GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Bioinformatics analysis with graph-based clustering to detect gastric cancer-related pathways

P. Liu[1], X. Wang[2], C.H. Hu[1] and T.H. Hu[2]

[1]Department of Oncology, Second Xiangya Hospital, Central South University, Changsha, Hunan, China
[2]Department of Thoracic Surgery, Second Xiangya Hospital, Central South University, Changsha, Hunan, China

Corresponding author: C.H. Hu
E-mail: huchunhongdr@gmail.com

Genet. Mol. Res. 11 (3): 3497-3504 (2012)
Received November 8, 2011
Accepted May 9, 2012
Published September 26, 2012
DOI http://dx.doi.org/10.4238/2012.September.26.5

ABSTRACT. Despite a dramatic reduction in incidence and mortality rates, gastric cancer still remains one of the most common malignant tumors worldwide, especially in China. We sought to identify a set of discriminating genes that could be used for characterization and prediction of response to gastric cancer. Using bioinformatics analysis, two gastric cancer datasets, GSE19826 and GSE2685, were merged to find novel target genes and domains to explain pathogenesis; we selected differentially expressed genes in these two datasets and analyzed their correlation in order to construct a network. This network was examined to find graph clusters and related significant pathways. We found that ALDH2 and CCNB1 were associated with gastric cancer. We also mined for the underlying molecular mechanisms involving these differently expressed genes. We found that ECM-receptor interaction, focal adhesion, and cell cycle were among the significantly associated pathways. We were able to detect genes and pathways that were not considered in previous research on

gastric cancer, indicating that this approach could be an improvement on the investigative mechanisms for finding genetic associations with disease.

Key words: Graph cluster; Gastric cancer; Bioinformatics analysis

INTRODUCTION

Despite its recent decline, gastric cancer still remains the second leading cause of cancer-related deaths worldwide, particularly prevalent in Asian countries (Cui et al., 2011a). Gastric cancers have been subdivided into two main subtypes based on histological appearance, mainly including the better differentiated intestinal-type and the poorly differentiated diffuse-type (Lauren, 1965; Stock and Otto, 2005; Crew and Neugut, 2006).

Transformation of a normal gastric epithelial cell to a malignant cell results from the accumulation of multiple gene abnormalities (Zheng et al., 2004). Currently, a number of molecular abnormalities have been identified, including the activation of oncogenes, inactivation of tumor suppressor genes, microsatellite and chromosomal instability, and alteration of growth factors and cytokines (Smith et al., 2006). Many proto-oncogenes are activated and overexpressed in gastric cancer, such as c-met (Inoue et al., 2004), c-erB2 (Song et al., 2004), K-sam (Toyokawa et al., 2009), Ras (Nishigaki et al., 2005), and c-myc (Calcagno et al., 2005). Inactivation of tumor suppressor genes due to mutations and/or loss of heterozygosity is also a frequent event in gastric carcinogenesis. For example, inactivation of p53 and p16 has been reported in both diffuse- and intestinal-type gastric cancers, whereas adenomatous polyposis coli and retinoic acid receptor β gene mutations seem to occur more frequently in intestinal-type gastric cancers. Besides, RUNX, fragile histidine triad, deleted in colon cancer, and trefoil factor family 1 are suggested as tumor suppressor candidate genes (Cervantes et al., 2007). Microsatellite has been recognized as one of the earliest changes in gastric carcinogenesis and results in genomic instability. Strong telomerase activity associated with human telomerase reverse transcriptase expression is also present in a majority of gastric carcinomas (Yasui et al., 2005). Numerous studies have demonstrated that many growth factors and their receptors are overexpressed in gastric cancer (Yokozaki et al., 2001), including the epidermal growth factor family, vascular endothelial growth factor, fibroblast growth factor, and insulin-like growth factor (Pavelic et al., 2003), and numerous cytokines, such as TGF, IL-1, and IL-18 (Kim et al., 2006).

Bioinformatics analysis provides a powerful tool for analyzing microarray experiments by combining data from multiple studies, and presents unique computational challenges. The Bioconductor package RankProd provides a new and intuitive tool for this purpose in detecting differentially expressed genes (DEGs) under two experimental conditions (Hong et al., 2006).

