Identification of Biomarkers for Diagnosis of Gastric Cancer by Bioinformatics

Da-Guang Wang¹, Guang Chen²*, Xiao-Yu Wen³, Dan Wang⁴, Zhi-Hua Cheng², Si-Qiao Sun²

Abstract

Background: We aimed to discover potential gene biomarkers for gastric cancer (GC) diagnosis. Materials and Methods: Genechips of 10 GC tissues and 10 gastric mucosa (GM, para-carcinoma tissue, normal control) tissues were generated using an exon array of Affymetrix containing 30,000 genes. The differentially expressed genes (DEGs) between GC tissues and normal control were identified by the Limma package and analyzed by hierarchical clustering analysis. Gene ontology (GO) and pathway enrichment analyses were performed for investigating the functions of DEGs. Receiver operating characteristics (ROC) analysis was performed to measure the effects of biomarker candidates for diagnosis of GC. Results: Totals of 896 up-regulated and 60 down-regulated DEGs were identified to be differentially expressed between GC samples and normal control. Hierarchical clustering analysis showed that DEGs were highly differentially expressed and most DEGs were up-regulated. The most significantly enriched GO-BP term was revealed to be mitotic cell cycle and the most significantly enriched pathway was cell cycle. The intersection analysis showed that most significant DEGs were cyclin B1 (CCNB1) and cyclin B2 (CCNB2). The sensitivities and specificities of CCNB1 and CCNB2 were both significantly high (<0.0001). Areas under the ROC curve for CCNB1 and CCNB2 were both greater than 0.9 (p<0.0001). Conclusions: CCNB1 and CCNB2, which were involved in cell cycle, played significant roles in the progressions and development of GC and these genes may be potential biomarkers for diagnosis and prognosis of GC.

Keywords: Gastric cancer - differentially expressed genes - diagnosis - biomarkers

Introduction

Gastric cancer (GC), is a kind of digestive cancer developing from the lining of the stomach (http://www.cancer.gov/cancertopics/types/stomach). GC is often asymptomatic or nonspecific symptoms in its early stages (Leung et al., 2008). However, by the time symptoms occur, GC has commonly reached the advanced stage and may have metastasized. More than 60% of cases are caused by infection with Helicobacter pylori (Gonzalez et al., 2013). The incidence of GC has occurred the 4th and the mortality of GC ranked the second among all tumor diseases worldwide (Orditura et al., 2014). The 5-year survival rates are disappointingly less than 10% globally for the reason that most patients have reached advanced stage (Orditura et al., 2014). GC has posed a great threat to human health all over the world.

An increasing number of studies have focused on the prevention and diagnosis of GC. Some traditional tumor markers are usually used in diagnosing and staging GC progress, such as carbohydrate antigen and cancer embryo antigen (Carpelan-Holmstrom et al., 2001; Takahashi et al., 2003). Besides, miRNAs, including miR-1, miR-20a, miR-27a, miR-34 and miR-423-5p, are identified as biomarkers for GC detection and indicate tumor progression stages (Liu et al., 2011). In addition, cluster of differentiation 44 is used as a cell surface marker to identify GC initiating cells (Takaishi et al., 2009). Moreover, the protein expression of epidermal growth factor receptor-2, which is associated with serosal invasion and lymph node metastasis, is an important independent prognostic indicator in GC (Yonemura et al., 2009). In this paper, we generated genechips of 10 GC tissues and 10 gastric mucosa (GM, para-carcinoma tissue) tissues using an exon array of Affymetrix containing 30000 genes. Differentially expressed genes (DEGs) between GC and matched normal control were identified for hierarchial clustering analysis and functional analysis. Furthermore, the effects of selected biomarker candidates in distinguishing GC from normal control were detected by receiver operating characteristics (ROC) analysis.

¹Gastrointestinal surgery, ²Vascular surgery, ³Hepatobiliary and pancreatic, ⁴Gastrointestinal Department of Internal Medicine, the First Hospital of Jilin University, Changchun, China  *For correspondence: guangchenchdh@163.com

DOI:http://dx.doi.org/10.7314/APJCP.2015.16.4.1361
Identification of Biomarkers for Gastric Cancer Diagnosis by Informatics
We aimed to discover the potential biomarkers for GC diagnosis.

Materials and Methods

Patients
The study was performed in our hospital between February 2014 and April 2014. A total of 10 people histology diagnosed as GC aged 42-71 (mean age 57.8 ± 6.3 years) were undergoing surgery to remove tumors. These patients included 7 males and 3 females. All the patients were treated without radiotherapy or chemotherapy. GM tissue, more than 5 cm far from the edge of tumor, was selected and used as normal control in each sample.

