JBUON 2018; 23(4): 1013-1019
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
E-mail: editorial_office@jbuon.com

ORIGINAL ARTICLE

Identification modules of gastric cancer based on protein-protein interaction networks and gene expression data

Wei Cui[1], Zhenfang Gu[1], Haiying Liu[1], Chunmei Zhang[1], Jie Liu[2]

[1]Department of Oncology, Affiliated Hospital of Jining Medical University, Jining, Shandong Province, P.R. China; [2]Department of Nuclear Medicine, First People's Hospital of Jining, Jining, Shandong Province, P.R. China

Summary

Purpose: *The tumorigenesis of gastric cancer is an intricate process which contains genetic and epigenetic changes of proto-oncogenes and cancer-suppressor genes. The purpose of this study was to identify novel modules for gastric cancer based on protein-protein interaction networks and gene expression data.*

Methods: *Microarray data and corresponding annotated files of E-GEOD-15460 were downloaded from ArrayExpress database. All human protein-protein interactions were downloaded from STRING database. The fast depth-first assay was used to identify all maximal cliques of disease group and control group. Benjamini-Hochberg method was used to perform multiple corrections of p value.*

Results: *248 modules for the control group and 30 modules for the disease group were determined in this research, and 734 pairs of similar or same modules of these two groups were detected through calculating module correlation density. Protein-protein interaction (PPI) network was identified, which comprised of 7899 genes and 48469 interrelationship pairs of genes. Finally, 6 modules with remarkable difference were found to be closely related with gastric cancer.*

Conclusions: *Novel modules with significant difference and the related genes are useful biomarkers and therapeutic targets for gastric cancer.*

Key words: *biomarkers, differential modules, gastric cancer, gene expression data, protein-protein interaction*

Introduction

The tumorigenesis of gastric cancer (GC) is an intricate process which contains genetic and epigenetic changes of proto-oncogenes and cancer-suppressor genes [1]. GC is the fifth most common cancer in the world, with 984,000 cases registered in 2013 worldwide [2]. Moreover, GC ranks second in mortality, with 841,000 deaths in 2013 [3]. The incidence varies widely among countries, and in the developed countries it accounts for 23%, while in the developing countries this figure is 77% [4]. Because alcohol may increase the risk of GC and estrogen lowers the risk of GC, men are more likely to suffer GC than women [5,6]. Therefore, there is an urgent need to explore the underlying mechanisms of GC tumorigenesis.

According to histogenesis, GC can be divided into two main types: intestinal type and diffuse type [7]. Intestinal type GC is often caused by factors of regional environment, dietary habits and infection by *helicobacter pylori (H. pylori)* [8]. And diffuse type GC is often controlled by genetic alterations, such as chromosome deletion, duplication, inversion and translocation [9]. Marshall and Warren firstly demonstrated that *H.pylori* was a main cause for antral gastritis, duodenal and gastric ulcers [10]. Parkin reported that more than half of the popu-

Correspondence to: Jie Liu, MB. Department of Nuclear Medicine, First People's Hospital of Jining, No. 6 Jiankang Road, Jining, 272011, Shandong Province, P.R. China.
Tel: +86 18653798859, Email: pjf933317@163.com
Received: 08/01/2018; Accepted: 23/01/2018

lation worldwide were infected with *H.pylori* [11], and *H.pylori* led to more than 75% occurrence of GC cases [12].

Up to now, there have been various types of molecular targeted therapies for GC, as an ATP-dependent molecular chaperone protein, Heat-shock protein 90 (HSP90) was found to regulate folding and assembly of various proteins [13,14]. Furthermore, Liu et al. reported that ganetespid acting as a strong inhibitor of HSP90 regulated G2/M cell-cycle arrest and cell apoptosis in MGC-803, SGC-7901 and MKN-28 GC cell lines [15]. Mammalian target of rapamycin (mTOR) is negatively related to disease-free survival of patients with lymph node metastasis [16]. Aurora-A encoding threnonine/serine kinases had been reported to play an important role in mitosis and tumor progression [17] and could be used to diagnose patients with worse outcome [18]. In addition, the novel pathways and biomarkers could be identified through analysis of PPI networks, which also can deepen the understanding of underlying mechanisms of tumorigenesis. It was reported that SOX2, USP15, YWHAE and DISC1 were involved in pancreatic carcinogenesis through PPI analysis, gene expression and mutation data [19]. Moreover, 54 corresponding genes were identified to participate in the prognosis of breast cancer by analyzing PPI [20]. Additionally, notch and chemokine signaling pathways were identified to modulate various processes in prostate cancer tumorigenesis [21].

