Novel biomarker candidates for gastric cancer

SANGHWA YANG[1] and HYUN CHEOL CHUNG[1,2,3]

[1]Cancer Metastasis Research Center (CMRC), [2]Brain Korea 21 Project for Medical Sciences,
[3]Department of Internal Medicine, Yonsei University College of Medicine,
134 Shinchon-dong, Seodaemun-gu, Seoul 120-752, Korea

Received September 10, 2007; Accepted November 20, 2007

Abstract. Gastric cancer continues to be a major threat to human health. Molecular descriptions on the diverse phases of this disease will be valuable for a better diagnosis and development of therapeutic targets. Previously, a 92-gene classifier that distinguishes tumor from non-tumor gastric tissues was proposed. To corroborate this finding, independent approaches of gene selection and class prediction algorithm were applied to the dataset of 86 tissues profiled on 17K cDNA microarrays. As a result, 22 genes were selected, of which 18 were in common with 92 genes previously shown. The differential expression patterns of *Chromogranin A* (CHGA) and *Thy-1 cell surface antigen* (THY1) were further validated with immunohisto-staining on gastric tissue microarrays. The differential expression patterns of several of the proposed genes have been proven to be critical for tumor progression in other cancer models and will likely function as novel biomarkers for gastric cancer as well.

Introduction

Gastric cancer is a major cause of human cancer-related death and certain risk factors affecting the development of gastric cancer, including age, gender, race, diet, *Helicobacter pylori* infection and clinicopathological parameters are well characterized (1-3). However, the molecular mechanisms underlying this specific cancer are still not clear. Even though ongoing cancer drug trials have shown signs of benefits (4), identification of molecular markers in the diverse phases of gastric cancer will continuously be needed to supplement the mostly histopathology-based diagnosis and for the development of therapeutic tools.

Correspondence to: Drs Sanghwa Yang and Hyun Cheol Chung, Cancer Metastasis Research Center (CMRC), Yonsei University College of Medicine, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-752, Korea
E-mail: ysh@yumc.yonsei.ac.kr; unchung8@yumc.yonsei.ac.kr

Key words: gastric cancer, DNA microarray, tissue array, gene expression profiling, biomarker, immunohistostaining

Recent advancements in genomics, epigenomics and proteomics-based high throughput screening technologies are being actively applied to the dissection of gastric cancer to elucidate the molecular nature of the disease (5-8). A recent report on the gene expression profiling of 86 gastric tissues on 17K cDNA microarray led to the identification of 92 genes that showed a significant difference in expression levels between non-tumorous and tumor tissues (9). In the current study, in order to narrow down the list and corroborate the previous results, independent gene selection and class prediction algorithm were employed to select for the 22-gene classifier. Immunostaining of two of the genes, CHGA and THY1, on tissue microarrays confirms the relative expression level of the selected genes at the protein level.

Materials and methods

Tissue samples, RNA extraction and cDNA microarray hybridization. Tissue samples, RNA extraction and hybridization on 17K cDNA microarrays were detailed in a recent publication (9). Tissues used for the profiling experiments are composed of 29 pairs (29 normal tissues and 29 matching tumor tissues, n=58) and a test set of 28 samples (7 pairs, 8 non-paired normal tissues and 6 non-paired 72 tumor tissues). The microarray hybridizations were performed in a reference RNA-based indirect-design, in which each of the cDNA targets generated from tissue total RNAs (Cy5-labeled) was competitively hybridized with common reference RNA-generated cDNAs labeled with Cy3. Hybridized slides were scanned with a Gene Pix 4000B laser scanner (Axon Instrument Inc, Union City, CA) and the raw data were saved in a Gene Pix Result (GPR) format.

Microarray data analysis. To exclude any possible bias due to the use of the k-nearest neighbor-based class prediction method used before, a different approach in class prediction implemented in the BRB ArrayTools developed by Dr Richard Simon and Amy Peng Lam was tested. The 17K cDNA microarray contains 15,723 gene probes. All the GPRs were imported into the BRB ArrayTools bioinformatics program. Probes with fluorescent intensities <100 in the Cy3 and Cy5 channels, or those containing a flag signal were removed. Probes with a spot diameter <10 or any missing values were also removed from further analysis in the filtering step. Print tip, lowess normalization was applied to

the entire dataset to adjust dye-dependent bias. Following the filtering step, 10,012 genes were left to be used for the analysis. In the class prediction function of BRB ArrayTools, gene selection was performed using a 'greedy pairs method' (10), which identifies pairs of genes that separates two classes (in this case, tumor and non-tumor), to select 11 pairs of genes. The prediction error (misclassification) rate in the training set was estimated by leave-one-out cross-validation.