RNA isolation and genechip hybridization
Total RNA from GC and normal control samples was isolated using TRIzol reagent (Invitrogen, CA, USA) according to the manufacturer’s instructions. RNA purity was determined using an ultraviolet and visible spectrophotometer (UNIC, Shanghai, China). Samples with RNA purity between 1.8-2.2 were used for genechip hybridization and scanning. Labeling, hybridization and staining of these samples were performed according to the Eukaryotic Target Preparation protocol in the Affymetrix Technical Manual (701021 rev. 4, Affymetrix Santa Clara, CA, USA). In summary, 1 μg of purified total RNA was used in the synthesis of cDNA and the cDNA was purified using the Genechips Sample Cleanup Module (Affymetrix Santa Clara, CA, USA). The purified cDNA was amplified to produce biotin labeled cRNA using BioArray HighYield RNA transcript labeling kit (T7, Enzo Life Sciences, Farmingdale, USA). Labeled cRNA was fragmented and hybridized to GeneChip Human Gene 2.0 ST according to the manufacturer’s protocol (Affymetrix Santa Clara, CA, USA). The hybridized genechips were washed and stained for antibody amplification stain. Then, the genechips were scanned using Gene array Scanner 3000 7G (Affymetrix Santa Clara, CA, USA).

Data preprocessing and screening of DEGs
GeneChip® Operating Software (Affymetrix Santa Clara, CA, USA) was used to gather signal value. Normalization algorithms were used to adjust sample signals by minimizing the effects of variation caused by non-biological factors. Quantile normalization was performed by the robust multiarray average (Irizarry et al., 2003) algorithm with application of Affy package in R statistical software program (Bolstad et al., 2003). Gene expression values of samples were log2-transformed and median-centered for further analysis. If multiple probes corresponded to a same gene, the mean value was calculated as the expression value of this gene. The Limma package (Smyth, 2004) in R language was used to screen DEGs. The DEGs with llog2 fold change (FC) ≥1.0 and p value<0.05 were considered to be significant.

Hierarchical clustering analysis of DEGs
To generate an overview of the gene expression profile, we used hierarchical clustering analysis, which produces a unique set of nested categories or clusters by sequentially pairing variables, clusters, or variables and clusters (BRIDGES JR, 1966). The gene expression profiles of the selected DEGs were performed hierarchical clustering analysis based on Euclidean distance using the “pheatmap” package in R language (Team, 2012) and then the heat map was generated.

Gene Ontology (GO) functional and pathway enrichment analysis of DEGs
GO functional analysis is a commonly used approach for functional studies of genomic or transcriptomic data (Consortium, 2004). GO categories include molecular function (MF), biological process (BP), and cellular component (CC). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a knowledge base which is used for systematic analysis of enzymatic pathways, gene functions, and linking genomic information with higher order functional information (Kanehisa and Goto, 2000). Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Da Wei Huang and Lempicki, 2008) provides exploratory visualizaion tools that promote discovery through functional classification, biochemical pathway maps, and conserve protein domain architectures (Dennis et al., 2003). In order to analyze the DEGs in function level, we performed GO-BP and KEGG pathway enrichment analysis for DEGs by DAVID. The p-value less than 0.05 was chosen as cut-off criterion. DEGs enriched in the most significant GO terms and pathways were intersection analyzed to selected biomarker candidates.

ROC analysis of biomarker candidates
A ROC graph is a technique for visualizing, organizing and selecting classifiers based on their performance (Fawcett, 2006). A diagnostic test was firstly performed in order to estimate the diagnostic value of candidate biomarkers in GC. First, the 10 couples of samples were diagnosed to be GC or no-GC according to histological examination. Then, the 10 couples of genechips were divided into cancer group and normal control group according to the expression value of candidate biomarkers, which was used as a test threshold. Sensitivity (true positive rate) and specificity (true negative rate) of each biomarker in this diagnostic test were calculated. Finally, corresponding ROC curves were obtained by plotting the sensitivity, against the 1- specificity using the Medcalc statistical software (MedCalc Software, Mariakerke, Belgium). Area under the ROC curve (AUC) was calculated to estimate the accuracy of this diagnostic test. A test with AUC bigger than 0.9 gets high accuracy, 0.7-0.9, moderate accuracy and 0.5-0.7, low accuracy.