In our study we aimed to identify the molecules related with GC, based on the protein-protein interaction networks and gene expression data. With new PPI networks significant modules would be identified, helping in depth understanding the potential mechanisms involved in GC tumorigenesis.

Methods

The research method used in this study mainly consisted of the following steps (Figure 1):

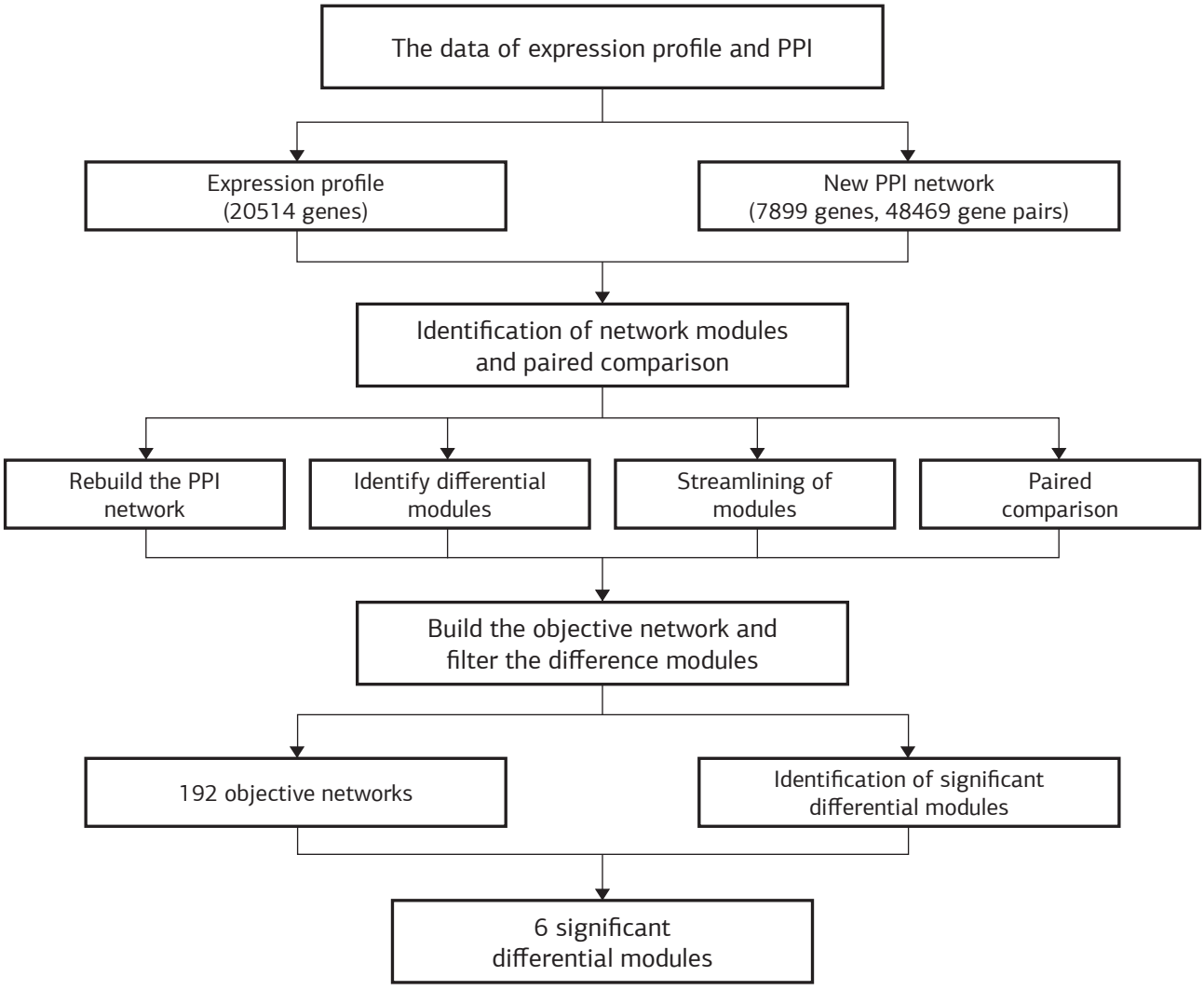

Figure 1. Flow chart showing overall methods and data flow used in this study.

Gene expression data processing

Microarray data and corresponding annotated files of E-GEOD-15460 were downloaded from ArrayExpress database. ArrayExpress database is affiliated to European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI). Three control samples and 227 disease samples were selected from HG-U133_Plus_2 to carry out this experiment.

Robust multichip average (RMA) was employed to correct background. Quantiles assay was used to perform data standardization. Expression data summarization was calculated using medianpolish method. Mapping was done between probes and genes.

PPI data

All human PPIs were downloaded from STRING database. Protein IDs were transited into gene symbols and PPIs without corresponding gene names were eliminated. PPIs were collected if their values of combined score were ≥ 0.8. We further selected PPI pair if its node was the corresponding gene of expression profiling.

Identification of network modules

Pearson correlation coefficient of each pair of the new PPI network was calculated, and its absolute value was regarded as weight value. Subsequently, the PPI network with weight values of control group and disease group was obtained.

The fast depth-first assay [22] was used to identify all maximal cliques of two groups. Maximal cliques were excluded if their nodes were less than 4 or more than 20.

Weight interaction density (WID) of each maximal clique in the control and disease group was calculated and ranked in descending order. Highly overlapping cliques were merged into one module. WID was calculated as follows:

$$d_w(C) = \frac{\sum_{(p,q)\in C} r(p,q)}{\binom{|c|}{2}}$$

(p,q) denotes a gene pair, r (p,q) denotes a weight (between 0 and 1) of one interaction, C denotes a clique.

Module correlation density values of control and disease groups were calculated to find similar or same modules in this study. The module correlation density was calculated as follows:

$$d_{cc}(S_i) = \frac{\sum_{p,q\in S_i} PCC((p,q),N)}{\binom{|S_i|}{2}}$$

S_i represents a set of modules, PCC ((p,q), N) denotes Pearson correlation of (p,q) under normal condition.

Construction of differential modules

Modules with the same genes were chosen from two groups to form objective networks. Local entropy of each node in each network was calculated as follows:

$$S_i = -\frac{1}{\log k_i}\sum_{j\in N(i)} p_{ij}\log p_{ij}$$

k_i represents degree of node i, N(i) denotes adjacent node set of node i, p_{ij} represents network random probability matrix. p_{ij} was defined as follows:

$$p_{ij} = \frac{|c_{ij}|}{\sum_{k\in N(i)} c_{ik}}$$

c_{ij} represents Pearson correlation coefficient between protein i and protein j. The global entropy of two groups were calculated as follows:

$$S = \sum_{i=1}^{n} C_i S_i$$

n denotes number of nodes in networks, C_i represents degree centrality of node i and C_i was calculated as follows:

$$C_i = \frac{k_i}{n-1}$$

The difference of global entropy of control and disease groups was calculated as follows:

$$\Delta S = S^I - S^N$$

S^I and S^N denote global entropy of disease group and control group, respectively.

Statistics

Permuted sample labels were used to replicate the analysis. Global entropy of control and disease groups was recalculated. Benjamini-Hochberg method [23] was used to do multiple correction of p value. Modules were considered as significant at $p<0.05$.

Results

Construction of PPI networks

Gene data was preprocessed by using RMA, quantiles and medianpolish methods, and the expression profile containing 20514 genes was obtained.