In this study, we used bioinformatics analysis to detect the DEGs and then used the graph clustering approach to further identify gene expression profiles that distinguish gastric tumors from normal gastric samples. In addition, the relevant pathways in the cluster were also analyzed to explain potential mechanisms in response to gastric cancer.

MATERIAL AND METHODS

Bioinformatics analysis for the expression profiles and DEG analysis

Two gastric cancer-related expression profiles GSE19826 and GSE2685 were obtained from a public functional genomic data repository GEO (http://www.ncbi.nlm.nih.gov/geo/), which are based on the Affymetrix Human Genome U133 Plus 2.0 Array and the Affymetrix Human Full-Length HuGeneFL Array, separately. In this study, a total 34 gastric tumors and 20 controls were selected to identify DEGs. For detail descriptions, see Table 1.

Table 1. Description of two gastric datasets.

Dataset	Gastric tumors	Normal samples	Platform
GSE19826	12	12	Affymetrix Human Genome U133 Plus 2.0 Array
GSE2685	22	8	Affymetrix Human Full-Length HuGeneFL Array
Total	34	20	

Statistical analysis

For the GSE19826 and GSE2685 datasets, the RankProd package was used to identify DEGs. The DEGs only with a percentage of false-positives (PFP) ≤ 0.05 (Hong et al., 2006) were considered to be differentially expressed between treatments and controls.

For demonstrating the potential relationship between DEGs, Spearman's rank correlation (r) was used for comparative target gene correlations. The significance level was set at $r > 0.9$. All statistical tests were performed with the R program (http://www.r-project.org/).

Network analyses and graph clustering

To identify co-expressed groups, we used DPClus (Altaf-Ul-Amin et al., 2006), a graph clustering algorithm that can extract densely connected nodes as a cluster. It is based on density and periphery tracking of clusters. DPClus is freely available from http://kanaya.naist.jp/DPClus/. In this study, we used the overlapping-mode with the DPClus settings. We set the parameter settings of cluster property (cp); density values were set at 0.5 (Fukushima et al., 2011).

Pathway enrichment analysis

The PATHWAY (Kanehisa, 2002) database records networks of molecular interactions in cells and their variants specific to particular organisms (http://www.genome.jp/kegg/) was used.

DAVID (Huang et al., 2009) was used to identify over-represented pathways. The $P < 0.05$ was considered to be significant.

RESULTS

Selection of DEGs and a correlation network construction

To get DEGs of gastric cancer, we searched GEO and obtained two publicly available

microarray data sets, GSE19826 and GSE2685. After microarray analysis, the DEGs with a fold-change >1.5 and $P < 0.05$ were selected. A total of 4530 genes from GSE19826 and 1723 genes from GSE2685 were selected as DEGs. Using the RankProd packages for meta-analysis, 303 upregulated genes and 439 downregulated genes with a PFP ≤ 0.05 were considered to be differentially expressed. Finally, 742 DEGs were collected after meta-analysis and the expression values of DEGs are displayed in Figure 1. We then quantified the correlation values between DEGs, and the co-expressed value $r > 0.9$ and $P < 0.05$ were selected as the threshold. Finally, 1163 relationships between 248 DEGs were used to construct a correlation network (Figure 1).

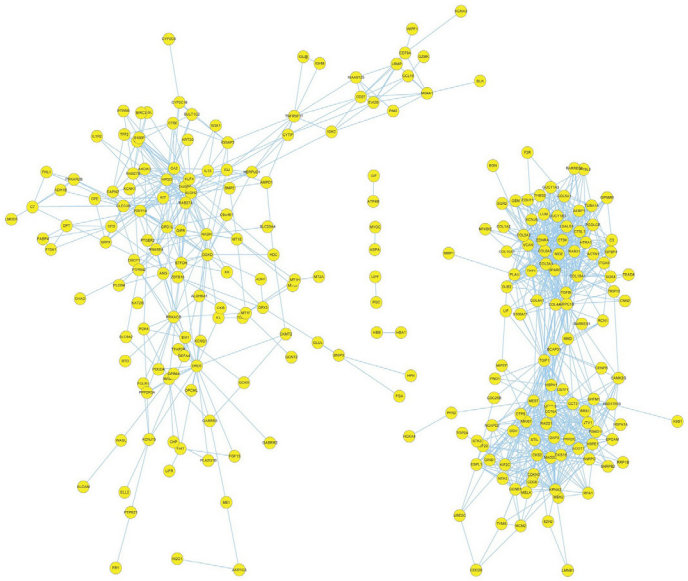

Figure 1. Correlation network of gastric cancer. Yellow points stand for DEG and blue lines stand for the correlation of two neighbor points with the $r > 0.9$.