Four different class prediction methods of the linear discriminant analysis, the 1-nearest neighbor, the 3-nearest neighbors and the nearest centroid were individually tested (refer to the BRB ArrayTool manual for descriptions and references for each of the methods). A statistical test of cross-validated misclassification rate was performed by 100 permutations. For each predictor, this test estimates a p-value for the global test of the hypothesis that the predictor is picking up the random noise in the data whereas classes do not differ at all with regard to expression profiles. A permutation analysis is used for the computation of the p-value for the global test. Class labels of the samples are randomly permuted 100 times. For each permutation, samples are classified and the cross-validated misclassification rate of each classifier is computed as a proportion of correctly predicted samples. The p-value of the predictor is the proportion of permutations with a misclassification rate smaller than the misclassification rate of the original labeling.

Table I. Sensitivity and specificity of the 22-gene classifier.

Class	Sensitivity	Specificity	PPV	NPV
Performance of the compound covariate predictor classifier				
Normal	1	1	1	1
Tumor	1	1	1	1
Performance of the linear diagonal discriminant analysis classifier				
Normal	1	1	1	1
Tumor	1	1	1	1
Performance of the nearest centroid classifier				
Normal	1	1	1	1
Tumor	1	1	1	1
Performance of the support vector machine classifier				
Normal	1	1	1	1
Tumor	1	1	1	1

PPV (positive predictive value) is the probability that a sample predicted as class A actually belongs to class A. NPV (negative predictive value) is the probability that a sample predicted as non class A actually does not belong to class A.

Based on 100 random permutations, classifiers (11 pairs or 22 genes) from each of the 4 prediction methods had a p-value of <0.01.