8590 nodes and 53975 interrelationship pairs of genes were obtained by analyzing PPI network. Based on these data, a new PPI network containing 7899 genes and 48469 interrelationship pairs of genes was obtained after mapping between genes and nodes.

Identification of gene expression modules

Through using fast depth-first method, 26416 maximal cliques in the control and disease groups were observed. Moreover, 7951 maximal cliques were identified in each group after excluding the maximal cliques which were less than 4 or more than 20 nodes.

The refinement of modules was performed by calculating WID values. Subsequently, 248 and 30

modules were determined in the control and disease groups, respectively. 734 pairs of similar or same modules of these two groups were obtained by calculating module correlation density.

Construction of objective modules and identification of significant differential modules

In total, 192 objective networks containing genes from the control and disease groups were identified, and the local entropy of each node in every objective network was calculated to obtain global entropy.

P values of modules were corrected using Benjamini-Hochberg method. Finally, 6 modules were identified as significant differential modules (Figure 2 and Table 1). The nodes of 6 significant differential modules ranged from 7 to 9 and edges ranged from 21 to 36. Differences of global entropy

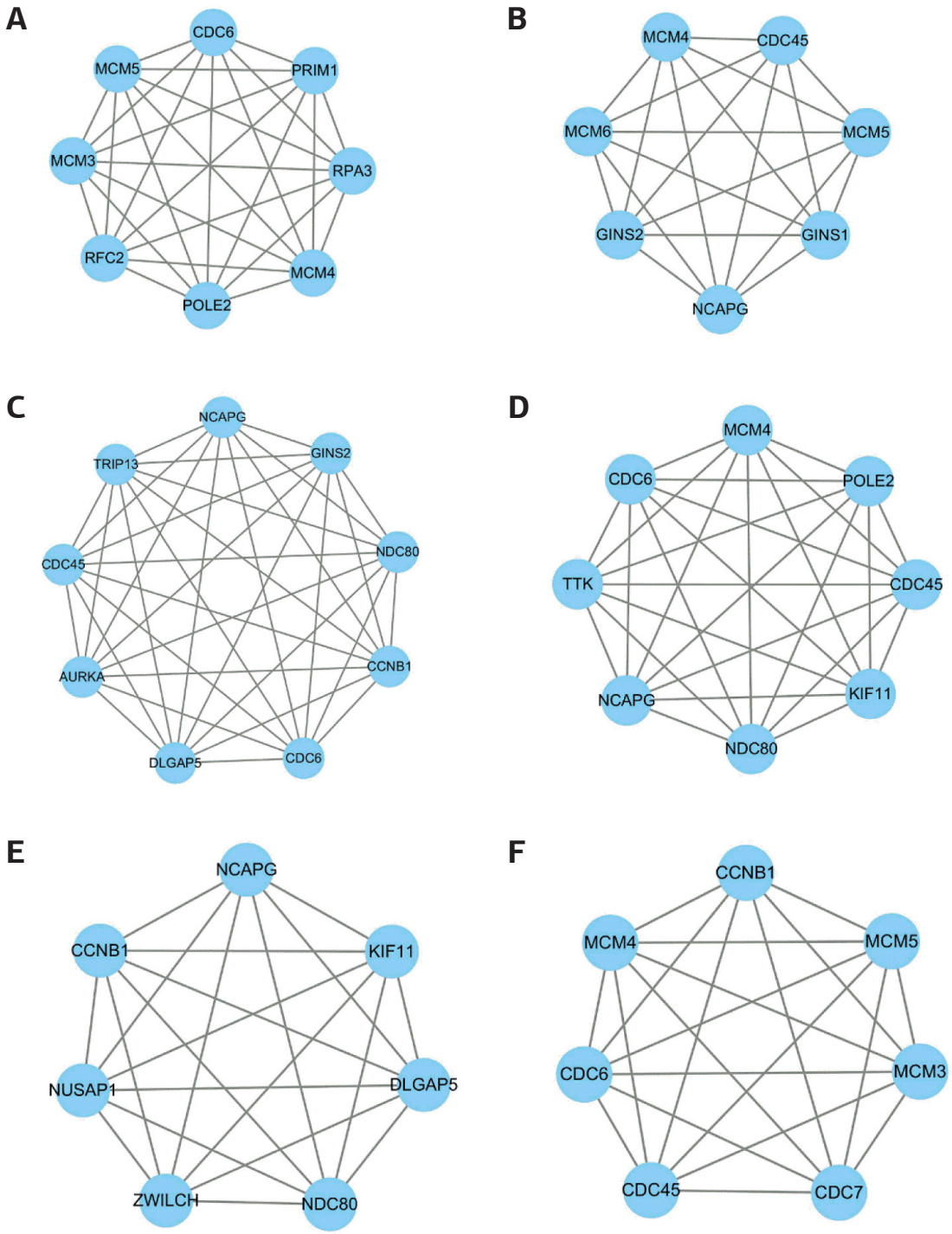

Figure 2. Six significant differential modules were identified for GC. **A:** 8 genes, **B:** 7 genes, **C:** 9 genes, **D:** 8 genes, **E:** 7 genes and **F:** 7 genes formed significant modules related with GC.

Table 1. The 6 significant differential modules identified

Modules	*Change of entropy*	*Number of nodes*	*Number of edges*	*p value*
module 1	0.483	8	28	0.014
module 2	1.395	7	21	0
module 3	2.368	9	36	0
module 4	1.327	8	28	0
module 5	0.872	7	21	0.007
module 6	1.263	7	21	0

ranged from 0.483 to 2.368. Moreover, we also found some highly frequent genes involved in 6 significant modules.

Discussion

In our study, PPI networks and gene expression data analyses were performed to identify novel pathways associated with GC. Six significant differential modules were identified for the control and disease groups and also some highly frequent genes were detected in 6 modules, such as Cell division cycle protein 6 (Cdc6), Cell division cycle protein 45 (Cdc45), KIF11, Ndc80, CCNB1, GINS2, and MCM4.