Graph clustering identifies modules significantly enriched in biochemical pathways

At $r > 0.9$, DPClus (Altaf-Ul-Amin et al., 2006) identified 4 clusters in the correlation network (Figure 1) for response to gastric cancer; they ranged in size from 13 to 32 genes. Especially clusters 1, 2 and 4 had a connection between each other. Graph clustering results are presented in Figure 2. To assess the significance of the clusters, we used the over-represented KEGG pathways (so-called KEGG enrichment analysis) in the clusters. The results of graph clustering with KEGG enrichment analysis are presented in Table 2.

Significant pathways ($P < 0.05$ using hypergeometric test) were related to ECM-receptor interaction, cell cycle, butanoate metabolism, and so on (Table 2). Among them, COL6A3, COL3A1, and COL5A2 were enriched for the ECM-receptor interaction and focal adhesion pathways; CCNB1, MAD2L1 and ESPL1 were overexpressed in the cell cycle and oocyte meiosis pathways; additionally, ALDH2 and HADH were upregulated in several pathways, such as: the butanoate metabolism, fatty acid metabolism, and so on (not displayed in Table 2).

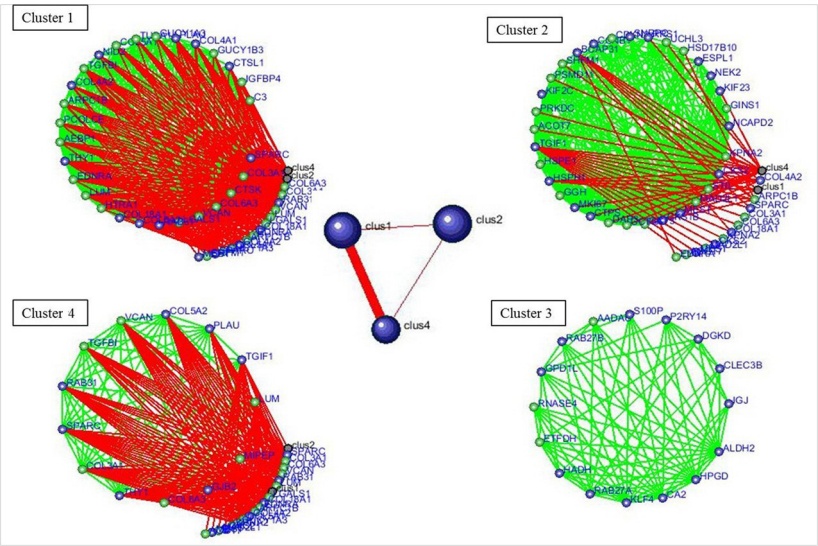

Figure 2. Graph clustering of correlated modules in gastric cancer (threshold r ≥ 0.9).

Table 2. List of enriched KEGG pathways in clusters 1 and 5 detected by DPClus.

Category	Term ID	Term	P	FDR
Cluster 1	hsa04512	ECM-receptor interaction	5.78E-06	0.004398
	hsa04510	Focal adhesion	3.85E-04	0.292749
	hsa04540	Gap junction	0.034741	23.5945
	hsa04270	Vascular smooth muscle contraction	0.05273	33.78681
Cluster 2	hsa04110	Cell cycle	0.002071	1.431166
	hsa04114	Oocyte meiosis	0.022454	14.60601
Cluster 3	hsa00650	Butanoate metabolism	0.033	22.23604
	hsa00071	Fatty acid metabolism	0.038732	25.62488
	hsa00380	Tryptophan metabolism	0.038732	25.62488
	hsa00280	Valine, leucine and isoleucine degradation	0.042539	27.80379
	hsa00310	Lysine degradation	0.042539	27.80379
	hsa00561	Glycerolipid metabolism	0.043488	28.33874
Cluster 4	hsa04512	ECM-receptor interaction	0.003874	1.84068
	hsa04510	Focal adhesion	0.020994	9.656048

FDR = false discovery rate; ECM = extracellular matrix.

