# Identification of differentially expressed genes in gastric cancer by high density cDNA microarray

**Park Se Won**

**Department of Medical Science**

**The Graduated School, Yonsei University**

# Identification of differentially expressed genes in gastric cancer by high density cDNA microarray

**Park Se Won**

**Department of Medical Science**

**The Graduated School, Yonsei University**

# Identification of differentially expressed genes in gastric cancer by high density cDNA microarray

**Directed by Professor Rha Sun Young**

**The Master's Thesis**

**Submitted to the Department of Medical Science,**

**the Graduate School of Yonsei University**

**in partial fulfillment of the requirements**

**for the degree of Master of Medical Science**

**Park Se Won**

**December 2003**

This certifies that the Master's Thesis
of Park Se Won is approved.

---

**Thesis Supervisor**

---

**Thesis Committee Member #1**

---

**Thesis Committee Member #2**

**The Graduate School**

**Yonsei University**

**December 2003**

# ACKNOWLEDGEMENTS

미숙한 저의 논문이 완성되기까지 세심한 지도와 열성으로 이끌어 주신 라선영 교수님께 어떤 형용사로도 표현 할 수 없을 정도로 감사드립니다. 환자들 돌보시며 저희들까지 신경 써 주시고 관심 가져주시는 정현철 선생님께 진심으로 감사 드립니다. 또한, 어려운 일이 생길 때마다 아낌없는 충고와 격려로 든든한 인생선배가 되어주신 양상화 선생님께 감사 드립니다.

석사초기에 미숙한 저를 가르치느라 고생하신 박규현 선생님, 자신이 맞은 일에 항상 성실하신 김태수 선생님, 맨주먹으로 동물실을 개척하며 '살쥐왕'으로 키워주신 심웅호 선생님과 후계자였던 재희, 스타크래프트 지존의 자리를 위협하던 유근이 형님, 정신적 지주이셨던 우영이 형님, 패션의 선두주자 정범이 형님, 나날이 아름다워지는 연호, 모범 연구원 정옥이, 지금은 안계시지만 연구동의 대부와 대모이셨던 태문이 형님과 여말희 누님, 언제나 사람을 공평하게 대하시는 하진이 형

님, 군기의 표상인 씩씩한 박찬희 이병, 너무나 순수하여 가슴 아팠던 영이, 쥐와 생사고락을 함께하는 면희 형님, 항상 조용하고 성실한 주혜누님, 분석의 제왕 상철이, 졸업 후에 더 고생했던 민영이, 주어진 일에 최선을 다하는 지혜, 연구동의 떠오르는 대모 귀연이, 꿋꿋하게 살아가는 다크호스 연주, 연구동의 영원한 킹카 재준이, 자꾸 이름을 잊어버려서 미안했던 경남이, 가까이하기엔 너무 멀었던 Helena, 의기투합하여 살아갔던 재휘, 이 못난 선배와 함께 같은 길을 걸어주었던 후배들, 마음의 안식처가 되어주었던 사람들… 지금의 제가 있을 수 있었던 것은 모두 여러분의 덕분입니다.

2년입니다. 벌써 2년이라는 시간이 지났다는 것이 믿어지지 않을 정도로 시간이 빨리 흐른 것 같습니다. 짧은 시간이라면 짧은 시간이고, 긴 시간이라면 긴 시간이었지만, 노인의 1년과 청년의 1년은 다르다는 말처럼, 그 동안 저에게 찾아온 변화를 받아들이고 앞으로 나가려고 합니다.

마지막으로 지금까지 저를 믿고 지켜주신 가족들과 제가 결심을 굳히는데 큰 힘이 되어 주었던 죽마고우 의섭이에게 감사의 말을 전합니다.

박 세 원 드림

# Table of contents

Abstract ····· 1

I. INTRODUCTION ····· 3

II. MATERIALS AND METHODS ····· 6

1. Tissue samples and clinical data ····· 6

2. Reference RNA and sample RNA preparation ···· 8

A. Cell culture

B. Total RNA extraction

C. RNA purification and quality control

3. High density cDNA microarray experiment ····· 11

A. cDNA microarray

B. Labeling and hybridization

C. Washing and scanning

4. Data analysis ····· 15

A. Normalization and filtering

B. Selection of differentially expressed genes

5. Gene annotation ····· 20

III. RESULTS ····· 21

1. Identification of gastric cancer classifier genes ··· 21

A. Training and cross-validation of expression profile

B. Choosing the amount of shrinkage and selection of classifier genes

2. Validation of identified classifier genes ····· 25

3. Characteristics of 238 classifier genes ····· 29

4. Identification of lymph node metastasis related genes ····· 50

IV. DISCUSSION ····· 56

V. CONCLUSION ····· 64

REFERENCES ····· 65

국문요약 ····· 74

# LIST OF FIGURES

Figure 1. Total RNA for microarray hybridization……10

Figure 2. Scanned microarray image and scatter plot……14

Figure 3. MA-plots after within pin group normalization…17

Figure 4. Box-plots after within pin group normalization…18

Figure 5. The training error rate in a training set……23

Figure 6. Cross-validation probabilities of each sample with different threshold……26

Figure 7. Hierarchical clustering of the training set with 238 selected genes……27

Figure 8. Tumor classification performance with selected 238 genes at threshold 3……28

Figure 9.Verified probabilities of common 18 genes……63

# LIST OF TABLES

Table 1. Patients characteristics·····7

Table 2. List of 238 classifier genes·····31

Table 3. Ontological information of 238 classifier genes····49

Table 4. Sixty-six lymph node metastasis related genes among 238 classifier genes·····51

Table 5. Commonly identified differentially expressed genes between prevous reports and current 238 classifiers·····62

## Abstract

# Identification differentially expressed genes in gastric cancer by high-density cDNA miroarray

**Park Se Won**

**Department of Medical Science**

**The Graduate School, Yonsei University**

**<Directed by Professor Rha Sun Young>**

To obtain molecular signatures of gastric cancer, we compared the expression profiles of 18 gastric cancers and 17 normal gastric tissues using 17,000 human gene containing cDNA microarray. After normalization and filtering, we divided the samples into two sets, 11 pairs as a training set and unpaired 7 gastric cancer and 6 normal gastric tissues as a test set. We selected significant genes in the training set and validated the significance of

the genes in the test set. To identify significant genes in gastric cancer, we used the nearest shrunken centroid method. We obtained 238 classifier genes, which discriminated gastric cancer from normal gastric tissue. The 238 classifiers showed a maximum cross-validation probability, clear hierarchical clustering pattern in the training set, and showed excellent class prediction probability in the independent test set. The classifier genes consisted of known genes related to biological features of cancer and 28% unknown genes. From this study, we obtained genome wide molecular signatures of gastric cancer, which provides preliminary exploration data for the pathophysiology of gastric cancer, and which might serve as basic information for development of biological markers in gastric cancer.

Key words: gastric cancer, cDNA microarray, expression profile, classifier genes

**Identification differentially expressed genes in gastric cancer by high-density cDNA miroarray**

**Park Se Won**

**Department of Medical Science**

**The Graduate School, Yonsei University**

**<Directed by Professor Rha Sun Young>**

# I. INTRODUCTION

Gastric cancer is one of the leading causes of cancer related death in the world[1], and is the most frequent malignancy in Korea. Even though advances in diagnosis and treatment have improved the survival rate for early gastric cancer, advanced gastric cancer retains its dismal prognosis[2]. Therefore, the

identification of molecular markers for diagnosis, treatment and prognosis is a significant issue in management of gastric cancer patients.

Gastric cancer shows various pathological features, resulting from different epidemiologies, biological behavior and clinical properties. Recent molecular studies reported genetic alterations in gastric cancer such as p53[3], β-catenin[4], E-cadherin [5], trefoil factor 1[6], and c-met[7]. However, it is difficult to understand gastric cancer carcinogenesis and progression based on the role of individual gene. Hence, the need for genome wide analysis has been concerned to identify and understand the genetic alteration of gastric cancer.