Among them, Cdc45, MCM2-7 and GINS proteins are the three main parts of eukaryotic replicative helicase CMG. Overexpression of Cdc45 induced an early S phase arrest and significantly increased firing rate of replication of Hela cells [24]. Cdc45 protein was found to be not involved in the cell cycle of proliferating cells and disappeared in senescent and terminally differentiated cells. Initiation of replication was reported to be limited through binding with Cdc45 protein [25]. Based on these previous studies, we speculated that Cdc45 level might decrease in GC. However, there were no studies about the effect of Cdc45 in GC until now. The potential role of Cdc45 in cell cycle suggested that it might participate in tumorigenesis of GC.

KIF11 had been reported to be overexpressed in many human cancers, such as lung cancer, breast cancer, pancreatic cancer and ovarian cancer [26]. Consistent with the above studies, the KIF11 expression was also found to be significantly increased in GC tissues compared with normal tissues in a previous study [27]. And KIF11 could act as a biomarker for intestinal mucin GC phenotype. In the next experiment, we will further investigate the potential role of KIF11 in cell proliferation, migration and invasion in GC.

As a part of kinetochores, Ndc80 exhibited an important role in G2/M phase of the cell cycle during mitosis [28]. Moreover, it had been reported that upregulation of Ndc80 was identified in GC which promoted cell migration and invasion [29]. Further research indicated that overexpression of Ndc80 caused chromosome instability and imbalance of cell division [29]. Therefore, we speculated that Ndc80 could function as a biomarker for treatment of GC. Additionally, Tang et al. reported that Ndc80 overexpression could absorb microtubule assembling factors, which could interfere with the microtubule balance and mitosis by Ndc80 internal loop [30].

As a member of cyclin family, CCNB1 is a highly conserved and constitutive expression protein [31]. Consistent with our study, CCNB1 had been found to be a prognostic biomarker for estrogen receptor positive breast cancer [32]. In addition, G2/M phase arrest could be induced by inhibition of CCNB1 in colorectal cancer and inhibition of CCNB1 induced apoptosis of certain type of colorectal cancer cells [33]. More importantly, CCNB1, CCNB2 and c-MYC genes could increase GC cell proliferation which are regulated by ISL1 [34]. These studies confirmed our observations as well. However, we found that CCNB1 could regulate GC cell proliferation via other transcriptional factors as well. According to these findings, we predicted that CCNB1 might be significantly increased in GC which needs further confirmation.