DISCUSSION

According to our analysis results, we could find that many targets and pathways closely related to gastric cancer were linked by our method. Among them, ALDH2, CCNB1, ECM-receptor and cell cycle pathways displayed higher degrees (Figures 1 and 2) in the clusters, suggesting that these genes may play more important roles in gastric cancer. We discuss the relationships between gastric cancer and identified genes as follows based on previous reports.

ALDH2 showed 22 degrees in the correlation network, but only displayed in the cluster 3. The ALDH2 protein belongs to a low-Km mitochondrial ALDH, which is the second enzyme to eliminate most of the acetaldehyde generated during alcohol metabolism *in vivo*. The increased exposure to acetaldehyde in individuals with the catalytically inactive form may

confer greater susceptibility to many types of cancer. ALDH2 polymorphisms were found to modify the susceptibility to the development of gastric cancer associated with alcohol intake, especially in the case of ALDH2 *1/*2 genotype in the Korean population (Shin et al., 2011). These findings have suggested an alcohol-ALDH2 genotype interaction in gastric carcinogenesis. However, ALDH2 polymorphisms do not seem to play an important role in the development of stomach cancer in Chinese males (Cao et al., 2010).

CCNB1 revealed 16 degrees in the correlation network (Figure 1), but only displayed in cluster 2. CCNB1 is a regulatory protein involved in mitosis. The gene product complexes with p34(cdc2) to form the maturation-promoting factor (MPF). There are two alternative transcripts, a constitutively expressed transcript and a cell cycle-regulated transcript, which is expressed predominantly during G2/M phase. Cell cycle dysregulation plays a major role in gastric carcinogenesis. The study showed that CCNB1 were frequently expressed in diffuse gastric cancer. In addition, the expression of CCNB1 was associated with regional lymph node metastasis and poor prognosis. These findings suggested that CCNB1 expression is closely associated with poor outcome in gastric cancer (Kim, 2007; Begnami et al., 2010).

To identify the relevant altered pathways, we used the hypergeometric distribution approach at the pathway level. Pathways can provide an alternative way to relax the significance threshold applied to single genes and may lead to a better biological interpretation. Finally, using the pathway-based significant analysis method, only ECM-receptor, focal adhesion, and cell cycle pathways were identified as the significant ones.

There is a substantial amount of evidence that the ECM-receptor pathway is related to the progression of gastric cancer (Cui et al., 2011b). For instance, the ECM receptor integrin (αvβ1) shows strong expression in gastric carcinoma tissues, which may be involved in the lymphatic metastasis of gastric carcinomas (Kawashima et al., 2003). In addition, ECM-degrading enzymes such as MMP-7 are also believed to play a significant role in invasion and metastasis of gastric cancer. MMP-7-positive tumors have been found to be significantly more frequent in diffuse-type lesions. MMP-7-positive lesions increase with the progression of gastric epithelial tumors, including adenomas, mucosal cancers, and cancers invading the submucosal layer or deeper (Kitoh et al., 2004; Ii et al., 2006).

There is also evidence that the focal adhesion pathway plays an important role in gastric cancer. The focal adhesion complex contains an interacting matrix of numerous proteins, which includes non-receptor tyrosine kinases, such as Src and focal adhesion kinase (FAK), and adaptor and actin-binding proteins, such as talin and paxillin, as well as cytosolic phosphatases and proteases. In particular, the calpain proteases have been implicated in the cleavage of focal adhesion proteins, which promotes focal adhesion turnover (Barbero et al., 2009). Wnt-5a was identified to have the abilities to stimulate cell migration and invasion in gastric cancer by regulating the turnover of focal adhesion complexes (Kurayoshi et al., 2006). Besides, focal adhesion kinase has been shown to be positively amplified in gastric cancer tissues and significantly correlated with gastric cancer progression (Park et al., 2010). PTEN inhibits focal adhesion, spreading, and migration by dephosphorylating focal adhesion kinase in gastric cancer (Kang et al., 2002).