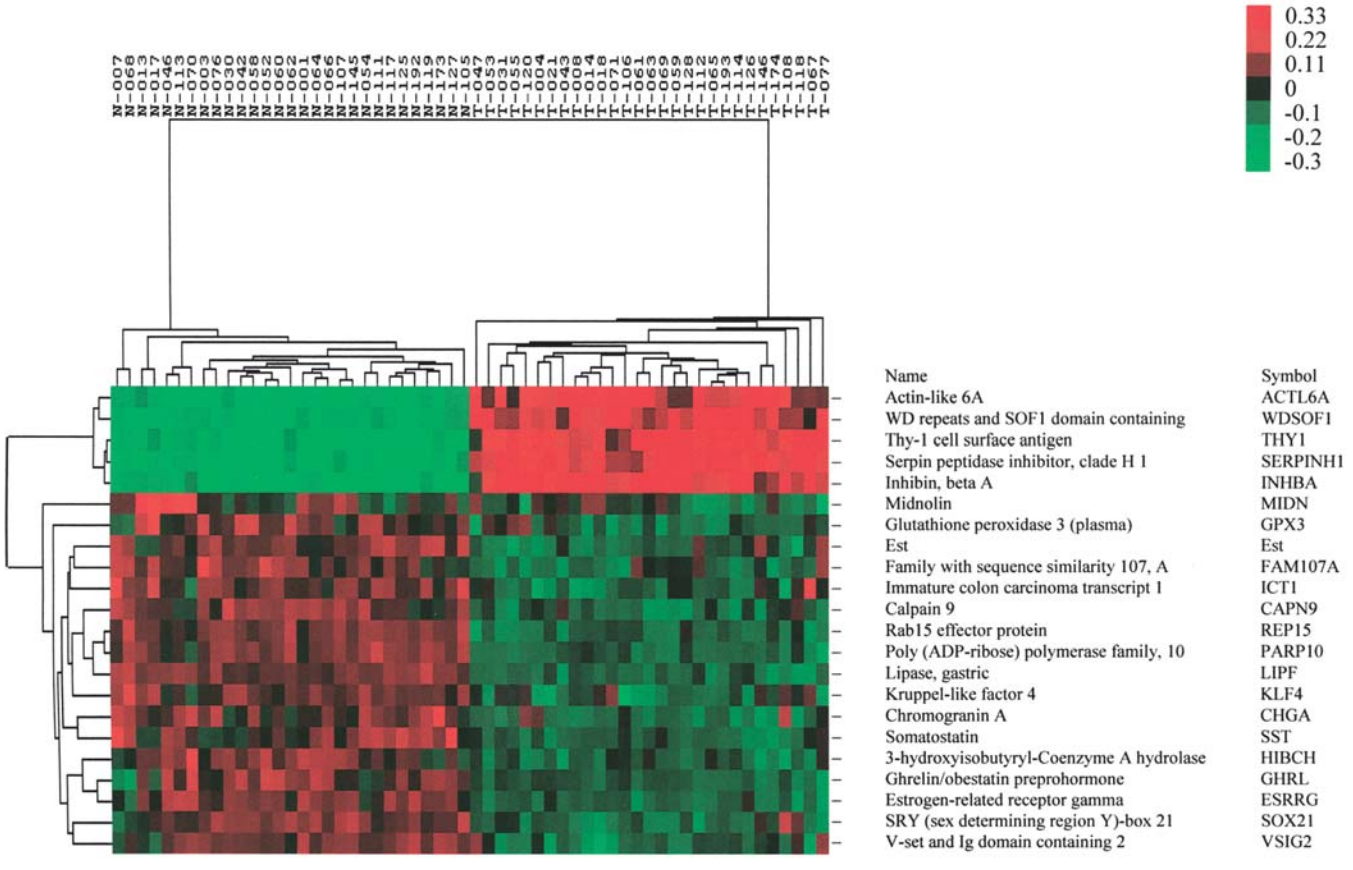

Figure 1. Heatmap representation of the 22-gene classifier. The log gene expression ratios for the 22 genes in the 58 tissues in the training set were subjected to a two-way hierarchical clustering and the result is in a Treeview format. Each of the tissue samples are labeled with a unique number preceded with N- (normal) or T- (tumor) labels. The red and green colors represent up-regulation and down-regulation, respectively and the relationship between the degree of color change and expression ratio variations are shown in the scale bar.