Results
Screening of DEGs and hierarchical clustering analysis
According to the cut-off criteria of p value<0.05 and llog2 FCi ≥1.0, a total of 956 DEGs were obtained, including 60 down-regulated and 896 up-regulated DEGs.

The heat map of hierarchical clustering analysis for the DEGs was showed in Figure 1. It was obvious to see that the expressions of DEGs between GC and normal
controls were significantly different and most DEGs were up-regulated.

**GO-BP and pathway enrichment analysis**

The top 15 GO terms enriched by DEGs in GO-BP category were listed in Figure 2. Terms related to cell cycle got obvious statistical significance. The most enriched GO term was mitotic cell cycle. The top 15 KEGG pathways were showed in Figure 3. The most enriched pathway was cell cycle. DEGs in mitotic cell cycle and cell cycle were listed in Table 1. The intersection analysis showed that 14 DEGs were involved in both mitotic cell cycle and cell cycle (Table 2). The most significant DEGs were cyclin B1 (CCNB1) and cyclin B2 (CCNB2) with the FC value of 3.92 and 3.76, respectively.

**ROC analysis**

The sensitivities of both CCNB1 and CCNB2 were 100% and the specificities of both these two DEGs were 90% ($p<0.0001$). AUCs for CCNB1 and CCNB2 were 0.940 and 0.950, respectively ($p<0.0001$).

**Discussion**

GC is estimated to be one of the most common and frequent malignant tumor of the digestive system. Cancer-related morbidity of GC ranks the second after lung cancer.

**Table 1. DEGs Enriched in Mitotic Cell Cycle or Cell Cycle**

<table>
<thead>
<tr>
<th>DEGs</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitotic cell cycle term</td>
<td>2.30E-25</td>
</tr>
<tr>
<td>KIF23, PRC1, KNTC1, TTK, AURKA, PTTG1, KIF2C, SEH1L, OIP5, MTBP, CDCA2, PSMD3, NUP37, CCNA2, ASPM, CDC3, CDC6, KIF11, SGOL1, KIF15, TPX2, NUSAP1, PBK, CDK2, INHBA, MAD2L1, PSEME1, SPAG5, PSMA5, PSMA4, C14orf106, PSMA3, BUB1B, CCDC99, BLM, NEK3, TIPIN, ANLN, CHEK1, CEP55, SPC25, TUBB, NCAPH, PSMB1, NCAPG, PSMB3, PSMB2, CDC123, BUB1, FBXO5, UBE2D1, ZWILCH, HILLS, ERC6C6L, PINX1, YEATS4, TXNL4B, NUF2, KIF18A, CENPF, CDC20, NDC80, CENPE, CDK5, SUGT1, SMG2, CCNB1, PSMD14, CCNB2, NOLC1, PLK1, PSMD10, C13orf34, ESI</td>
<td>2.30E-25</td>
</tr>
<tr>
<td>Cell cycle pathway</td>
<td>7.15E-09</td>
</tr>
<tr>
<td>YWHAZ, TTK, PRKDC, CHEK1, SFN, PTTG1, CHEK2, CCNE2, ORC6L, BUB1, CCNA2, TFDP1, CDC7, CDC6, YWHAH, TP53, CDC20, YWHAE, MCM4, CDK2, MCM5, ORC1L, MCM6, CCNB1, CCNB2, MAD2L1, PLK1, BUB1B</td>
<td>7.15E-09</td>
</tr>
</tbody>
</table>

**Figure 1. Hierarchical Clustering of DEGs.** Abscissa axis: 10 couples of tissue samples, where “C” stands for gastric cancer samples, “N” stands for the normal control. Vertical axis: DEGs, where red color, up-regulated DEGs; green color, down-regulated DEGs.

**Figure 2. GO-BP Enrichment Analysis for DEGs.** Abscissa axis: the value of -Log2P. Vertical axis: the top 15 GO terms. -Log2P: -log2 transformed of $p$-value.

**Figure 3. KEGG Pathway Enrichment Analysis for DEGs.** Abscissa axis: the value of -Log2P. Vertical axis: the top 15 pathways. -Log2P: -log2 transformed of the $p$ value.

**Figure 4. ROC Curve for CCNB1 and CCNB2.** Abscissa axis: 1- specificity; vertical axis: sensitivity.
in China (Yang, 2006). In recent study, bioinformatics method has been used to predict the potential microRNA biomarkers for early detection of GC (Liu et al., 2012). In our work, we used bioinformatics approach to screen the potential gene biomarkers for GC diagnosis. Our results suggested that CCNB1 and CCNB2 were the most significant biomarker candidates. They were both involved in mitotic cell cycle GO-BP term and cell cycle pathway. ROC analysis showed that the sensitivities and specificities of CCNB1 and CCNB2 in diagnostic tests were high (>0.9, 0.9) and AUCs for CCNB1 and CCNB2 were both greater than 0.9, which indicated that this diagnostic tests got high accuracy.