Array technologies are accurate and comprehensive ways of simultaneously analyzing the genome wide expression and have been applied in various research fields[8]. Excellent advances in microarray techniques as a high-thoughput technology showed sufficient sensitivity and specificity in cancer classification, molecular targeting and prognosis prediction[9-11]. Several studies of gastric cancer using microarray technology

were reported[12-16]. The majority of previous studies were focused on the identification of significant genes for the discrimination of tumor from normal gastric tissues. Moreover, identification of molecular signatures related to clinical features and histological subtype were reported and several efforts have been made to identify expression profiles related to prognosis such as lymph node metastasis, treatment resistance and survival in gastric cancer[11,18-23]. However, these studies have shortcomings concerning the propriety of the supporting statistical analysis, experimental design, and gene selection.

In this study, we performed genome-wide analysis of gastric cancer using qualified tissue samples accompanying clinical information by a high-density cDNA microarray. In addition, we incorporated an optimal experimental design by using two different sets of training and test samples and systemic analysis to identify significant genes in gastric cancer.

## II. MATERIALS AND METHODS

### 1. Tissue samples and clinical data

Eighteen gastric cancer and 17 normal gastric tissues were obtained from gastric cancer patients who underwent curative surgery. Informed consent was obtained from all patients and approval was obtained from the Internal Review Board of the Cancer Metastasis Research Center at Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, Korea. Patient profiles are summarized in Table 1. The fresh samples were snap-frozen in liquid nitrogen immediately after the resection and stored at -70°C until analysis.

**Table 1. Patients characteristics (n=18)**

| | |
|---|---|
| **Gender** | Male : Female 13:5 |
| **Age (years)** | Median (Range) 62 (48-89) |
| **Borrman type** | |
| **II** | 6 |
| **III** | 11 |
| **IV** | 1 |
| **Tumor size(cm)** | Median (Range) 6 (3-15) |
| **Differentiation** | |
| **Well** | 2 |
| **Moderate** | 8 |
| **Poor** | 8 |
| **Lauren's classification** | |
| **Intestinal** | 9 |
| **Diffuse** | 3 |
| **Mixed** | 6 |
| **Infiltration type** | |
| **Infiltrative** | 8 |
| **Expanding** | 3 |
| **Infiltrative and expanding** | 7 |
| **Depth of invasion** | |
| **Pm** | 2 |
| **Ss** | 4 |
| **Se** | 9 |
| **Si** | 3 |
| **Number of lymph node metastasis** | |
| **Negative** | 5 |
| **Positive** | 13 |
| **Stage** | |
| **II** | 5 |
| **IIIA** | 7 |
| **IIIB** | 5 |
| **IV** | 1 |

Histological grading of each gastric cancer tissue was decided according to the classification of WHO. Pm denotes proper muscle, Ss denotes subserosa, Se denotes erosa extended and Si denotes serosa invasion.

**2. Reference RNA and sample RNA preparation.**

**A. Cell culture**

As a reference RNA for microarray experiments, we pooled the 11 cancer cell line RNA to prepare a single batch of Yonsei reference RNA (Cancer Metastasis Research Center, Yonsei University College of Medicine, Seoul, Korea). The cell lines were AGS, MDA-MB-231, HCT116, SK-Hep-1, A549, HL-60, MOLT-4, HeLa, HT1080, Caki-2 and U87MG. Cells were cultured and maintained in minimal essential medium with 10% fetal bovine serum (GIBCO, Grand Island, NY, USA), in 100units/ml of penicillin and 0.1mg/ml of streptomycin (GIBCO, Grand Island, NY, USA), at 37°C in a 5% CO2 incubator. After extracting total RNA from each cell line, we mixed the equal amount of RNA from each.

**B. Total RNA extraction**

Extraction of total RNA from samples was carried out using Trizol reagent (Invitrogen, USA) according to the manufacturer's protocol. Briefly, 1ml

Trizol were added to 50~100mg tissue and homogenized by inverting. After addition of 200ul chloroform, tissue homogenates were centrifuged 10,000xg for 25min at 4°C, and the aqueous phases were collected. Total RNA was precipitated with same volume of isopropanol for 30min at -20°C and centrifuged 10,000xg for 30min. The pellets were washed once with 1ml 70% ethanol and suspended in RNase free water.

**C. RNA purification and quality control**

For better quality of RNA, we purified RNA using Rneasy kit (Qiagen, Germany). 350ul of RLT buffer, 250ul of 95% ethanol were added to 100ul of samples, applied to RNeasy mini column and centrifuged at 8000xg for 15 sec. 500ul of RPE buffer was added to column and centrifuged at 8000xg for 15 sec (repeated 2 times). The purified RNA was eluted by addition of 30ul RNase free water. The quantity and quality of RNA were evaluated using Gene Spec III (Hitachi, Japan) and a Gel Documentation-Photo System (Vilber Lourmat, France), respectively (Fig. 1).

**A.** **B.**

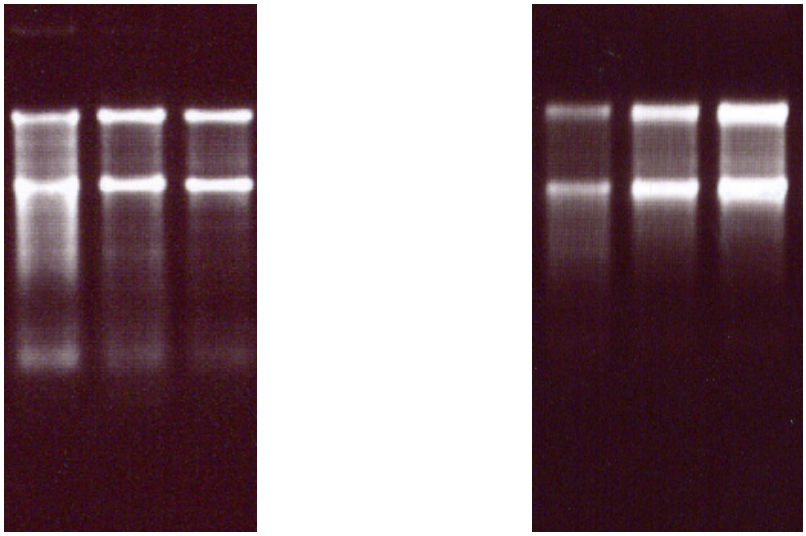

**Figure 1. Total RNA for microarray hybridization.**

A. Before the purification. B. After the purification.

### 3. High-density cDNA microarray experiment

#### A. cDNA microarray preparation

Human cDNA microarray (Yonsei-CMRC-GenmicTree, Seoul, Korea) containing 17,000 genes were used for hybridization. The microarray was pre-hybridized with the blocking solution consisted of 3.5X SSC, 0.1% SDS, 10mg/ml BSA and $dH_2O$. The solution was filtered and incubated at 42°C for 2 hours. The blocked microarray was dipped in water and isopropanol serially and completely dried at 1000rpm for 5 mins.

#### B. Labeling and hybridization

The microarray hybridization was performed following the protocol established at Yonsei CMRC (Yonsei Cancer Metastasis Research Center) based on the protocol of Brown, P. O. *et al*[24-25]. Briefly, total RNA (40μg) were labeled using 6μl of 5X first strand cDNA buffer, 3μl of 0.1M DTT, 0.6μl of low-dTTP dNTP mix (25mM each dATP, dGTP, dCTP and 10mM

dTTP), 400units of Superscript II (Invitrogen, Carlsbad, California, USA). and dUTP-Cy5 for test sample RNA or dUTP-Cy3 for reference RNA. The RT reaction was performed at 42 °C for 2 hours and stopped by addition of 15ul 0.1N NaOH. After incubation at 65°C for 30 mins, 15ul of 0.1N HCl was added for neutralization. The labeled probes were mixed with 30ug human Cot-1 DNA (GIBCOBRL, Gaithersburg, MD, USA), 20ug poly A RNA (Sigma, Saint Louis, Missouri, USA), and 100ug yeast tRNA (GIBCOBRL, Gaithersburg, MD, USA). A Microcon-30 filter (Amicon, Bedford, MA, USA) was used to purify and concentrate the hybridization mixture, which was then adjusted to contain 3.4X SSC and 0.3% SDS in a final volume of 90ul. Following denaturation at 100°C for 2 mins, the probe was hybridized to the microarray under a glass cover slip at 65°C for 8~16 hours.