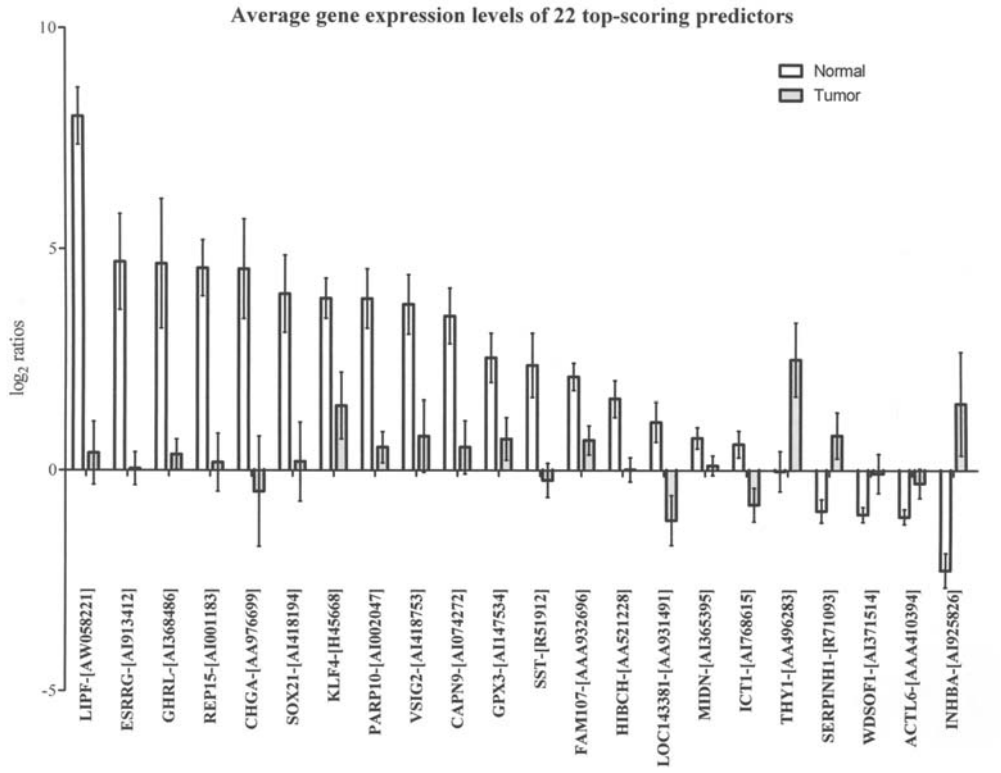

Figure 2. A graphic presentation of the average expression ratios for the 22 genes. A $\log_2$ expression ratio represents an average value from 29 samples of normal and tumor tissues, respectively.

Immunohistochemical staining on tissue microarrays. Mouse monoclonal antibodies against CHGA (1:500) (Dako, Denmark) and THY1 (Abcam Inc., Cambridge, MA) were used to validate the differential expression levels of these two genes at the protein level by immunostaining on tissue microarrays with provided protocol (Super Bio Chips, Seoul, Korea).

Results

Selection of the 11-pair class predictors. Twenty-two genes were selected that were differentially expressed between the gastric adenocarcinomas and adjacent, histologically normal gastric tissues. In the two-way (sample and genes) hierarchical clustering of the 58 tissues in the training set, the genes were clearly able to distinguish the two tissue types (Fig. 1). The 22-gene classifiers showed 100% efficiency in cross-validation in the training set and are highly sensitive (class A sample to be correctly predicted as class A) and specific (non class A sample to be correctly predicted as non-A) (Table I). These classifiers can also correctly predict the tissue types in the test set (Table II) as a group. Some of the selected genes are components of important biological pathways; *Ghrelin/obestatin preprohormone* (GHRL) and *Somatostatin* (SST) are members of the neuroactive ligand-receptor interaction pathway. *Inhibin, ß A* (INHBA) is a member of TGF-ß signaling pathway, whereas THY1 and INHBA are members of leukocyte transendothelial migration and cytokine-cytokine receptor interaction pathways, respectively.

Validation of differential gene expression by immunohistochemical staining on tissue microarrays. The relative expression ratios of 22 genes are shown in Fig. 2 and are detailed in Table III. The differential expression of two genes was further confirmed by immunostaining. CHGA shows about a 30-fold relative increase in gene expression ratios in normal gastric tissues. In the immunohistochemical staining on tissue microarrays with CHGA antibody, the over-expression in the normal tissue is confirmed (Fig. 3). CHGA shows high expression in each of the 59 normal gastric tissues, whereas it shows only weak expression in 5 tissues and relatively high expression in 2 of the tissues (data not shown). THY1 was barely observable in non-tumor tissues, whereas it shows both cytoplasmic and nuclear staining in most of the tumor tissues (data not shown).

Discussion

The purpose of this study was to identify a smaller number of genes that can be used as model genes in a biochemical study of gastric cancer at the molecular level. Additionally, these genes may become assets in the development of molecular probes in the diagnosis of gastric cancer. What is different from an earlier study was that independent gene selection and class prediction algorithms were employed in order to minimize statistical bias. The gene selection in this study is based on the selection of differentially expressed between the classes at a univariate parametric significance level of $p<0.001$. In a comparison with the 92 genes previously identified (9), 18 of the 22 genes are selected in common, the 4 different genes being *Actin-like 6A* (ACTL6A), *WD repeats*

Table II. Predictions of classifiers for new samples in the test set.

Array id in the test set	Compound covariate predictor	Linear discriminant analysis	Nearest centroid	Support vector machines
Y-GC-01-002	N	N	N	N
Y-GC-01-019	N	N	N	N
Y-GC-01-020	T	T	T	T
Y-GC-01-026	N	N	N	N
Y-GC-01-027	T	T	T	T
Y-GC-01-048	N	N	N	N
Y-GC-01-049	N	N	N	N
Y-GC-01-090	T	T	T	T
Y-GC-01-092	N	N	N	N
Y-GC-01-097	N	N	N	N
Y-GC-01-098	T	T	T	T
Y-GC-01-129	N	N	N	N
Y-GC-01-130	T	T	T	T
Y-GC-01-147	N	N	N	N
Y-GC-01-148	T	T	T	T
Y-GC-01-155	N	N	N	N
Y-GC-01-156	T	T	T	T
Y-GC-01-180	N	N	N	N
Y-GC-01-181	N	N	N	N
Y-GC-01-184	T	T	T	T
Y-GC-01-185	T	T	T	T
Y-GC-01-188	T	T	T	T
Y-GC-01-190	N	N	N	N
Y-GC-01-191	N	N	N	N
Y-GC-01-196	T	T	T	T
Y-GC-01-198	T	T	T	T
Y-GC-01-200	N	N	N	N
Y-GC-01-201	N	N	N	N

