Comparative analysis of gene expression profiles of gastric cardia adenocarcinoma and gastric non-cardia adenocarcinoma

BIN SONG¹, JUAN DU², NENG DENG¹, JI-CHEN REN² and ZHEN-BO SHU¹

¹Department of Gastrointestinal Surgery, China-Japan Union Hospital, Jilin University, Changchun, Jilin 130033; ²Second Department of Internal Medicine, The Tumor Hospital of Jilin, Changchun, Jilin 130012, P.R. China

Received April 24, 2015; Accepted August 4, 2016

DOI: 10.3892/ol.2016.5161

Abstract. In the present study, gene expression profiles were analyzed to identify the molecular mechanisms underlying gastric cardia adenocarcinoma (GCA) and gastric non-cardia adenocarcinoma (GNCA). A gene expression dataset (accession number GSE29272) was downloaded from Gene Expression Omnibus, and consisted of 62 GCA samples and 62 normal controls, as well as 72 GNCA samples and 72 normal controls. The two groups of differentially-expressed genes (DEGs) were compared to obtain common and unique DEGs. A differential analysis was performed using the Linear Models for Microarray Data package in R. Functional enrichment analysis was conducted for the DEGs using the Database for Annotation, Visualization and Integrated Discovery. Protein-protein interaction (PPI) networks were constructed for the DEGs with information from the Search Tool for the Retrieval of Interacting Genes. Subnetworks were extracted from the whole network with Cytoscape. Compared with the control, 284 and 268 genes were differentially-expressed in GCA and GNCA, respectively, of which 194 DEGs were common between GCA and GNCA. Common DEGs [e.g., claudin (CLDN)7, CLDN4 and CLDN3] were associated with cell adhesion and digestion. GCA-unique DEGs [e.g., MAD1 mitotic arrest deficient like 1, cyclin (CCN)B1, CCNB2 and CCNE1] were associated with the cell cycle and the regulation of cell proliferation, while GNCA-unique DEGs (e.g., GATA binding protein 6 and hyaluronoglucosaminidase 1) were implicated in cell death. A PPI network with 141 nodes and 446 edges were obtained, from which two subnetworks were extracted. Genes [e.g., fibronectin 1, collagen type I α2 chain (COL1A2) and COL1A1] from the two subnetworks were implicated in extracellular matrix organization. These common DEGs could advance our understanding of the etiology of gastric cancer, while the unique DEGs in GCA and GNCA could better define the properties of specific cancers and provide potential biomarkers for diagnosis, prognosis or therapy.

Introduction

Gastric cancer is the third leading cause of cancer-associated mortality (1). The most common cause of gastric cancer is infection by the bacteria Helicobacter pylori, which accounts for ~60% of cases (1,2). Smoking also increases the risk significantly. The prognosis of stomach cancer is generally poor due to the fact that the tumor has often metastasized by the time of diagnosis (3), which makes it necessary to identify biomarkers for an early diagnosis.

Stomach cancers are overwhelmingly adenocarcinomas. Gastric adenocarcinomas are a heterogeneous group of tumors. The cardia lies between the end of the esophagus and the body of the stomach, and is a small macroscopically indistinct zone that lies immediately distal to the gastroesophageal junction. Gastric cardia adenocarcinoma (GCA) and esophageal squamous cell carcinoma (ESCC) share certain etiological risk factors. Abnet et al reported a shared susceptibility locus in PLCE1 at 10q23 for GCA and ESCC (4), GCA may have a distinct etiology. Substantially higher TP53 mutation rates have been detected in cases with GCA than gastric non-cardia adenocarcinoma (GNCA) (5). Kamangar et al indicated that H. pylori is a strong risk factor for non-cardia gastric cancer, but that it is inversely associated with the risk of gastric cardia cancer (6). Kim et al found differences in the clinicopathology and protein expression in cardia carcinoma and non-cardia carcinoma (7).