**C. Washing and scanning**

After hybridization, slides were serially washed with 2X SSC-0.1% SDS, 1X SSC-0.1% SDS, 0.2X SSC, 0.05X SSC at room temperature and dried by centrifugation at 800rpm for 5 min. Hybridized microarrays were scanned using a GenePix 4000B (Axon Instruments, USA) scanner and GenePix Pro 4.1 software (Axon Instruments, USA) (Fig. 2).

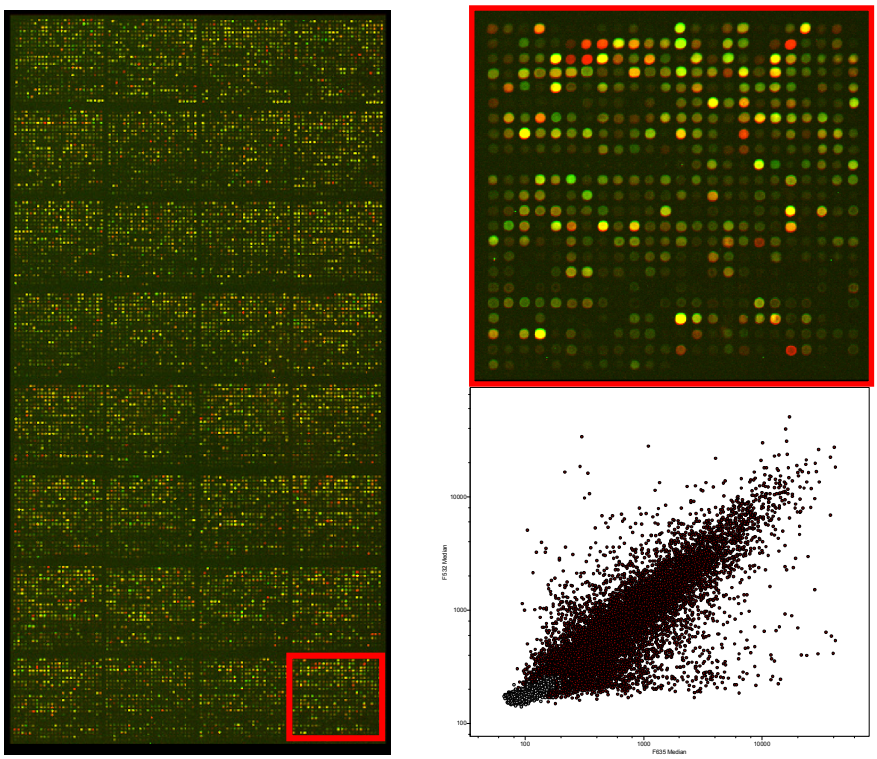


**Figure 2. Scanned microarray image and scatter plot.** A. Scanning image.

B. Scatter plot. : X axis denotes Cy-5 intensity, Y axis denotes Cy-3 intensity.

## 4. Data analysis

### A. Normalization and filtering

To normalize microarray data, we used within pin group, intensity dependent Loess method[26]. An 'MA-plot' was used to represent the (R, G) data, where $M=\log_2 R/G$ and $A=\log_2(RG)^{1/2}$ ; R means F635 signal from Cy-5 and G means F532 signal from Cy-3 labeling. With MA-plots, we detected intensity-dependent patterns for the purpose of normalization (Fig. 3). To correct the pin variation, the 'within pin group normalization'[26] was performed (Fig. 4). A raw data was simply normalized relative to a (pin tip+A),

i.e. $\log_2 R/G \quad \log_2 R/G - c_i(A) = \log_2 R/[ki(A)G]$

where $c_i(A)$ is the Lowess fit to the MA-plot for the *i*th pin group only, $i=1,2,\ldots I$ and $I$ denotes the number of pin groups. The normalized data was filtered using the criteria of missing value >20% in studied samples. Remained data with missing values were adjusted using the K nearest

neighborhood method[27].