N, normal tissue; T, tumor tissue.

Table III. Information on the 22-gene classifier.

GB acc[d]	Description	Symbol	N[a]	T[b]	N/T[c]
AW058221	Lipase, gastric	LIPF	257.5	1.3	196.6
AI074272	Calpain 9	CAPN9	11.2	1.4	7.8
AI001183	Rab15 effector protein	REP15	23.8	1.1	21.0
AI913412	Estrogen-related receptor γ	ESRRG	26.2	1.0	25.6
AI002047	Poly (ADP-ribose) polymerase family, member 10	PARP10	14.7	1.4	10.2
AA410394	Actin-like 6A	ACTL6A	0.5	0.8	0.6
AA521228	3-hydroxyisobutyryl-Coenzyme A hydrolase	HIBCH	3.1	1.0	3.0
H45668	Kruppel-like factor 4	KLF4	14.9	2.8	5.4
R51912	Somatostatin	SST	5.2	0.9	6.0
AI147534	Glutathione peroxidase 3	GPX3	5.8	1.6	3.6
AI925826	Inhibin, ß A	INHBA	0.2	2.9	0.1
AA976699	Chromogranin A	CHGA	23.4	0.7	32.5
AI418194	SRY-box 21	SOX21	15.9	1.2	13.8
R71093	Serpin peptidase inhibitor, clade H, member 1	SERPINH1	0.5	1.7	0.3
AA931491	Hypothetical protein LOC143381	LOC143381	2.1	0.5	4.6
AI365395	Midnolin	MIDN	1.7	1.1	1.5
AA932696	Family with sequence similarity 107, member A	FAM107A	4.3	1.6	2.7
AA496283	Thy-1 cell surface antigen	THY1	1.0	5.7	0.2
AI368486	Ghrelin/obestatin preprohormone	GHRL	25.5	1.3	19.9
AI768615	Immature colon carcinoma transcript 1	ICT1	1.5	0.6	2.6
AI418753	V-set and immunoglobulin domain containing 2	VSIG2	13.4	1.7	7.9
AI371514	WD repeats and SOF1 domain containing	WDSOF1	0.5	1.0	0.5

[a]N, average gene expression ratios for the 29 normal gastric tissues in the training set. [b]T, average gene expression ratios for the 29 gastric tumor tissues in the training set. [c]N/T, relative ratios of N[a] over T[b]. GB acc[d], GenBank accession number.

