression of cardia cancer. Therefore, the target gene was backtracked according to ncRNA and the Module_target gene_ncRNA interaction network diagram was constructed. The network revealed m1 has the strongest regulation effect in the network. The core gene of m1 is TMOD1, so we believe that TMOD1 plays a key role in the development of cardia cancer. To verify whether the core gene TMOD1 of m1 is central to cardia cancer, we used PPI analysis to verify. In result, the core gene TMOD1 of m1 was at TOP10 in the algorithm, which indicates the importance of TMOD1 to the interaction network.

Conclusions

We suggest that the targeted regulation of miR-17-5p on TMOD1 gene expression affects the neuroactive ligand-receptor interaction pathway and promotes proliferation and apoptosis of cardia cancer cells.

Conflict of Interest

The Authors declare that they have no conflict of interests.

References

- 1) AN JY, BAIK YH, CHOI MG, NOH JH, SOHN TS, BAE JM, KIM S. The prognosis of gastric cardia cancer after R0 resection. *Am J Surg* 2010; 199: 725-729.
- 2) BLASER MJ, SAITO D. Trends in reported adenocarcinomas of the oesophagus and gastric cardia in Japan. *Eur J Gastroenterol Hepatol* 2002; 14: 107-113.
- 3) MAEDA H, OKABAYASHI T, NISHIMORI I, SUGIMOTO T, NAMIKAWA T, DABANAKA K, TSUJII S, ONISHI S, KOBAYASHI M, HANAZAKI K. Clinicopathologic features of adenocarcinoma at the gastric cardia: is it different from distal cancer of the stomach? *J Am Coll Surg* 2008; 206: 306-310.
- 4) TAKEDA J, HASHIMOTO K, UMETANI H, TANAKA T, KOUFUJI K, KAKEGAWA T. Invisible cardia cancer metastasis to the diaphragm. *Kurume Med J* 1992; 39: 77-82.
- 5) TANAKA T, TAKEDA J, HASHIMOTO K, KOUFUJI K, YANO S, KUROIWA T, KAKEGAWA T. Clinical and pathological evaluation of early cancer in the gastric cardia. *Kurume Med J* 1990; 37: 265-269.
- 6) ZHAO H, CHEN W, LIN Y, QIN J, WANG L. Analysis of surgery for incurable gastric cancer. *World J Surg Oncol* 2015; 13: 339.
- 7) MATTIOLI S, DI SIMONE MP, FERRUZZI L, D'OVIDIO F, PILOTTI V, CARELLA R, D'ERRICO A, GRIGIONI WF. Surgical therapy for adenocarcinoma of the cardia: modalities of recurrence and extension of resection. *Dis Esophagus* 2001; 14: 104-109.
- 8) ZHANG CH, HE YL, ZHAN WH, SONG W, CHEN CQ, CAI SR, HUANG MJ. [Impact of spleen preservation on the outcome of radical resection for cardia cancer]. *Zhonghua Wei Chang Wai Ke Za Zhi* 2007; 10: 531-534.
- 9) PARK CH, PARK JC, CHUNG H, SHIN SK, LEE SK, LEE YC. A specific role of endoscopic ultrasonography for therapeutic decision-making in patients with gastric cardia cancer. *Surg Endosc* 2016; 30: 4193-4199.
- 10) CHEN LX, ZHU YR, DOU P. Phagoquosonogram (PASG) diagnosis for cancer of gastric cardia. A preliminary study of 223 patients. *Chin Med J (Engl)* 1989; 102: 844-850.
- 11) HOFFMANN W. [Laser therapy and tube implantation for palliative treatment of inoperable esophageal and cardia cancer]. *Bildgebung* 1993; 60: 157-160.
- 12) BARRETT T, WILHITE SE, LEDOUX P, EVANGELISTA C, KIM IF, TOMASHEVSKY M, MARSHALL KA, PHILLIPPY KH, SHERMAN PM, HOLKO M, YEFANOV A, LEE H, ZHANG N, ROBERTSON CL, SEROVA N, DAVIS S, SOBOLEVA A. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 2013; 41: D991-995.