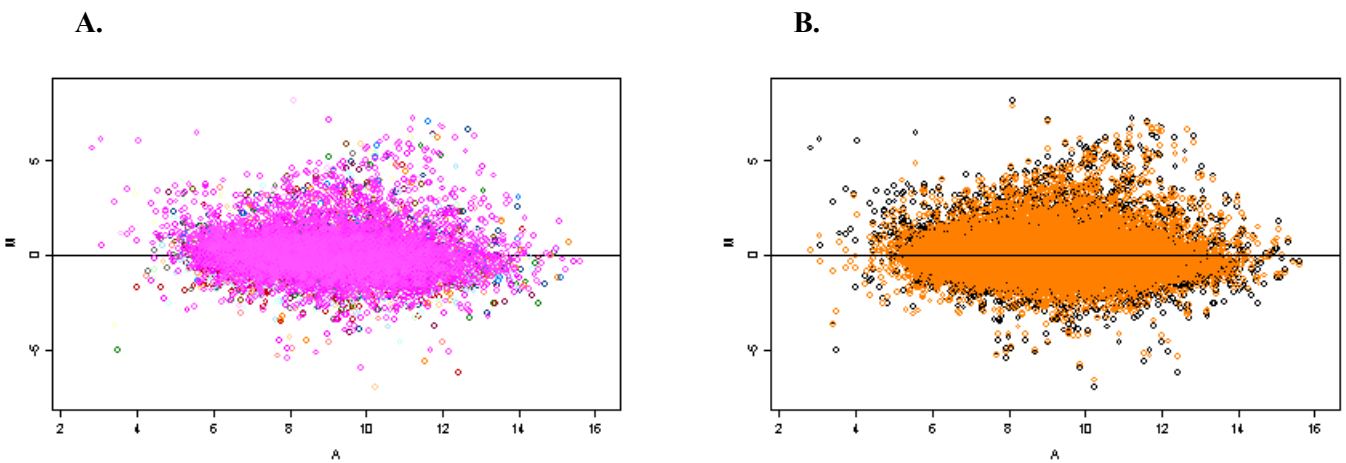


**Figure 3. MA-plots after within pin group normalization.** A. Before the normalization. B. After the normalization.

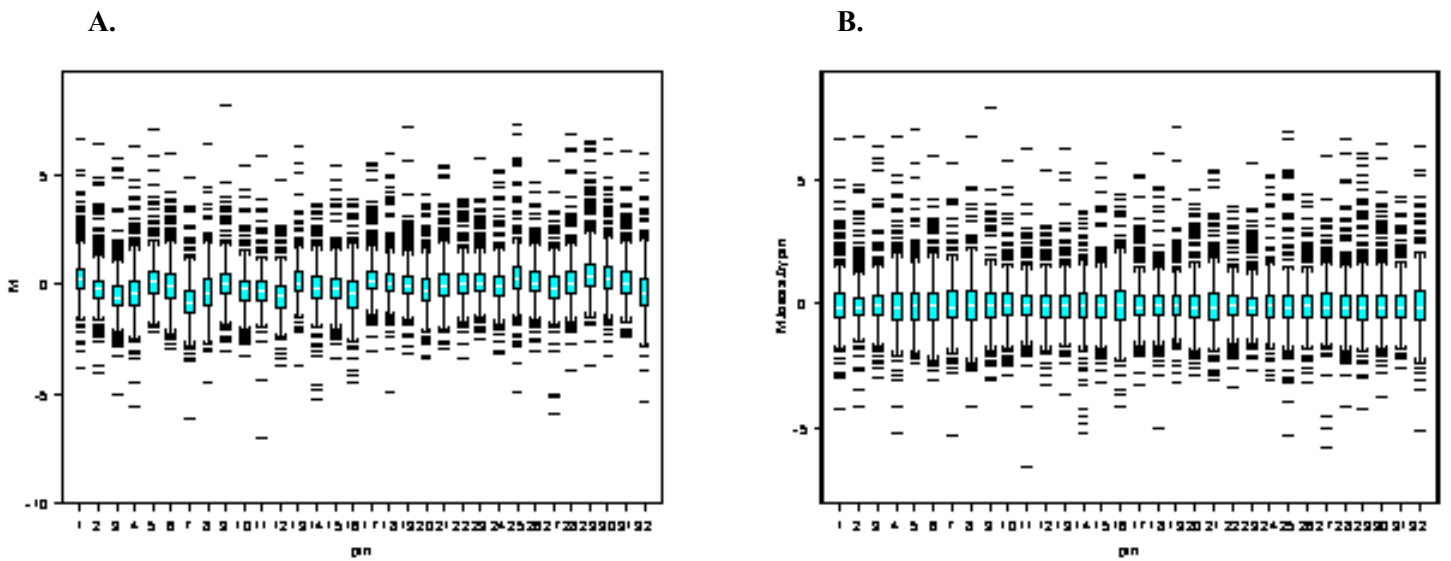


**Figure 4. Box-plots after within pin group normalization.** A. Before the normalization. B. After the normalization.

**B. Selection of differentially expressed genes**

We used 11 paired samples as a training set and 13 unpaired samples (7 gastric cancer and 6 normal gastric tissues) as a test set among 35 samples. To identify differentially expressed genes, which discriminate gastric cancer from normal tissues, namely classifier genes, we used nearest shrunken centroid classification by PAM (Prediction Analysis of Micorarray)[28]. The classifier genes were determined based on minimum training error rate, maximum cross-validation probability and clear hierarchical clustering pattern in the training set. The selected classifier genes were validated by class prediction using Gaussian linear discriminant analysis in independent test set. CLUSTER and TREEVIEW[29] were used for data clustering and visualization. To evaluate the changes of relative gene expressions, we divided mean log R/G of all tumor samples by mean log R/G of all normal samples. To identify the significant genes between patients with or without lymph node metastasis, we compared the mean log R/G of each patient

group.

**5. Gene Annotation**

The ontological information was mined in http://source.stanford.edu and http://apps1/niaid.nih.gov/david. The classifier genes were categorized by their biological process.

# III. RESULTS

## 1. Identification of gastric cancer classifier genes

### A. Training and cross-validation of expression profiles

We selected 11 paired samples as a training set from among the 35 samples to reduce the effect of individual variability. After normalization and filtering as described above, we obtained 12,856 gene expression data, which were available for further analysis. We trained the gene expression profile of the training set to discriminate cancer from normal tissues, and followed this with a cross-validation and a test error rate calculation of various number of genes using PAM. We used balanced 10-fold cross-validation in training set, ensuring that the classes were distributed evenly among the 10 folds. Figure 5 showed the training error rate and cross-validation for different values of shrinkage parameter, i.e. threshold. The shrinkage amount of reducing number of genes played as reducing the effect of noisy genes to make the classifier more accurate. Our data showed a remarkable discrimination

capability between cancer and normal tissues due to improved accuracy of classifier genes, from no shrinkage (Left) to complete shrinkage (Right) (Fig. 5).

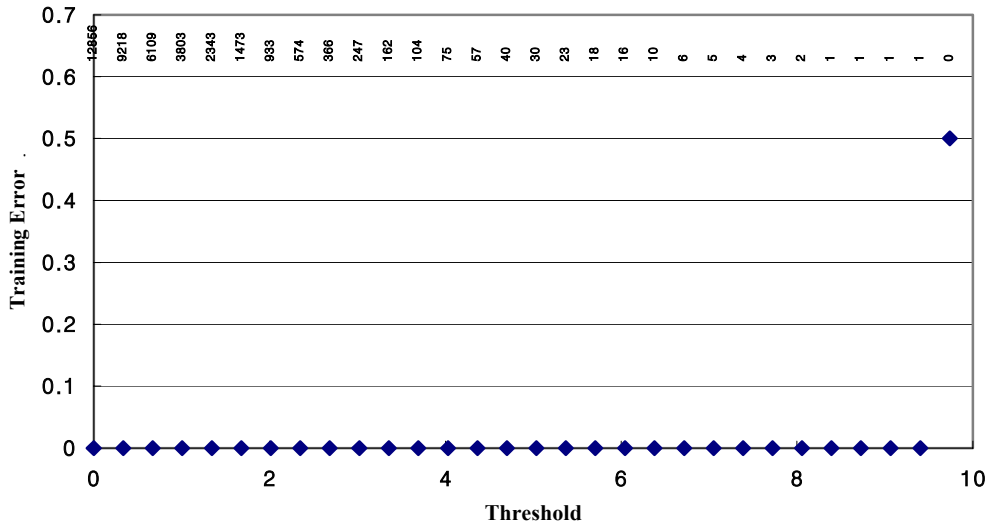


Figure 5. The training error rate in a training set.

The training errors were shown as a function of the shrinkage parameter, threshold. From no shrinkage (left) to complete shrinkage (right), training error rate of “0” was observed at different threshold.

**B. Choosing the amount of shrinkage and selection of classifier**

To select the best classifier genes, we determined shrinkage amount based on cross-validation probability and hierarchical clustering pattern. Even though training error rate showed constant error rate of "0" for different shrinkage parameter, the cross-validation probabilities were different in each threshold (Fig. 6). We observed that the cross-validation probabilities were maximally maintained from threshold 5 to 0. However, low shrinkage amount below threshold 2, which included more than 946 genes, left too many genes to manage. Then, we performed the two way hierarchical clustering with genes of threshold 5 to 3 to evaluate the differential expression patterns of cancer and normal samples. The threshold 5 and 4 showed unclear hierarchical clustering patterns while threshold 3 showed the best clustering pattern. Therefore, we chose threshold 3 as the optimal shrinkage amount, the point, which showed maximum cross-validation probability, with manageable number of 238 genes and clear hierarchical clustering pattern (Fig. 7). We

confirmed that these classifier genes clearly discriminate between tumor and normal group in the training set (Fig. 8-a).

**2. Validation of identified classifier genes**

To verify selected classifier genes, we performed class prediction with 238 genes in a test set, which is independent with a training set and consisted of unpaired 7 gastric cancer and 6 normal tissues. At first, we calculated prediction probabilities of all samples in the test set at each threshold. Compare to cross-validation probabilities of the training set, the prediction probabilities of test set were maximized at threshold 3. In addition, the prediction probabilities of each sample at threshold 3 were showed 0% error rate (Fig. 8-b).

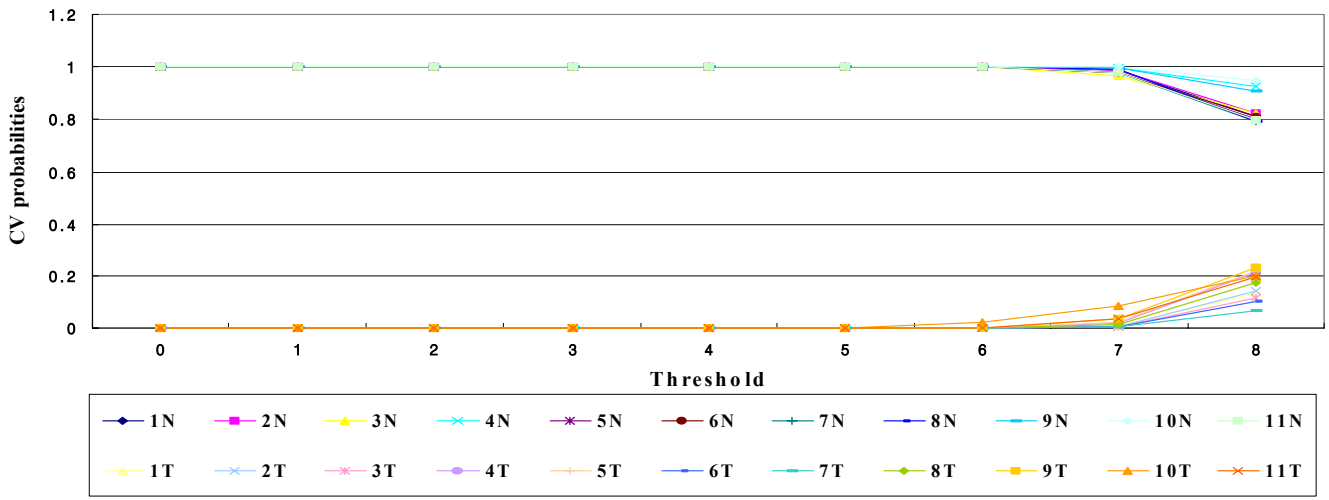


**Figure 6. Cross-validation probabilities of each sample with different threshold.** Estimated cross-validation probabilities of individual samples in a training set for different values of threshold were displayed. Samples were partitioned by scores which representing normal. Cross-validation probabilities were maximized from threshold 5 to 0, showing that all normal samples scored maximum “1” and all tumor samples scored minimum “0”. N: normal, T: tumor