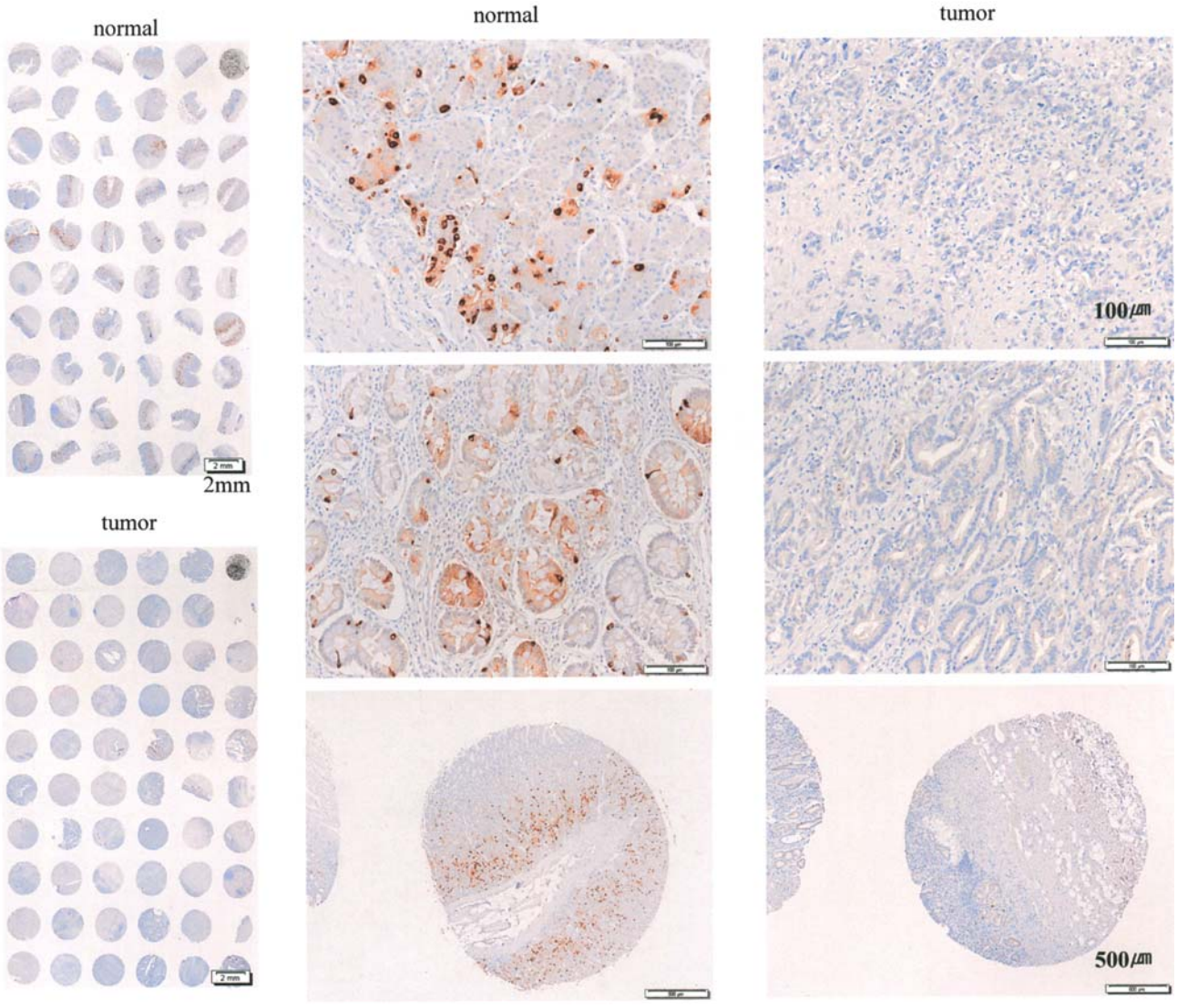

Figure 3. Immunohistostaining of *Chromogranin A*. In the left panel are the overall pictures of stained tissue slides, with normal tissues (top) and tumor tissues (bottom). Three examples of normal or tumor samples stained with CHGA are shown, in the middle panel and on the right, respectively.

and *SOF1 domain containing* (WDSOF1), *Midnolin* (MIDN) and *V-set* and *immunoglobulin domain containing 2* (VSIG2). Of the 22 genes, the gene products of *Glutathione peroxidase 3 (plasma)* (GPX3), INHBA, *Lipase, gastric* (LIPF) and CHGA are soluble proteins that deserve to be studied more for diagnosis purposes. The expression patterns of the selected genes coincide well with those reported in other publications. For example, GPX3, a secretory protein with an antioxidant activity, shows an average of a 3.6-fold increase in expression level in normal gastric tissues compared to gastric adenocarcinoma in the current analysis (Table III). In Barrett's adenocarcinomas (BAs), consistently reduced levels of GPX3 mRNA and protein in tumor samples were observed and the decreased gene expression in the tumor was due to hypermethylation in the promoter region of GPX3 (11). Similar results have been observed in prostate cancer, where GPX3 is hypermethylated and inactivated during prostate cancer progression (12). Furthermore, in a search for cancer-specific CpG methylation that may serve as a tumor marker, Lodygin *et al* found that GPX3 showed the highest frequency of promoter CpG methylation in primary prostate cancer samples (13). INHBA and INHA (α subunit of Inhibin), form a pituitary follicle-stimulating hormone (FSH) secretion inhibitor and have a tumor-suppressor activity. The expression level of INHBA is significantly overexpressed in mouse models of Wnt-induced tumors (14) and, likewise, its expression in gastric adenocacinoma is 10-fold higher than in normal gastric tissues (Table III). CHGA is a molecular marker for tumors expressing neuroendocrine (NE) cell differentiation. It was reported as being overexpressed in small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) tumors and cell lines (15) and immunohistostaining confirms its overexpression in gastric cancer (Fig. 2). *Estrogen-related receptor* γ (ESRRG), overexpressed ~25-fold in normal gastric tissues (Table III), was also reported as being expressed in normal human prostatic epithelial cells, though its expression decreased with the prostate cancer cells and tissues (16,17). Several genes listed here have also been reported previously as showing concordant relative expression profiles in normal and tumor tissues in gastric cancer, including *Kruppel-like factor 4* (KLF4) (18-20), GHRL overexpression in non-neoplastic gastric mucosa (21,22), or