Although a number of genetic alterations have been identified in gastric cancer, including those in cadherin 1 (CDH1) (8,9), β-catenin (10), CDH17 (11) and Met (12), no study has distinguished these alterations by anatomical subsite. A number of previous gene expression profiling studies have also ignored the differences in diverse anatomical subsite (11,13,14), such as in GCA and GNCA. Therefore, a comparative analysis of the gene expression profiles of GCA and GNCA would provide more accurate and valuable information on GCA.

Key words: gastric cardia adenocarcinoma, gastric non-cardia adenocarcinoma, gene expression data, differentially-expressed genes, functional enrichment analysis, protein-protein interaction network
Based upon the gene expression data from Wang et al (15), the present study adopted functional enrichment analysis and protein-protein interaction (PPI) network analysis to obtain a greater understanding of the common pathogenesis of GCA and GNCA, as well as the unique molecular mechanisms underlying GCA and GNCA, which could facilitate the development of targeted strategies for early detection, prognosis, and therapy.

**Materials and methods**

**Gene expression data.** A gene expression dataset (access number GSE29272) (15) was downloaded from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/), and consisted of 62 GCA samples and 62 normal controls, as well as 72 GNCA samples and 72 normal controls. Gene expression levels were measured using the Affymetrix Human Genome U133A Array (Affymetrix Inc., Santa Clara, CA, USA).

**Pre-treatment and differential analysis.** Raw data were treated with the Robust Multi-array Analysis method from the Affy package (16). Values of probes mapping to a same Entrez gene ID were averaged as a final expression level for the specific gene. A total of 22,283 probes and 12,495 genes were obtained. Differential analysis was performed with the Linear Models for Microarray Data (17) in R to identify DEGs in GCA and GNCA. The cut-offs were set as log (fold change) > 1 and a P-value of < 0.01. Overlapping DEGs of GCA and GNCA, as well as unique DEGs, were further selected out.

**Functional enrichment analysis.** Gene Ontology (GO; http://www.geneontology.org/) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg) pathway enrichment analysis were applied on the overlapping DEGs and unique DEGs using the Database for Annotation, Visualization and Integration Discovery (http://david.abcc.ncifcrf.gov/) online tool (18). P < 0.05 was set as the threshold.
Construction of the PPI network. A PPI network was constructed for the overlapping DEGs using information from the Search Tool for the Retrieval of Interacting Genes (19). Interactions with a score of >0.4 were retained and then visualized by Cytoscape (20). The proteins in the network serve as the ‘nodes’, and each pairwise protein interaction is represented by an undirected link and the degree of a node corresponds to the number of interactions of a protein. Degree was calculated for each node. Hub genes were then selected out according to the degree.

Subnetworks were also identified by Cytoscape (20) and its plugin MCODE (21), on which functional enrichment analysis was then applied.

Results

Differentially-expressed genes (DEGs). Gene expression data of GCA and GNCA prior to and after normalization are shown in Figs. 1 and 2. A good performance of normalization was achieved.

A total of 284 DEGs, 151 upregulated and 133 downregulated, were identified in GCA, while 268 DEGs, 126 upregulated and 142 downregulated, were revealed in GNCA. A total of 194 DEGs, 90 upregulated and 104 downregulated, were common between GCA and GNCA (Fig. 3).

Functional enrichment analysis result. Functional enrichment analysis was performed for the overlapping DEGs and unique DEGs. As shown in Table I, upregulated overlapping DEGs were involved in cell adhesion, the response to wounding and the regulation of cell proliferation. Extracellular matrix (ECM)-receptor interaction and focal adhesion were significantly over-represented. As for downregulated overlapping DEGs, digestion, oxidation reduction and the homeostatic process were enriched. The DEGs were associated with the metabolism of xenobiotics by cytochrome P450 and nitrogen metabolism.

GCA-unique DEGs were closely associated with the cell cycle and the regulation of cell proliferation (Table II).
Pathways such as the cell cycle, the p53 signaling pathway and the Toll-like receptor signaling pathway were enriched (Table II).