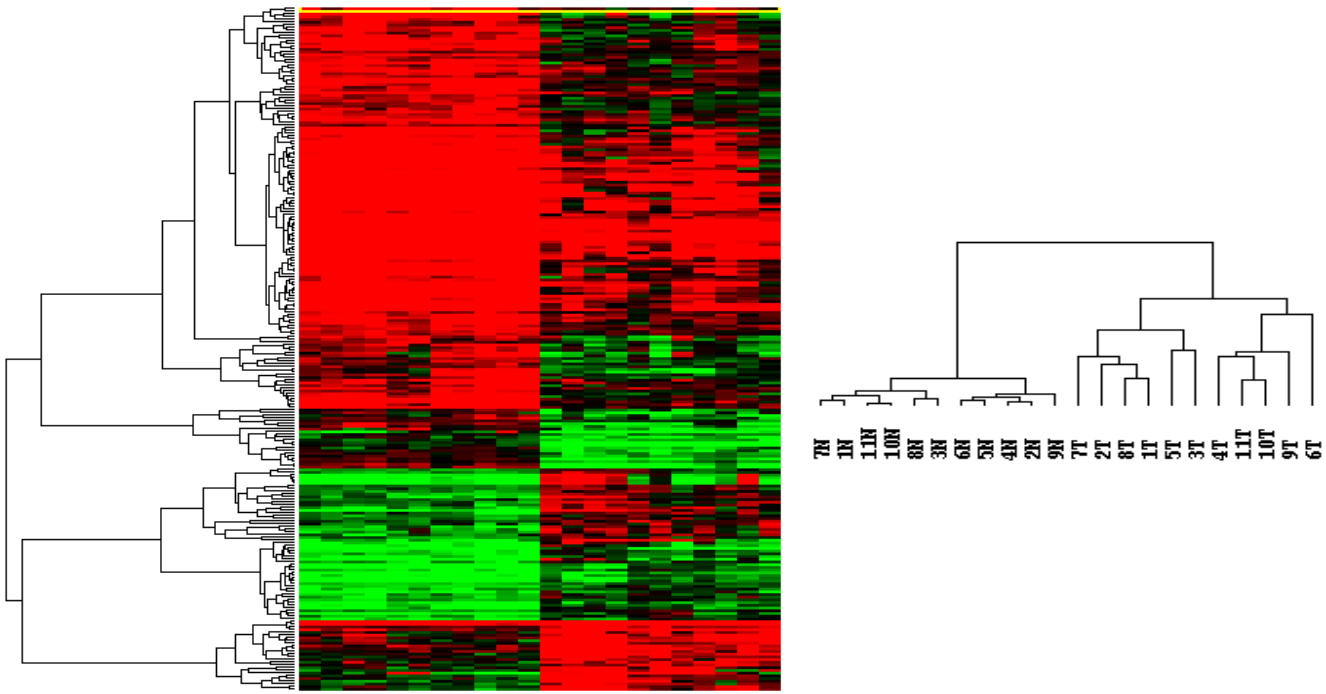


**Figure 7. Hierarchical clustering of the training set with 238 selected genes.** Eleven paired samples were clearly discriminated with 238 genes at threshold 3. 161 genes were down regulated and 77 genes are up regulated in gastric cancers. N: normal, T: tumor2

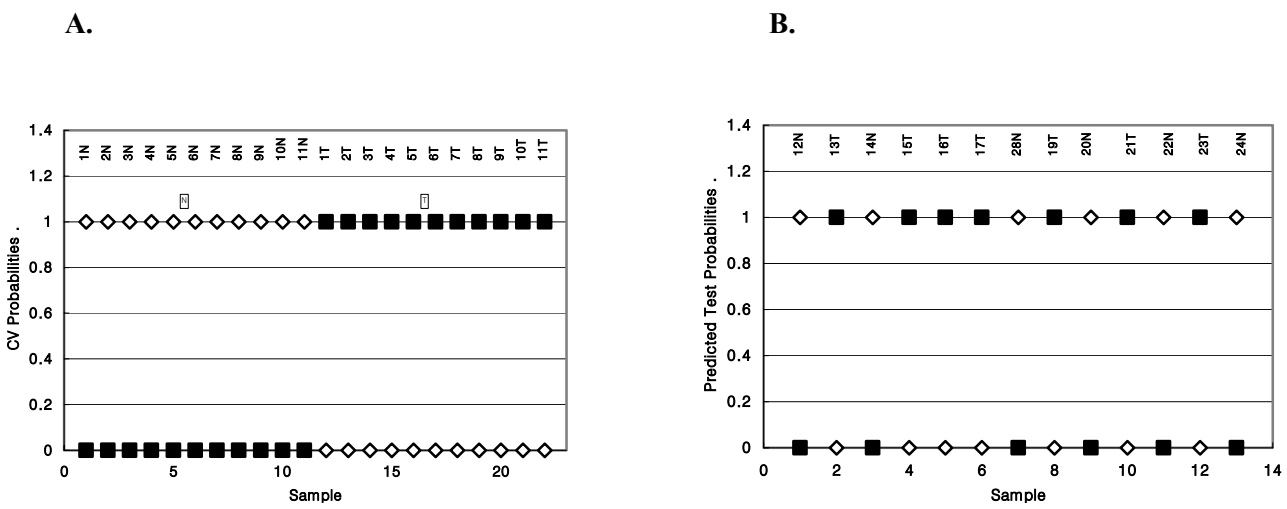


**Figure 8. Tumor classification performance with selected 238 genes at threshold 3.** Normal score (white) and tumor score (black) were voted for each sample. The higher score decided the class of samples and the probability represented the distance between two scores. All samples were correctly classified and showed maximum probability. A. Training set, B. Test set, N: normal, T: tumor.

### 3. Characteristics of 238 classifier genes

The classifier genes consisted of 161 down-regulated genes and 77 up-regulated genes (Table 2). In this study, the most significant gene was "AI001183", of which function is not clearly identified, and it was 16 fold down-regulated in gastric cancer. Calpain 9, which was reported to be related in gastric cancer, showed 7.5 fold down-regulation. Meanwhile, Inhibin, the most significant gene among up-regulated genes, showed 14.5 fold up-regulation in cancer. Endotheial cell specific molecule 1 and thrombspondin 2, which related to angiogensis, 5.3 fold and 7.9 fold up-regulated respectively. The ontological information of 238 classifier genes was categorized by biological process (Table 3). Except 68 unknown genes (28%), the remained 170 classifiers were related to known biological behavior of cancer such as cell cycle, cell growth, cell motility, cell adhesion and extracellular matrix remodeling. Twenty eight genes involved in cell growth and cell adhesion (11%) were mostly up regulated. Other genes showed various biological processes such as angiogenesis (1%), apoptosis (1%), metabolism (14%),

proteolysis (4%), transcription (6%) and signaling (6%).