in other cancer models, and *family with sequence similarity 107, member A* (FAM107A), one of the down-regulated genes in colon cancer (23). THY1 is a membrane glycoprotein precursor and is one of the up-regulated genes in colorectal cancer compared to normal mucosa (24), whose general pattern of overexpression in tumors coincides with the current report showing its overexpression in gastric cancer by microarray and immunostaining (data not shown). The relative expression profiles of *Poly (ADP-ribose) polymerase family, member 10* (PARP10) (25), *3-hydroxyisobutyryl-Coenzyme A hydrolase* (HIBCH), SST, LIPF in gastric cancer and normal gastric tissues have not been previously documented elsewhere. The gene expression data, including these four genes presented in this report, would be informative as model genes in the functional studies of gastric cancer.

Acknowledgements

This study was funded by the Korean Science and Engineering Fund (KOSEF) through the Cancer Metastasis Research Center (CMRC) at Yonsei University College of Medicine and also supported by a faculty research grant of Yonsei University College of Medicine for 2007 (No. 6-2007-0187).

References

1. Dicken B, Bigam D, Cass C, Mackey J, Joy A and Hamilton S: Gastric adenocarcinoma. Ann Surg 241: 27-39, 2005.
2. Smith M, Hold G, Tahara E and El-Omar E: Cellular and molecular aspects of gastric cancer. World J Gastroenterol 12: 2979-2990, 2006.
3. Shiraishi N, Sato K, Yasuda K, Inomata M and Kitano S: Multivariate prognostic study on large gastric cancer. J Surg Oncol 96: 14-18, 2007.
4. Hampton T: Cancer drug trials show modest benefit: drugs target liver, gastric, head and neck cancers. JAMA 298: 273-275, 2007.
5. Ellmark P, Ingvarsson J, Carlsson A, Lundin B, Wingren C and Borrebaeck C: Identification of protein expression signatures associated with *Helicobacter pylori* infection and gastric adenocarcinoma using recombinant antibody microarrays. Mol Cell Proteomics 5: 1638-1646, 2006.
6. Yamashita K, Park H, Kim M, Osada M, Tokumaru Y, Inoue H, Mori M and Sidransky D: PGP9.5 methylation in diffuse-type gastric cancer. Cancer Res 66: 3921-3927, 2006.
7. Mikata R, Yokosuka O, Fukai K, Imazeki F, Arai M, Tada M, Kurihara T, Zhang K, Kanda T and Saisho H: Analysis of genes upregulated by the demethylating agent 5-aza-2'-deoxycytidine in gastric cancer cell lines. Int J Cancer 119: 1616-1622, 2006.
8. Chen YR, Juan HF, Huang HC, Huang HH, Lee YJ, Liao MY, Tseng CW, Lin LL, Chen JY, Wang MJ, Chen JH and Chen YJ: Quantitative proteomic and genomic profiling reveals metastasis-related protein expression patterns in gastric cancer cells. J Proteome Res 5: 2727-2742, 2006.
9. Yang S, Shin J, Park KH, Jeung H-C, Rha SY, Noh SH, Yang WI and Chung HC: Molecular basis of the differences between normal and tumor tissues of gastric cancer. Biochim Biophys Acta 1772: 1033-1040, 2007.
10. Bø T and Jonassen I: New feature subset selection procedures for classification of expression profiles. Genome Biol 3: research 0017, 2002.
11. Lee O, Schneider-Stock R, McChesney P, Kuester D, Roessner A, Vieth M, Moskaluk C and El-Rifai W: Hypermethylation and loss of expression of glutathione peroxidase-3 in Barrett's tumorigenesis. Neoplasia 7: 854-861, 2005.