GNCA-unique DEGs were implicated in the regulation of angiogenesis, cell death and small GTPase-mediated signal transduction (Table III). Pathways such as focal adhesion and vascular smooth muscle contraction were enriched (Table III).

**PPI network.** A PPI network was constructed for the overlapping DEGs (Fig. 4), including 141 nodes and 446 edges.

Subnetwork 1 consisting of nodes with a degree >10 were extracted from the whole network (Fig. 5). The subnetwork contained 14 nodes and 75 edges. Three hub genes with a degree >25 were identified: Fibronectin 1 (FN1), collagen type I α2 (COL1A2) and COL1A1. The degrees of connectivity were 38, 28 and 26, respectively. The 3 hub genes interacted with each other and their neighboring nodes were selected as subnetwork 2, which included 45 nodes and 245 edges (Fig. 5).

Functional enrichment analysis was performed for the genes in the two subnetworks (Tables IV and V). GO enrichment analysis showed that the genes in subnetwork 1 were associated with collagen fibril organization, ECM organization and cell adhesion (Table IV). ECM-receptor interaction and focal adhesion were significantly enriched (Table IV). As for genes from subnetwork 2, they were involved in ECM organization, cell adhesion and extracellular structure organization (Table V). ECM-receptor interaction, focal adhesion, and complement and coagulation cascades were enriched (Table V).

**Discussion**

A comparative analysis of gene expression data of GCA and GNCA was performed in the present study. A total of 284 DEGs were identified in GCA, while 268 DEGs were revealed in GNCA. As many as 194 DEGs were shared by GCA and GNCA, while 90 DEGs were unique in GCA.

Upregulated common DEGs were involved in cell adhesion and the regulation of cell proliferation. Several members of the claudin family were revealed, including claudin 7 (CLDN7), CLDN4 and CLDN3. CLDN7 expression is an early event in gastric tumorigenesis (22). Zavala-Zendejas et al also reported that the overexpression of CLDN7 in the human gastric adenocarcinoma AGS cell line increased its invasiveness, migration and proliferation rate (23). CLDN4 and CLDN3 may also play roles in the pathogenesis of gastric cancer. Cadherin 11 (CDH11) and CDH17 were also found to be upregulated in GCA and GNCA in the present study. CDH17 is reported as a prognostic marker in early-stage gastric cancer (11). Zhang et al blocked the proliferation and migration of gastric cancer via targeting CDH17 with an artificial microRNA (24), suggesting that CDH17 is a potential target for the control of gastric cancer progression. In the present study, downregulated common DEGs were associated with digestion and metabolism, suggesting that the function of the stomach was impaired by the cancer. Several enzymes were included in the affected list, such as progastricsin, calpain 9 and aldo-keto reductase family 1 member C2. Mucin 6 (MUC6) and MUC5AC, playing essential roles in epithelial cytoprotection from acids and proteases, were also downregulated.

In the present study, a PPI network, including 141 nodes and 446 edges, was constructed for the common DEGs, from which two subnetworks were disclosed. Genes from the two subnetworks were associated with cell adhesion and ECM organization. The ECM provides a microenvironment for cell proliferation, cell adhesion and cell motion, and thus plays a critical role in cancer development (25) and metastasis (26). FN1, COL1A2 and COL1A1 were the top 3 hub genes in the whole network. FN1 is involved in cell adhesion and migration processes such as metastasis. David et al found that fibronectin expression is significantly associated with the expanding growth pattern of the gastric cancer (27). Yang et al found that Twist regulates cell motility and invasion in gastric cancer cell lines through N-cadherin and fibronectin production (28).