It is widely believed that the cell cycle plays a crucial role in gastric progression (Smith et al., 2006). The cell cycle regulator, cyclin G2, has been found to be positively expressed in 66.3% human gastric cancer tissues. Moreover, cyclin G2 expression has been shown to be inversely correlated with the more advanced stages, the presence of lymph node

metastasis, and the presence of perineural invasion (Choi et al., 2009). Furthermore, cyclin-dependent kinase 6 (CDK6) has been found to be upregulated in gastric cancer tissues. The expression of the CDK6 inhibitor miR-107, which binds to CDK6 3'-untranslated region could inhibit gastric cancer cell proliferation, induce G1 cell cycle arrest, and block invasion by gastric cancer cells. Reduced miR-107 expression correlates with tumor invasion and nodal metastasis (Feng et al., 2012).

In conclusion, we used network analysis as a conceptual framework to explore the pathobiology of gastric cancer based on the assumption that gastric cancer is a contextual attribute of distinct patterns of interactions between multiple genes. The salient result of our study is finding the ALDH2 and CCNB1 genes and ECM-receptor, focal adhesion, and cell cycle pathways differentially expressed in gastric cancer, which have all been related to gastric cancer in a direct or indirect manner. We anticipate numerous advances in gastric cancer research in the coming years based on our bioinformatics analysis.

ACKNOWLEDGMENTS

Research supported by the China Hunan Provincal Science and Technology Department (#2007fj3057), the Hunan Provincial Natural Foundation (#08jj6010) and the "125 Talent" Project of the Third Xiangya Hospital of Central South University.

REFERENCES

Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K, et al. (2006). Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinformatics* 7: 207.

Barbero S, Mielgo A, Torres V, Teitz T, et al. (2009). Caspase-8 association with the focal adhesion complex promotes tumor cell migration and metastasis. *Cancer Res.* 69: 3755-3763.

Begnami MD, Fregnani JH, Nonogaki S and Soares FA (2010). Evaluation of cell cycle protein expression in gastric cancer: cyclin B1 expression and its prognostic implication. *Hum. Pathol.* 41: 1120-1127.

Calcagno DQ, Leal MF, Taken SS, Assumpcao PP, et al. (2005). Aneuploidy of chromosome 8 and C-MYC amplification in individuals from northern Brazil with gastric adenocarcinoma. *Anticancer Res.* 25: 4069-4074.

Cao HX, Li SP, Wu JZ, Gao CM, et al. (2010). Alcohol dehydrogenase-2 and aldehyde dehydrogenase-2 genotypes, alcohol drinking and the risk for stomach cancer in Chinese males. *Asian Pac. J. Cancer Prev.* 11: 1073-1077.

Cervantes A, Rodriguez BE, Perez FA and Chirivella G, I (2007). Molecular biology of gastric cancer. *Clin. Transl. Oncol.* 9: 208-215.

Choi MG, Noh JH, An JY, Hong SK, et al. (2009). Expression levels of cyclin G2, but not cyclin E, correlate with gastric cancer progression. *J. Surg. Res.* 157: 168-174.

Crew KD and Neugut AI (2006). Epidemiology of gastric cancer. *World J. Gastroenterol.* 12: 354-362.

Cui J, Chen Y, Chou WC, Sun L, et al. (2011a). An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer. *Nucleic Acids Res.* 39: 1197-1207.

Cui J, Li F, Wang G, Fang X, et al. (2011b). Gene-expression signatures can distinguish gastric cancer grades and stages. *PLoS One* 6: e17819.

Feng L, Xie Y, Zhang H and Wu Y (2012). miR-107 targets cyclin-dependent kinase 6 expression, induces cell cycle G1 arrest and inhibits invasion in gastric cancer cells. *Med. Oncol.* 29: 856-863.

Fukushima A, Kusano M, Redestig H, Arita M, et al. (2011). Metabolomic correlation-network modules in *Arabidopsis* based on a graph-clustering approach. *BMC Syst. Biol.* 5: 1.

Hong F, Breitling R, McEntee CW, Wittner BS, et al. (2006). RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22: 2825-2827.