GINS consist of four parts: PSF1, GINS2 (PSF2), PSF3 and SLD5, and GINS were found to play a pivotal role in opening double strands of DNA ahead of replication fork [35]. Liang et al. reported that expression of GINS2 was increased in triple negative breast cancer (TNBC) cells and knockdown of GINS2 inhibited the cell invasion and growth in TNBC [36]. These studies implied that GINS2 might be increased and influence the GC cell growth. However, the specific function of GINS has not been identified in GC. And in the further study, we will confirm the above inference via biological experiments.

Minichromosome maintenance-deficient 4 (MCM4) is a component of MCM2-7 complex, which plays an important role in DNA replication

and maintenance of chromosome structures. It had been reported that MCM4-deficiency led to human adrenal insufficiency, NK cell deficiency and short stature [37]. We considered that MCM4 might play a role in GC initiation and further validation about the change in its expression will been investigated in our following experiments which has not been found in previous studies.

In summary, we performed an integrated analysis by combining PPI networks and gene expression data for GC. Fast depth-first and Benjamini-Hochberg algorithm were used to perform data analysis. And 6 significant differential modules were identified in this study. Moreover, genes of these modules play an important role in biological processes, such as DNA replication, chromosome stability, and adrenal secretion. These differential modules and corresponding genes may provide outstanding therapeutic targets and biomarkers for the treatment of GC.

Conflict of interests

The authors declare no conflict of interests.

References

1. Parkin DM, Bray F, Ferlay J et al. Estimating the world cancer burden: Globocan 2000. Int J Cancer 2001;94:153-6.
2. Fitzmaurice C, Dicker D, Pain A et al. The Global Burden of Cancer 2013. JAMA Oncol 2015;1:505-27.
3. Ajani JA, Bentrem DJ, Besh S et al. Gastric cancer, version 2.2013: featured updates to the NCCN Guidelines. J Natl Compr Canc Netw 2013;11:531-46.
4. Figueiredo C, Constanza Camargo M, Leite M et al. Pathogenesis of Gastric Cancer: Genetics and Molecular Classification. Curr Top Microbiol Immunol 2017;400:277-304.
5. Makarem N, Lin Y, Bandera EV et al. Concordance with World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) guidelines for cancer prevention and obesity-related cancer risk in the Framingham Offspring cohort (1991-2008). Cancer Causes Control 2015;26:277-86.
6. Camargo MC, Goto Y, Zabaleta J et al. Sex hormones, hormonal interventions, and gastric cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2012;21:20-38.
7. Lauren P. The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. Acta Pathol Microbiol Scand 1965;64:31-49.
8. Park WS, Oh RR, Park JY et al. Frequent somatic mutations of the beta-catenin gene in intestinal-type gastric cancer. Cancer Res 1999;59:4257-60.
9. Hudler P. Genetic aspects of gastric cancer instability. Sci World J 2012;2012:761909.
10. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1984;1:1311-5.
11. Parkin DM. International variation. Oncogene 2004;23:6329-40.
12. de Martel C, Ferlay J, Franceschi S et al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. Lancet Oncol 2012;13:607-15.
13. Neckers L, Workman P. Hsp90 molecular chaperone inhibitors: are we there yet? Clin Cancer Res 2012;18:64-76.
14. Trepel J, Mollapour M, Giaccone G et al. Targeting the dynamic HSP90 complex in cancer. Nat Rev Cancer 2010;10:537-49.
15. Liu H, Lu J, Hua Y et al. Targeting heat-shock protein 90 with ganetespib for molecularly targeted therapy of gastric cancer. Cell Death Dis 2015;6:e1595.
16. An JY, Kim KM, Choi MG et al. Prognostic role of p-mTOR expression in cancer tissues and metastatic lymph nodes in pT2b gastric cancer. Int J Cancer 2010;126:2904-13.
17. Marumoto T, Zhang D, Saya H. Aurora-A - a guardian of poles. Nat Rev Cancer 2005;5:42-50.
18. Wang J, Yang S, Zhang H et al. Aurora-A as an independent molecular prognostic marker in gastric cancer. Oncol Rep 2011;26:23-32.
19. Srihari S, Ragan MA. Systematic tracking of dysregulated modules identifies novel genes in cancer. Bioinformatics 2013;29:1553-61.
20. Chen J, Sam L, Huang Y et al. Protein interaction network underpins concordant prognosis among heterogeneous breast cancer signatures. J Biomed Inform 2010;43:385-96.
21. Wang JM, Wu JT, Sun DK et al. Pathway crosstalk analysis based on protein-protein network analysis in prostate cancer. Eur Rev Med Pharmacol Sci 2012;16:1235-42.
22. Feng MJ, Chu HM, Shen CJ et al. Attractors of hypertrophic cardiomyopathy using maximal cliques and attract methods. Comput Biol Chem 2017;67:194-9.
23. Pawluk-Kolc M, Zieba-Palus J, Parczewski A. Application of false discovery rate procedure to pairwise comparisons of refractive index of glass fragments. Forensic Sci Int 2006;160:53-8.
24. Kohler C, Koalick D, Fabricius A et al. Cdc45 is limiting for replication initiation in humans. Cell Cycle 2016;15:974-85.
25. Pollok S, Bauerschmidt C, Sanger J et al. Human Cdc45 is a proliferation-associated antigen. FEBS J 2007;274:3669-84.
26. LoRusso PM, Goncalves PH, Casetta L et al. First-in-human phase 1 study of filanesib (ARRY-520), a kinesin