GCA-unique DEGs were implicated in the cell cycle and the regulation of cell proliferation in the present study. MADI mitotic arrest deficient-like 1 (MAD1L1) is a component of the mitotic spindle-assembly checkpoint that prevents the onset of anaphase until all chromosomes are properly aligned at the metaphase plate, and it may play a role in cell cycle control and tumor suppression. Coe et al report that MAD1L1 is the most frequent copy number gain in small cell lung cancer cell lines (29). Guo et al further suggested that genetic variants in MAD1L1 and MAD2L1 confer a susceptibility to lung cancer, which may result from reduced spindle checkpoint function due to the attenuated function of MAD1L1 and/or MAD2L1 (30). We supposed that MAD1L1 had a similar role in the pathogenesis of GCA. Three members of the cyclin family, CCNB1, CCNB2 and CCNE1, were also found to be upregulated in GCA. The B-type cyclins, CCNB1 and CCNB2, associate with p34cdc2 and are essential components of the cell cycle regulatory machinery. Begnami et al found that the expression of cyclin B1 is associated with regional lymph node metastasis and a poor prognosis in gastric cancer (31).

GNCA-unique DEGs were found to be associated with cell death in the present study. Abnormalities in cell death are closely associated with tumorigenesis. The GATA family
Figure 4. Protein-protein interaction network for the common differentially-expressed genes. Upregulated genes are in red, while downregulated genes are in green.

Figure 5. Two subnetworks extracted from the whole protein-protein interaction network.
of transcription factors participates in gastrointestinal development. GATA binding protein 6 (GATA6) is shown to be expressed in gastric cancer; it activates the expression of gastrointestinal protective trefoil genes, trefoil factor 1 (TFF1) and TFF2 (32). Haveri et al further reported that GATA6 expression is altered in neoplastic human gastrointestinal mucosa (33), suggesting that it may play a role in tumor progression. Hyaluronoglucosaminidase 1 (HYAL1) is a lysosomal hyaluronidase. Hyaluronan is believed to be involved in cell proliferation, migration and differentiation. HYAL1 is suggested to exhibit prognostic potential in prostate cancer (34). The study by Lokeshwar et al showed that, depending on the concentration,