12. Dobosy J, Roberts J, Fu V and DF J: The expanding role of epigenetics in the development, diagnosis and treatment of prostate cancer and benign prostatic hyperplasia. J Urol 177: 822-831, 2007.
13. Lodygin D, Epanchintsev A, Menssen A, Diebold J and Hermeking H: Functional epigenomics identifies genes frequently silenced in prostate cancer. Cancer Res 65: 4218-4227, 2005.
14. Labbé E, Lock L, Letamendia A, Gorska A, Gryfe R, Gallinger S, Moses H and Attisano L: Transcriptional cooperation between the transforming growth factor-beta and Wnt pathways in mammary and intestinal tumorigenesis. Cancer Res 67: 75-84, 2007.
15. Gazdar AF, Helman LJ, Israel MA, Russell EK, Linnoila RI, Mulshine JL, Schuller HM and Park J-G: Expression of neuroendocrine cell markers L-Dopa decarboxylase, chromogranin A and dense core granules in human tumors of endocrine and nonendocrine origin. Cancer Res 48: 4078-4082, 1988.
16. Cheung CP, Yu S, Wong KB, Chan LW, Lai FMM, Wang X, Suetsugi M, Chen S and Chan FL: Expression and functional study of estrogen receptor-related receptors in human prostatic cells and tissues. J Clin Endocrinol Metab 90: 1830-1844, 2005.
17. Yu S, Wang X, Ng C-F, Chen S and Chan FL: ERR{gamma} suppresses cell proliferation and tumor growth of androgen-sensitive and androgen-insensitive prostate cancer cells and its implication as a therapeutic target for prostate cancer. Cancer Res 67: 4904-4914, 2007.
18. Wei D, Gong W, Kanai M, Schlunk C, Wang L, Yao JC, Wu T-T, Huang S and Xie K: Drastic down-regulation of Kruppel-like factor 4 expression is critical in human gastric cancer development and progression. Cancer Res 65: 2746-2754, 2005.
19. Kanai M, Wei D, Li Q, Jia Z, Ajani J, Le X, Yao J and Xie K: Loss of Kruppel-like factor 4 expression contributes to Sp1 overexpression and human gastric cancer development and progression. Clin Cancer Res 12: 6395-6402, 2006.
20. Cho YG, Song JH, Kim CJ, Nam SW, Yoo NJ, Lee JY and Park WS: Genetic and epigenetic analysis of the KLF4 gene in gastric cancer. APMIS 115: 802-808, 2007.
21. An JY, Choi M-G, Noh JH, Sohn TS, Jin D-K and Kim S: Clinical significance of ghrelin concentration of plasma and tumor tissue in patients with gastric cancer. J Surg Res (In press), doi:10.1016/j.jss.2007.1002.1017.
22. Mottershead M, Karteris E, Barclay JY, Suortamo S, Newbold M, Randeva H and Nwokolo CU: Immunohistochemical and quantitative mRNA assessment of ghrelin expression in gastric and oesophageal adenocarcinoma. J Clin Pathol 60: 405-409, 2007.
23. Ancona N, Maglietta R, Piepoli A, D'Addabbo A, Cotugno R, Savino M, Liuni S, Carella M, Pesole G and Perri F: On the statistical assessment of classifiers using DNA microarray data. BMC Bioinformatics 7: 387, 2006.
24. Chiu S-T, Hsieh F-J, Chen S-W, Chen C-L, Shu H-F and Li H: Clinicopathologic correlation of up-regulated genes identified using cDNA microarray and real-time reverse transcription-PCR in human colorectal cancer. Cancer Epidemiol Biomarkers Prev 14: 437-443, 2005.
25. Chou H-YE, Chou HT and Lee S-C: CDK-dependent activation of poly(ADP-ribose) polymerase member 10 (PARP10). J Biol Chem 281: 15201-15207, 2006.