**Table 2. List of 238 classifier genes**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 1 | AI001183 | similar to RIKEN cDNA 2210417D09 | unknown | 0.25 | 4.37 | -4.1 |
| 2 | AI002047 | hypothetical protein FLJ14464 | unknown | 0.58 | 3.75 | -3.2 |
| 3 | AI074272 | calpain 9 (nCL-4) | ETC | 0.67 | 3.53 | -2.9 |
| 4 | AI924357 | aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III) | metabolism | -5.22 | -0.22 | -5.0 |
| 5 | R93124 | aldo-keto reductase family 1, member C3 (3-alpha hydroxysteroid dehydrogenase, type II) | unknown | -4.56 | -0.56 | -4.0 |
| 6 | AI301329 | aldo-keto reductase family 1, member B10 (aldose reductase) | unknown | -4.48 | 0.94 | -5.4 |
| 7 | AA159577 | mucin 5, subtype B, tracheobronchial | ETC | 1.37 | 5.78 | -4.4 |
| 8 | AI933187 | protein inhibitor of activated STAT protein PIASy | transcription | -1.88 | 0.55 | -2.4 |
| 9 | AI925826 | inhibin, beta A (activin A, activin AB alpha polypeptide) | cell growth | 1.33 | -2.53 | 3.9 |

Cancer denotes the mean log R/G of all cancer samples. Normal denotes the mean log R/G of all normal samples.
Ratio denotes the expression ratio between cancer and normal samples.