| Group         | Category          | Term                                                                 | Count | P-value       |
|---------------|-------------------|----------------------------------------------------------------------|-------|---------------|
| Upregulated   | GOTERM_BP_FAT     | GO:0007155, cell adhesion                                            | 25    | 3.75x10^-13  |
|               | GOTERM_BP_FAT     | GO:0022610, biological adhesion                                       | 25    | 3.86x10^-13  |
|               | GOTERM_BP_FAT     | GO:0001501, skeletal system development                                | 12    | 1.76x10^-6   |
|               | GOTERM_BP_FAT     | GO:0009611, response to wounding                                       | 14    | 8.15x10^-6   |
|               | GOTERM_BP_FAT     | GO:0042127, regulation of cell proliferation                           | 13    | 1.54x10^-3   |
|               | GOTERM_CC_FAT     | GO:0044421, extracellular region part                                  | 40    | 1.83x10^-13  |
|               | GOTERM_CC_FAT     | GO:0005576, extracellular region                                       | 48    | 6.77x10^-10  |
|               | GOTERM_CC_FAT     | GO:0005578, proteinaceous ECM                                          | 23    | 1.45x10^-7   |
|               | GOTERM_CC_FAT     | GO:0031012, ECM                                                        | 23    | 7.16x10^-7   |
|               | GOTERM_CC_FAT     | GO:0005615, extracellular space                                       | 24    | 1.16x10^-11  |
|               | GOTERM_MF_FAT     | GO:0005201, ECM structural constituent                                 | 12    | 1.03x10^-12  |
|               | GOTERM_MF_FAT     | GO:0005539, glycosaminoglycan binding                                  | 12    | 2.27x10^-10  |
|               | GOTERM_MF_FAT     | GO:0001871, pattern binding                                            | 12    | 6.34x10^-10  |
|               | GOTERM_MF_FAT     | GO:0005198, structural molecule activity                               | 18    | 3.61x10^-8   |
|               | GOTERM_MF_FAT     | GO:0005509, calcium ion binding                                        | 14    | 1.28x10^-3   |
|               | KEGG_PATHWAY     | hsa04512: ECM-receptor interaction                                    | 13    | 5.02x10^-13  |
|               | KEGG_PATHWAY     | hsa04510: Focal adhesion                                              | 13    | 1.60x10^-8   |
|               | KEGG_PATHWAY     | hsa04670: Leukocyte transendothelial migration                          | 5     | 1.21x10^-2   |
|               | KEGG_PATHWAY     | hsa04514: Cell adhesion molecules                                     | 5     | 1.77x10^-2   |
| Downregulated | GOTERM_BP_FAT     | GO:0007586, digestion                                                  | 14    | 1.39x10^-14  |
|               | GOTERM_BP_FAT     | GO:0055114, oxidation reduction                                        | 15    | 4.63x10^-5   |
|               | GOTERM_BP_FAT     | GO:0010035, response to inorganic substance                             | 8     | 3.23x10^-4   |
|               | GOTERM_BP_FAT     | GO:0010033, response to organic substance                              | 11    | 1.59x10^-2   |
|               | GOTERM_BP_FAT     | GO:0042592, homeostatic process                                        | 11    | 2.05x10^-2   |
|               | GOTERM_CC_FAT     | GO:0005576, extracellular region                                       | 36    | 1.54x10^-4   |
|               | GOTERM_CC_FAT     | GO:0005615, extracellular space                                       | 16    | 4.01x10^-3   |
|               | GOTERM_CC_FAT     | GO:0044421, extracellular region part                                  | 19    | 4.81x10^-3   |
|               | GOTERM_CC_FAT     | GO:0045177, apical part of cell                                        | 6     | 6.78x10^-3   |
|               | GOTERM_CC_FAT     | GO:0016324, apical plasma membrane                                     | 5     | 1.19x10^-2   |
|               | GOTERM_MF_FAT     | GO:0008289, lipid binding                                              | 12    | 2.71x10^-4   |
|               | GOTERM_MF_FAT     | GO:0048037, cofactor binding                                           | 8     | 1.70x10^-3   |
|               | GOTERM_MF_FAT     | GO:0004175, endopeptidase activity                                      | 8     | 1.52x10^-2   |
|               | GOTERM_MF_FAT     | GO:0070011, peptidase activity, acting on L-amino acid peptides        | 9     | 3.64x10^-4   |
|               | GOTERM_MF_FAT     | GO:0008233, peptidase activity                                         | 9     | 4.53x10^-2   |
|               | KEGG_PATHWAY     | hsa00980: Metabolism of xenobiotics by cytochrome P450                 | 5     | 6.98x10^-4   |
|               | KEGG_PATHWAY     | hsa00982: Drug metabolism                                             | 4     | 8.56x10^-3   |
|               | KEGG_PATHWAY     | hsa00910: Nitrogen metabolism                                          | 3     | 1.06x10^-2   |

GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes; ECM, extracellular matrix.
HYAL1 functions as a tumor promoter or as a suppressor in prostate cancer (35). We speculated that it may exert a similar function in the progression of GNCA.

Overall, in the present study, a number of common DEGs were identified in GCA and GNCA, which could advance our understanding of the etiology of gastric cancer. Furthermore,
DEGs unique to GCA and to GNCA were identified, which supplemented our knowledge on the pathogenetic mechanisms involved and provided potential biomarkers for the diagnosis, prognosis or treatment of the disease.

References

1. Fock KM and Ang TL: Epidemiology of Helicobacter pylori infection and gastric cancer in Asia. J Gastroenterol Hepatol 25: 479-486, 2010.
2. Plummer M, Franceschi S, Vignat J, Forman D and de Martel C: Global burden of gastric cancer attributable to Helicobacter pylori. Int J Cancer 136: 487-490, 2015.

3. Wilkinson N: Management of gastric cancer. Surgical Oncology Clinics. Saunders, Philadelphia, PA, 2012.

4. Abnet CC, Freedman ND, Hu N, Wang Z, Yu K, Shu XO, Yuan JM, Zheng W, Dawsey SM, Dong LM, et al: A shared susceptibility locus in PCLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nat Genet 42: 764-767, 2010.