Huang dW, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.

Ii M, Yamamoto H, Adachi Y, Maruyama Y, et al. (2006). Role of matrix metalloproteinase-7 (matrilysin) in human

cancer invasion, apoptosis, growth, and angiogenesis. *Exp. Biol. Med.* 231: 20-27.
Inoue T, Kataoka H, Goto K, Nagaike K, et al. (2004). Activation of c-Met (hepatocyte growth factor receptor) in human gastric cancer tissue. *Cancer Sci.* 95: 803-808.
Kanehisa M (2002). The KEGG database. *Novartis Found. Symp.* 247: 91-101.
Kang YH, Lee HS and Kim WH (2002). Promoter methylation and silencing of PTEN in gastric carcinoma. *Lab. Invest.* 82: 285-291.
Kawashima A, Tsugawa S, Boku A, Kobayashi M, et al. (2003). Expression of alphav integrin family in gastric carcinomas: increased alphavbeta6 is associated with lymph node metastasis. *Pathol. Res. Pract.* 199: 57-64.
Kim DH (2007). Prognostic implications of cyclin B1, p34cdc2, p27(Kip1) and p53 expression in gastric cancer. *Yonsei Med. J.* 48: 694-700.
Kim J, Kim C, Kim TS, Bang SI, et al. (2006). IL-18 enhances thrombospondin-1 production in human gastric cancer via JNK pathway. *Biochem. Biophys. Res. Commun.* 344: 1284-1289.
Kitoh T, Yanai H, Saitoh Y, Nakamura Y, et al. (2004). Increased expression of matrix metalloproteinase-7 in invasive early gastric cancer. *J. Gastroenterol.* 39: 434-440.
Kurayoshi M, Oue N, Yamamoto H, Kishida M, et al. (2006). Expression of Wnt-5a is correlated with aggressiveness of gastric cancer by stimulating cell migration and invasion. *Cancer Res.* 66: 10439-10448.
Lauren P (1965). 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.* 64: 31-49.
Nishigaki M, Aoyagi K, Danjoh I, Fukaya M, et al. (2005). Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. *Cancer Res.* 65: 2115-2124.
Park JH, Lee BL, Yoon J, Kim J, et al. (2010). Focal adhesion kinase (FAK) gene amplification and its clinical implications in gastric cancer. *Hum. Pathol.* 41: 1664-1673.
Pavelic K, Kolak T, Kapitanovic S, Radosevic S, et al. (2003). Gastric cancer: the role of insulin-like growth factor 2 (IGF 2) and its receptors (IGF 1R and M6-P/IGF 2R). *J. Pathol.* 201: 430-438.
Shin CM, Kim N, Cho SI, Kim JS, et al. (2011). Association between alcohol intake and risk for gastric cancer with regard to ALDH2 genotype in the Korean population. *Int. J. Epidemiol.* 40: 1047-1055.
Smith MG, Hold GL, Tahara E and El-Omar EM (2006). Cellular and molecular aspects of gastric cancer. *World J. Gastroenterol.* 12: 2979-2990.
Song HS, Do YR, Kim IH, Sohn SS, et al. (2004). Prognostic significance of immunohistochemical expression of EGFR and C-erbB-2 oncoprotein in curatively resected gastric cancer. *Cancer Res. Treat.* 36: 240-245.
Stock M and Otto F (2005). Gene deregulation in gastric cancer. *Gene* 360: 1-19.
Toyokawa T, Yashiro M and Hirakawa K (2009). Co-expression of keratinocyte growth factor and K-sam is an independent prognostic factor in gastric carcinoma. *Oncol. Rep.* 21: 875-880.
Yasui W, Oue N, Aung PP, Matsumura S, et al. (2005). Molecular-pathological prognostic factors of gastric cancer: a review. *Gastric. Cancer* 8: 86-94.
Yokozaki H, Yasui W and Tahara E (2001). Genetic and epigenetic changes in stomach cancer. *Int. Rev. Cytol.* 204: 49-95.
Zheng L, Wang L, Ajani J and Xie K (2004). Molecular basis of gastric cancer development and progression. *Gastric Cancer* 7: 61-77.