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 10 | AA976699 | chromogranin A (parathyroid secretory protein 1) | ETC | -0.28 | 4.51 | -4.8 |
| 11 | AI271987 | hypothetical protein DKFZp761G0122 | unknown | 0.49 | 4.34 | -3.8 |
| 10 | AA976699 | chromogranin A (parathyroid secretory protein 1) | ETC | -0.28 | 4.51 | -4.8 |
| 11 | AI271987 | hypothetical protein DKFZp761G0122 | unknown | 0.49 | 4.34 | -3.8 |
| 12 | AI913412 | estrogen-related receptor gamma | transcription | 0.36 | 4.02 | -3.7 |
| 13 | W88655 | sulfotransferase family, cytosolic, 1C, member 1 | metabolism | 1.35 | 4.70 | -3.3 |
| 14 | AW058221 | lipase, gastric | metabolism | 1.32 | 8.00 | -6.7 |
| 15 | AI674972 | progastricsin (pepsinogen C) | proteolysis | 2.83 | 8.22 | -5.4 |
| 16 | R72097 | similar to Pepsin A precursor | ETC | 1.74 | 8.29 | -6.5 |
| 17 | AI003367 | gap junction protein, alpha 7, 45kDa (connexin 45) | transport | 1.14 | 4.51 | -3.4 |
| 18 | AW028846 | trefoil factor 2 (spasmolytic protein 1) | ETC | 2.37 | 7.77 | -5.4 |
| 19 | R51912 | somatostatin | signaling | -0.34 | 1.81 | -2.1 |
| 20 | AI418194 | Homo sapiens transcribed sequences | unknown | -0.38 | 3.78 | -4.2 |
| 21 | R91396 | annexin A10 | ETC | 0.90 | 6.09 | -5.2 |
| 22 | AA025150 | Sapiens, clone IMAGE:5478062, mRNA | unknown | 0.06 | 2.80 | -2.7 |
| 23 | AI368486 | ghrelin precursor | signaling | 0.45 | 3.71 | -3.3 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 24 | H23187 | carbonic anhydrase II | metabolism | 1.64 | 5.98 | -4.3 |
| 25 | AA630584 | prostate stem cell antigen | ETC | 1.59 | 6.39 | -4.8 |
| 26 | AA775223 | hydroxyprostaglandin dehydrogenase 15-(NAD) | metabolism | -0.84 | 2.97 | -3.8 |
| 27 | H94487 | cathepsin E | proteolysis | 1.66 | 5.50 | -3.8 |
| 28 | AA844831 | carboxypeptidase A2 (pancreatic) | proteolysis | -0.04 | 4.19 | -4.2 |
| 29 | AI936084 | secretoglobin, family 2A, member 1 | unknown | -1.80 | 1.72 | -3.5 |
| 30 | AI768615 | immature colon carcinoma transcript 1 | ETC | -1.03 | 0.44 | -1.5 |
| 31 | AI333599 | 18 kDa antrum mucosa protein | unknown | 1.98 | 8.96 | -7.0 |
| 32 | AA496997 | lamin A/C | unknown | 0.16 | 2.84 | -2.7 |
| 33 | H38240 | thrombospondin 2 | cell adhesion | 2.76 | -0.22 | 3.0 |
| 34 | AA436401 | TU3A protein | unknown | 0.88 | 2.90 | -2.0 |
| 35 | AI337340 | coenzyme Q7 homolog, ubiquinone (yeast) | ETC | -0.09 | 1.23 | -1.3 |
| 36 | AA486324 | proteasome (prosome, macropain) activator subunit 3 (PA28 gamma; Ki) | ETC | 0.22 | 2.16 | -1.9 |
| 37 | AI375428 | EST | unknown | 0.46 | 2.79 | -2.3 |
| 38 | W46577 | endothelial cell-specific molecule 1 | cell growth | 0.90 | -1.49 | 2.4 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 39 | AI991902 | zinc finger protein 145 (Kruppel-like, expressed in promyelocytic leukemia) | cell growth | -0.04 | 1.62 | -1.7 |
| 40 | AI745626 | EST | unknown | -0.93 | 2.17 | -3.1 |
| 41 | AI924634 | involucrin | unknown | -1.15 | 0.38 | -1.5 |
| 42 | R62603 | collagen, type VI, alpha 3 | extracellular matrix | 0.94 | -1.35 | 2.3 |
| 43 | N53136 | cytochrome P450, family 2, subfamily C, polypeptide 8 | transport | 1.80 | 4.20 | -2.4 |
| 44 | AA977679 | somatostatin receptor 1 | signaling | 0.45 | 3.62 | -3.2 |
| 45 | AA496283 | Thy-1 cell surface antigen | unknown | 2.90 | 0.48 | 2.4 |
| 46 | AA280692 | diacylglycerol kinase, delta 130kDa | signaling | -0.57 | 1.21 | -1.8 |
| 47 | AW009769 | trefoil factor 1 | cell growth | 2.89 | 7.21 | -4.3 |
| 48 | AA683077 | LOC150225 | unknown | 1.59 | -0.45 | 2.0 |
| 49 | R71093 | serine (or cysteine) proteinase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1) | unknown | 0.61 | -1.41 | 2.0 |
| 50 | AI032392 | RAB27A, member RAS oncogene family | transport | 0.20 | 2.24 | -2.0 |
| 51 | H45668 | Kruppel-like factor 4 (gut) | transcription | 1.11 | 3.51 | -2.4 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 52 | AA931491 | hypothetical protein LOC143381 | unknown | -1.04 | 1.01 | -2.1 |
| 53 | AI657057 | glucosaminyl (N-acetyl) transferase 1, core 2 | metabolism | 0.89 | 2.69 | -1.8 |
| 54 | AI884731 | wingless-type MMTV integration site family, member 10B | signaling | -0.97 | 0.31 | -1.3 |
| 55 | AI868227 | alcohol dehydrogenase 1C (class I), gamma polypeptide | metabolism | 2.76 | 7.13 | -4.4 |
| 56 | AI418753 | cortical thymocyte receptor (X. laevis CTX) like | unknown | 0.41 | 3.15 | -2.7 |
| 57 | AA485893 | ribonuclease, RNase A family, 1 (pancreatic) | extracellular matrix | 2.73 | 5.61 | -2.9 |
| 58 | W94629 | Sapiens cDNA FLJ11796 fis, clone HEMBA1006158, highly similar to Homo sapiens transcription factor forkhead-like 7 (FKHL7) gene. | unknown | -0.53 | -2.59 | 2.1 |
| 59 | AA521345 | BTB (POZ) domain containing 2 | ETC | 0.08 | 1.63 | -1.5 |
| 60 | AA521439 | synaptotagmin-like 2 | unknown | 0.65 | 2.73 | -2.1 |
| 61 | AA677706 | lactotransferrin | transport | 2.34 | 6.81 | -4.5 |
| 62 | AI366996 | immunoglobulin heavy constant mu | unknown | 2.86 | 6.31 | -3.4 |
| 63 | AI017442 | ectonucleoside triphosphate diphosphohydrolase 5 | unknown | 0.68 | 2.40 | -1.7 |
| 64 | AA458878 | agrin | unknown | -0.02 | -1.40 | 1.4 |
| 65 | AI802786 | UDP glycosyltransferase 2 family, polypeptide B17 | metabolism | 1.20 | 3.90 | -2.7 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 66 | N30372 | interferon regulatory factor 5 | cell growth | 1.70 | 5.34 | -3.6 |
| 67 | AA864299 | proapoptotic caspase adaptor protein | apoptosis | 0.35 | 1.72 | -1.4 |
| 68 | AA490172 | collagen, type I, alpha 2 | extracellular matrix | 3.98 | 1.79 | 2.2 |
| 69 | AA425217 | cadherin 3, type 1, P-cadherin (placental) | extracellular matrix | 1.21 | -1.13 | 2.3 |
| 70 | AA775616 | secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1) | ETC | -2.11 | -5.72 | 3.6 |
| 71 | T65736 | selenium binding protein 1 | ETC | 2.61 | 4.52 | -1.9 |
| 72 | AI538192 | hypothetical protein DKFZp761N1114 | unknown | 0.36 | 3.33 | -3.0 |
| 73 | AA400258 | Gene 33/Mig-6 | ETC | -0.05 | 1.57 | -1.6 |
| 74 | AI653116 | sulfatase 1 | apoptosis | 5.80 | 2.79 | 3.0 |
| 75 | AA911832 | Sapiens, clone IMAGE:4471726, mRNA | unknown | 0.29 | -1.60 | 1.9 |
| 76 | AI984082 | retinoic acid induced 3 | signaling | 0.41 | -1.45 | 1.9 |
| 77 | AA152347 | glutathione S-transferase A4 | ETC | 0.10 | 2.11 | -2.0 |
| 78 | AA045320 | arylacetamide deacetylase (esterase) | metabolism | -0.15 | 2.90 | -3.0 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 79 | AA911063 | Homo sapiens transcribed sequence with moderate similarity to protein ref:NP_060265.1 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens] | unknown | 0.28 | 2.03 | -1.7 |
| 80 | AI473884 | solute carrier family 16 | transport | -1.48 | 1.36 | -2.8 |
| 81 | AA490497 | ubiquitin-like 3 | unknown | 0.92 | 2.62 | -1.7 |
| 82 | AA953560 | Homo sapiens transcribed sequence with weak similarity to protein ref:NP_060265.1 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens] | unknown | -0.04 | -2.88 | 2.8 |
| 83 | AI828306 | collagen, type X, alpha 1 | extracellular matrix | 2.97 | 0.42 | 2.6 |
| 84 | AI057267 | EST | unknown | 2.81 | 0.82 | 2.0 |
| 85 | W72294 | chemokine (C-X-C motif) ligand 14 | signaling | 2.11 | 4.26 | -2.1 |
| 86 | AA449336 | protein regulator of cytokinesis 1 | ETC | -0.93 | -2.50 | 1.6 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 87 | AA911705 | Homo sapiens transcribed sequence with moderate similarity to protein ref:NP_060312.1 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens] | unknown | 0.94 | 3.81 | -2.9 |
| 88 | AA975430 | period homolog 1 (Drosophila) | signaling | 0.13 | 1.61 | -1.5 |
| 89 | H95960 | secreted protein, acidic, cysteine-rich (osteonectin) | extracellular matrix | 1.72 | -0.03 | 1.7 |
| 90 | T71349 | cytochrome P450, family 3, subfamily A, polypeptide 4 | metabolism | 1.78 | 4.33 | -2.5 |
| 91 | AA446259 | protein tyrosine phosphatase, non-receptor type 12 | ETC | 0.37 | -1.01 | 1.4 |
| 92 | AA923696 | hypothetical protein MGC11324 | unknown | -1.59 | 0.57 | -2.2 |
| 93 | T98612 | collagen, type III, alpha 1 | extracellular matrix | 5.30 | 3.67 | 1.6 |
| 94 | AA865554 | LBP protein; likely ortholog of mouse CRTR-1 | transcription | 0.48 | 1.86 | -1.4 |
| 95 | AI341050 | RecQ protein-like 4 | DNA repair | -0.34 | -2.12 | 1.8 |
| 96 | W46900 | chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha) | cell growth | 1.78 | -0.24 | 2.0 |
| 97 | AI000804 | EST | unknown | -1.92 | -0.13 | -1.8 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 98 | AA187351 | ribonucleotide reductase M2 polypeptide | DNA replication | -0.86 | -1.90 | 1.0 |
| 99 | R16712 | anillin, actin binding protein | ETC | -0.93 | -2.33 | 1.4 |
| 100 | AA991590 | apolipoprotein C-I | metabolism | 3.04 | 0.70 | 2.3 |