5. Gleeson CM, Sloan JM, McManus DT, Maxwell P, Arthur K, McGuigan JA, Ritchie AJ and Russell SE: Comparison of p53 and DNA content abnormalities in adenocarcinoma of the esophagus and gastric cardia. Br J Cancer 77: 277-286, 1998.

6. Kamangar F, Dawsey SM, Blaser MJ, Perez-Perez GI, Pietenpol NM, Neubacher CJ, Abnet CC, Albanes D, Virtamo J and Taylor PK: Opposing risks of gastric cardia and noncardia gastric adenocarcinomas associated with Helicobacter pylori seropositivity. J Natl Cancer Inst 98: 1445-1452, 2006.

7. Kim MA, Lee HS, Yang HK and Kim WH: Clinicopathologic and protein expression differences between cardia carcinoma and noncardia carcinoma of the stomach. Cancer 103: 1439-1446, 2005.

8. Ziozgas D, Baltogiannis G, Fatouros M and Roukos DH: Identifying and preventing high-risk gastric cancer individuals with CDH1 mutations. Ann Surg 247: 714-715; author reply 715-716, 2008.

9. Norton JA, Ham CM, Van Dam J, Jeffrey RB, Longacre TA, Huntsman DG, Chun N, Kurian AW and Ford JM: CDH1 truncating mutations in the E-cadherin gene: An indication for total gastrectomy to treat hereditary diffuse gastric cancer. Ann Surg 245: 873-879, 2007.

10. Yu J, Tao Q, Cheng YY, Lee KY, Ng SS, Cheung KF, Tian L, Rha SY, Neumann U, Röcken C, Kim MA, Lafleur BJ, Aburatani H, Lee PH, Yang CY, Kuo ML, Chang KJ, Song Y, Hyland C, et al: Comparison of global gene expression patterns in gastric cancer cells. J Proteome Res 5: 2727-2742, 2006.

11. Chen CN, Lin JJ, Chen JJ, Lee PH, Yang CY, Kuo ML, Chang KJ and Hsieh FJ: Gene expression profile predicts patient survival of gastric cancer after surgical resection. J Clin Oncol 23: 7286-7295, 2005.

12. Wang G, Hu N, Yang HH, Wang L, Su H, Wang C, Clifford R, Dawsey EM, Li JM, Ding T, et al: Comparison of global gene expression of gastric cardia and noncardia cancers from a high-risk population in China. PLoS One 8: e63826, 2013.

13. Irlizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP: Exploration, normalization and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249-264, 2003.

14. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article 3, 2004.

15. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Royoayi J, Stephens R, Baseel MW, Lane HC and Lempeicki RA: The DAVID gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 8: R183, 2007.

16. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C and Jensen LJ: STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41: D808-D815, 2013.

17. Kohl M, Wiese S and Warscheid B: Cytoscape: Software for visualization and analysis of biological networks. Methods Mol Biol 696: 291-303, 2011.

18. Bader GD and Hogue CW: An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4: 2, 2003.

19. Johnson AH, Friher HF, Zaika A, Powell SM, Roche J, Crowe S, Moskaluk CA and El-Rifai W: Expression of tight-junction protein claudin-7 is an early event in gastric tumorigenesis. Am J Pathol 167: 577-584, 2005.

20. Zavala-Zendejas VE, Torres-Martinez AC, Salas-Morales B, Fortoul TI, Montaño LF and Rendon-Huerta EP: Claudin-6, 7, or 9 overexpression in the human gastric adenocarcinoma cell line AGS increases its invasiveness, migration, and proliferation rate. Cancer Invest 29: 1-11, 2011.

21. Zhang J, Liu QS and Dong WG: Blockade of proliferation and migration of gastric cancer via targeting CDH17 with an artificial microRNA. Med Oncol 28: 494-501, 2011.