CCNB1 and CCNB2 are members of cyclin B family, which are critical for the cells checking into or out of M phase in the cell cycle (Ford and Pardee, 1999; Zhou et al., 2002). CCNB1 is the best studied and characterized member of the cyclin B family (Yuan et al., 2005). It is a regulatory protein involved in mitosis and predominantly expressed during G2/M phase (Chow et al., 2003). Our function analysis showed that CCNB1 was involved in the GO-BP term and KEGG pathway related to the cell cycle of GC. It is known that lacking of regulation in the cell cycle is one of the hallmarks of cancer. The overexpression of CCNB1 can result in uncontrolled cell growth through cell cycle in cancer cells. High expression level of CCNB1 has been detected in variety of cancers, including breast cancer (Kawamoto et al., 1997), colorectal cancer (Wang et al., 1997), prostate cancer (Masahal et al., 1996) and oral cancer (Kushner et al., 1999), as well as GC (Yasuda et al., 2005). It corresponded with our result that CCNB1 was 3.92 FC in GC compared with normal control. In addition, the suppression of CCNB1 by huanglian treatment can inhibit tumor cell growth in GC through retention of cells in G2 (Li et al., 2000). Furthermore, CCNB1 may be also involved in the genesis of GC, the overexpression of CCNB1 may play important roles in human gastric carcinogenesis (Kim, 2007). Thus, CCNB1 may be a critical target in the progression of GC.

Research also shows that the overexpression of CCNB1 predominantly occurs in the early stage of GC (Yasuda et al., 2002). It has been confirmed that high expression levels of CCNB1 usually occur before tumor cells get immortalization [22]. Therefore, high level of CCNB1 may be used as a biomarker to reveal the progression of GC in early stage. Previous research indicates aberrantly expressed CCNB1 in tumors and premalignant lesions should be further explored as diagnostic markers (Suzuki et al., 2005). Our results also showed that diagnosis test that used CCNB1 expression value as a test threshold got high sensitivity and specificity. This means that CCNB1 expression value maybe have the potential to be a biomarker in the diagnosis of GC. Moreover, high levels of CCNB1 also indicate lymph node metastasis and poor prognosis in GC (Begnami et al., 2010). Thus, CCNB1 may be a potential biomarker in the diagnosis and prognosis of GC.

Similar to CCNB1, CCNB2 is an essential component of the cell cycle regulatory machinery, too. The expression of CCNB2 carrying a mutation at arginine 32 arrests HeLa cells in a pseudo mitotic state (Gallant and Nigg, 1992). Furthermore, overexpression of CCNB2 alters the spindle checkpoint, which results into chromosomal instability (Sarafan-Vasseur et al., 2002). The progression of GC also intimately connects with cell cycle. Cancer represents a dysregulation of the cell cycle such as overexpress cyclins to undergo unregulated cell growth (Schwartz and Shah, 2005). The loss of cell cycle inhibitor can lead to lymph node metastasis of GC (Kim et al., 2000). Moreover, present evidences show that cell cycle arrest can inhibit cell growth and cell proliferation of GC (Lin et al., 2006; Otsubo et al., 2008), and G2/M cell cycle arrest is observed associated with a marked decreased cyclin B (Lin et al., 2006). Accordingly, CCNB2 may take part in the progression of GC through cell cycle. Although the present evidences of direct association between CCNB2 and GC progression are rare, CCNB2 may be another key regulator in GC development. At last, CCNB2 has been identified as a biomarker for the diagnosis of other cancers, such as lung cancer (Hofmann et al., 2004), colorectal cancer (Park et al., 2007) and cervical cancer (Garcia et al., 2013). These evidences suggest that CCNB2 may be a potential biomarker for diagnosis of GC.

In summary, CCNB1 and CCNB2, which were involved in cell cycle, played significant roles in the progression and development of GC. Our study proposes these genes to be potential biomarkers for diagnosis and prognosis of GC. However, further studies are necessary for verifying the clinical applications of these genes as biological markers for GC diagnosis.

**Acknowledgements**

We wish to express our warm thanks to Fenghe (Shanghai) Information Technology Co., Ltd. Their ideas and help gave a valuable added dimension to our research.

**References**