spindle protein inhibitor, in patients with advanced solid tumors. Invest New Drugs 2015;33:440-9.

27. Imai T, Oue N, Nishioka M et al. Overexpression of KIF11 in Gastric Cancer with Intestinal Mucin Phenotype. Pathobiology 2017;84:16-24.
28. Martin-Lluesma S, Stucke VM, Nigg EA. Role of Hec1 in spindle checkpoint signaling and kinetochore recruitment of Mad1/Mad2. Science 2002;297:2267-70.
29. Qu Y, Li J, Cai Q et al. Hec1/Ndc80 is overexpressed in human gastric cancer and regulates cell growth. J Gastroenterol 2014;49:408-18.
30. Tang NH, Toda T. MAPping the Ndc80 loop in cancer: A possible link between Ndc80/Hec1 overproduction and cancer formation. Bioessays 2015;37:248-56.
31. Miyazaki T, Arai S. Two distinct controls of mitotic cdk1/cyclin B1 activity requisite for cell growth prior to cell division. Cell Cycle 2007;6:1419-25.
32. Ding K, Li W, Zou Z et al. CCNB1 is a prognostic biomarker for ER+ breast cancer. Med Hypotheses 2014;83:359-64.
33. Fang Y, Yu H, Liang X et al. Chk1-induced CCNB1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. Cancer Biol Ther 2014;15:1268-79.
34. Shi Q, Wang W, Jia Z et al. ISL1, a novel regulator of CCNB1, CCNB2 and c-MYC genes, promotes gastric cancer cell proliferation and tumor growth. Oncotarget 2016;7:36489-500.
35. Zheng M, Zhou Y, Yang X et al. High GINS2 transcript level predicts poor prognosis and correlates with high histological grade and endocrine therapy resistance through mammary cancer stem cells in breast cancer patients. Breast Cancer Res Treat 2014;148:423-36.
36. Liang P, Song Z, Chen D et al. GINS2 regulates matrix metallopeptidase 9 expression and cancer stem cell property in human triple negative breast cancer. Biomed Pharmacother 2016;84:1568-74.
37. Hughes CR, Guasti L, Meimaridou E et al. MCM4 mutation causes adrenal failure, short stature, and natural killer cell deficiency in humans. J Clin Invest 2012;122:814-20.