- 13) LI WQ, HU N, BURTON VH, YANG HH, SU H, CONWAY CM, WANG L, WANG C, DING T, XU Y, GIFFEN C, ABNET CC, GOLDSTEIN AM, HEWITT SM, TAYLOR PR. PLCE1 mRNA and protein expression and survival of patients with esophageal squamous cell carcinoma and gastric adenocarcinoma. *Cancer Epidemiol Biomarkers Prev* 2014; 23: 1579-1588.
- 14) KITAZAWA S. [The science of the mental present: implications of temporal illusions]. *Brain Nerve* 2013; 65: 911-921.
- 15) RITCHIE ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
- 16) LAW CW, CHEN Y, SHI W, SMYTH GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014; 15: R29.
- 17) SMYTH GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; 3: Article3.
- 18) LIU W, LI L, YE H, TU W. [Weighted gene co-expression network analysis in biomedicine research]. *Sheng Wu Gong Cheng Xue Bao* 2017; 33: 1791-1801.
- 19) LANGFELDER P, HORVATH S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
- 20) YU G, WANG LG, HAN Y, HE QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; 16: 284-287
- 21) DONG Y, WANG ZG, CHI TS. Long noncoding RNA Lnc01614 promotes the occurrence and development of gastric cancer by activating EMT pathway. *Eur Rev Med Pharmacol Sci* 2018; 22: 1307-1314.
- 22) GUO W, DONG Z, SHI Y, LIU S, LIANG J, GUO Y, GUO X, SHEN S, WANG G. Methylation-mediated downregulation of long noncoding RNA LOC100130476 in gastric cardia adenocarcinoma. *Clin Exp Metastasis* 2016; 33: 497-508.
- 23) WANG J, ZHANG H, ZHOU X, WANG T, ZHANG J, ZHU W, ZHU H, CHENG W. Five serum-based miRNAs were identified as potential diagnostic biomarkers in gastric cardia adenocarcinoma. *Cancer Biomark* 2018; 23: 193-203.
- 24) PEI G, CHEN L, ZHANG W. WGCNA Application to proteomic and metabolomic data analysis. *Methods Enzymol* 2017; 585: 135-158.
- 25) WANG Y, FENG X, JIA R, LIU G, ZHANG M, FAN D, GAO S. Microarray expression profile analysis of long non-coding RNAs of advanced stage human gastric cardia adenocarcinoma. *Mol Genet Genomics* 2014; 289: 291-302.
- 26) GUO W, DONG Z, LIU S, QIAO Y, KUANG G, GUO Y, SHEN S, LIANG J. Promoter hypermethylation-mediated downregulation of miR-770 and its host gene MEG3, a long non-coding RNA, in the development of gastric cardia adenocarcinoma. *Mol Carcinog* 2017; 56: 1924-1934.
- 27) BHAYYA H, PAVANI D, AVINASH TEJASVI ML, GEETHA P. Oral lymphangioma: a rare case report. *Contemp Clin Dent* 2015; 6: 584-587.
- 28) LIU W, DONG Z, LIANG J, GUO X, GUO Y, SHEN S, KUANG G, GUO W. Downregulation of potential tumor suppressor miR-203a by promoter methylation contributes to the invasiveness of gastric cardia adenocarcinoma. *Cancer Invest* 2016; 34: 506-516.
- 29) JARRIS YS, BARTLEMAN A, HALL EC, LOPEZ L. A pre-clinical medical student curriculum to introduce health disparities and cultivate culturally responsive care. *J Natl Med Assoc* 2012; 104: 404-411
- 30) GUO W, Lv P, LIU S, Xu F, GUO Y, SHEN S, LIANG J, KUANG G, DONG Z. Aberrant methylation-mediated downregulation of long noncoding RNA C5orf66-AS1 promotes the development of gastric cardia adenocarcinoma. *Mol Carcinog* 2018; 57: 854-865.