| 101 | AA521228 | 3-hydroxyisobutyryl-Coenzyme A hydrolase | metabolism | 0.09 | 1.52 | -1.4 |
| 102 | AA133469 | keratin 20 | unknown | 0.64 | 4.88 | -4.2 |
| 103 | AA857098 | collagen, type V, alpha 2 | extracellular matrix | 1.49 | -0.36 | 1.8 |
| 104 | AI017394 | hypothetical protein LOC283445 | unknown | 0.57 | 1.97 | -1.4 |
| 105 | AI814383 | cathepsin L2 | extracellular matrix | 0.30 | -0.96 | 1.3 |
| 106 | AI031571 | epithelial cell transforming sequence 2 oncogene | cell growth | -0.75 | -2.22 | 1.5 |
| 107 | AI147534 | EST | unknown | 0.81 | 2.34 | -1.5 |
| 108 | AI422138 | Homo sapiens transcribed sequences | unknown | -0.44 | 2.58 | -3.0 |
| 109 | AA489587 | fibronectin 1 | cell motility | -0.60 | -3.31 | 2.7 |
| 110 | AA405569 | fibroblast activation protein, alpha | proteolysis | 1.92 | -0.20 | 2.1 |
| 111 | AA497002 | melanoma cell adhesion molecule | cell adhesion | -0.75 | -3.26 | 2.5 |
| 112 | AA488406 | mesothelin | cell adhesion | 0.63 | -1.83 | 2.5 |
| 113 | AI381043 | inositol 1,4,5-trisphosphate 3-kinase A | signaling | 0.28 | 1.85 | -1.6 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 114 | AI689831 | Sapiens, clone IMAGE:4816940, mRNA | unknown | -0.55 | 0.93 | -1.5 |
| 115 | AA971274 | hypothetical protein FLJ10916 | unknown | 0.95 | 2.51 | -1.6 |
| 116 | AA100036 | ectodermal-neural cortex | ETC | 0.45 | -1.21 | 1.7 |
| 117 | H53340 | metallothionein 1G | ETC | -0.29 | 2.42 | -2.7 |
| 118 | AA448261 | high mobility group AT-hook 1 | transcription | -0.71 | -2.29 | 1.6 |
| 119 | AI791122 | cholecystokinin B receptor | ETC | 0.68 | 2.92 | -2.2 |
| 120 | AA677534 | laminin, gamma 2 | cell adhesion | 0.76 | -0.59 | 1.4 |
| 121 | N24824 | v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog | cell growth | 0.07 | 2.08 | -2.0 |
| 122 | AI675465 | homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1 | ETC | 0.07 | 1.51 | -1.4 |
| 123 | AI244667 | transcription factor 2, hepatic; LF-B3; variant hepatic nuclear factor | transcription | -0.27 | 1.31 | -1.6 |
| 124 | AA086476 | adenosine monophosphate deaminase 1 | metabolism | 0.91 | 2.40 | -1.5 |
| 125 | AA446462 | BUB1 budding uninhibited by benzimidazoles 1 homolog | cell cycle | -0.91 | -2.49 | 1.6 |
| 126 | AA775091 | delta sleep inducing peptide, immunoreactor | transcription | 1.34 | 2.65 | -1.3 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 127 | AI473897 | Full-length cDNA clone CS0DM001YM08 of Fetal liver of Homo sapiens | unknown | -1.35 | 1.34 | -2.7 |
| 128 | AA070226 | similar to restin | unknown | 3.51 | 5.81 | -2.3 |
| 129 | AA971278 | cytochrome P450, family 2, subfamily S, polypeptide 1 | metabolism | 0.99 | 2.53 | -1.5 |
| 130 | AA844864 | regenerating islet-derived 1 beta | unknown | 2.64 | 7.03 | -4.4 |
| 131 | AI023541 | carbonic anhydrase IX | metabolism | 0.19 | 3.51 | -3.3 |
| 132 | AI221536 | tumor necrosis factor receptor superfamily, member 12A | apoptosis | -1.03 | -2.88 | 1.9 |
| 133 | AI813911 | Down syndrome critical region gene 1-like 1 | transcription | 2.44 | 4.58 | -2.1 |
| 134 | AA047778 | myosin IB | cell growth | -0.54 | -1.67 | 1.1 |
| 135 | R31701 | Homo sapiens transcribed sequences | unknown | 2.14 | -0.81 | 2.9 |
| 136 | AA872383 | metallothionein 1E | transport | -0.72 | 1.58 | -2.3 |
| 137 | AA459305 | procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3 | ETC | 0.02 | -1.23 | 1.3 |
| 138 | AW075162 | tissue inhibitor of metalloproteinase 1 | proteolysis | 0.95 | -1.00 | 2.0 |
| 139 | AA280832 | galactose-4-epimerase, UDP- | metabolism | 0.27 | 1.70 | -1.4 |
| 140 | AI245812 | potassium inwardly-rectifying channel, subfamily J, member 15 | transport | 0.67 | 2.79 | -2.1 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 141 | AA625655 | regenerating islet-derived 1 alpha | unknown | 2.42 | 6.83 | -4.4 |
| 142 | AA434115 | chitinase 3-like 1 | metabolism | 1.73 | -0.24 | 2.0 |
| 143 | AA284669 | plasminogen activator, urokinase | cell growth | -0.48 | -2.42 | 1.9 |
| 144 | AI859300 | Sapiens clone FLB9440 PRO2550 mRNA | unknown | 0.05 | 1.93 | -1.9 |
| 145 | AA857542 | ATPase, Ca++ transporting, ubiquitous | metabolism | -0.57 | 0.66 | -1.2 |
| 146 | T70999 | UDP glycosyltransferase 1 family, polypeptide A9 | metabolism | -0.79 | 1.64 | -2.4 |
| 147 | N64508 | podocalyxin-like | unknown | -0.21 | -1.52 | 1.3 |
| 148 | AA136125 | EST | unknown | -0.58 | -1.71 | 1.1 |
| 149 | N62179 | aldehyde dehydrogenase 6 family, member A1 | metabolism | 0.01 | 1.64 | -1.6 |
| 150 | AW057705 | fms-related tyrosine kinase 3 | signaling | -0.31 | 1.28 | -1.6 |
| 151 | AA913127 | glucosaminyl (N-acetyl) transferase 2, I-branching enzyme | ETC | -0.84 | 0.62 | -1.5 |
| 152 | AA873089 | cytochrome P450, family 3, subfamily A, polypeptide 5 pseudogene 2 | metabolism | 1.35 | 3.26 | -1.9 |
| 153 | R89492 | cytochrome P450, family 2, subfamily C, polypeptide 9 | metabolism | 0.74 | 2.47 | -1.7 |
| 154 | AI261741 | transmembrane protease, serine 2 | proteolysis | 2.48 | 4.64 | -2.2 |
| 155 | AA862465 | alpha-2-glycoprotein 1, zinc | immune response | 2.45 | 5.48 | -3.0 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 156 | AA047567 | progesterone receptor membrane component 2 | ETC | -0.56 | 0.58 | -1.1 |
| 157 | R50354 | leukemia inhibitory factor | signaling | -1.17 | -3.10 | 1.9 |
| 158 | AA932983 | UDP glycosyltransferase 1 family, polypeptide A10 | ETC | -1.09 | 1.12 | -2.2 |
| 159 | AW072778 | transcription factor CP2 | transcription | 4.11 | 6.34 | -2.2 |
| 160 | W90085 | nuclear receptor subfamily 0, group B, member 2 | transcription | 0.92 | 2.83 | -1.9 |
| 161 | AA479745 | hypothetical protein MGC27165 | unknown | 4.99 | 7.83 | -2.8 |
| 162 | AI521155 | G protein-coupled receptor 64 | ETC | -1.41 | 0.29 | -1.7 |
| 163 | AI289110 | nuclear transport factor 2 | transport | -0.82 | 1.54 | -2.4 |
| 164 | AA862966 | EST | unknown | 0.44 | 3.38 | -2.9 |
| 165 | R67275 | collagen, type XI, alpha 1 | extracellular matrix | 3.02 | 0.71 | 2.3 |
| 166 | AI375353 | serum/glucocorticoid regulated kinase | apoptosis | 0.73 | 2.50 | -1.8 |
| 167 | AI658727 | microseminoprotein, beta- | unknown | -0.07 | 1.54 | -1.6 |
| 168 | AA775257 | integral membrane protein 2A | ETC | -1.41 | 0.53 | -1.9 |
| 169 | AI989728 | serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5 | cell motility | 1.61 | -0.68 | 2.3 |
| 170 | AA954935 | matrix metalloproteinase 11 | extracellular | 2.14 | 0.10 | 2.0 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 171 | AA877213 | cytochrome P450, family 24, subfamily A, polypeptide 1 | transport | -1.88 | -3.33 | 1.5 |
| 172 | AA011096 | monoamine oxidase A | metabolism | 0.88 | 3.00 | -2.1 |
| 173 | AI291863 | fructosamine-3-kinase | metabolism | 4.01 | 5.73 | -1.7 |
| 174 | AI041729 | protein disulfide isomerase, pancreatic | transport | 0.39 | 3.25 | -2.9 |
| 175 | N68159 | ATP-binding cassette, sub-family C (CFTR/MRP), member 5 | transport | -0.21 | 0.99 | -1.2 |
| 176 | AI289178 | Sapiens cDNA FLJ31206 fis | unknown | 0.43 | 2.03 | -1.6 |
| 177 | AA459401 | kallikrein 10 | cell cycle | 1.77 | -0.45 | 2.2 |
| 178 | AA683578 | adenosine deaminase | metabolism | -2.88 | -0.94 | -1.9 |
| 179 | AA947730 | epsin 1 | ETC | 0.11 | 1.31 | -1.2 |
| 180 | AI244615 | alcohol dehydrogenase 6 (class V) | metabolism | 0.62 | 2.77 | -2.2 |
| 181 | AA903860 | homer homolog 1 | ETC | -0.83 | -2.10 | 1.3 |
| 182 | AA449742 | coagulation factor XIII, A1 polypeptide | ETC | 2.07 | 4.11 | -2.0 |
| 183 | AA150402 | collagen, type IV, alpha 1 | extracellular matrix | 0.16 | -1.08 | 1.2 |
| 184 | AA968896 | midkine (neurite growth-promoting factor 2) | cell cycle | 1.44 | 0.10 | 1.3 |
| 185 | AW005713 | serine protease inhibitor, Kazal type 1 | ETC | 3.40 | 6.18 | -2.8 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 186 | W73874 | cathepsin L | extracellular matrix | 0.13 | -1.22 | 1.4 |
| 187 | AI969643 | phosphoribosylaminoimidazole carboxylase | ETC | -1.19 | -2.48 | 1.3 |
| 188 | AA903201 | Notch homolog 1 | signaling | -0.47 | -1.71 | 1.2 |
| 189 | AI269774 | phytanoyl-CoA hydroxylase | metabolism | -0.11 | 1.02 | -1.1 |
| 190 | R92425 | cytochrome P450, family 3, subfamily A, polypeptide 5 | metabolism | 2.31 | 4.90 | -2.6 |
| 191 | AI971049 | myocilin | ETC | 0.05 | 1.35 | -1.3 |
| 192 | AA644448 | protein tyrosine phosphatase, receptor type, U | cell adhesion | -0.16 | -1.45 | 1.3 |
| 193 | AW029415 | DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 21 | unknown | -0.38 | -1.48 | 1.1 |
| 194 | AA934734 | hypothetical protein FLJ31819 | unknown | 0.76 | 2.79 | -2.0 |
| 195 | AI418638 | hypothetical protein LOC201895 | unknown | 0.39 | 1.76 | -1.4 |
| 196 | AA488324 | BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast) | cell cycle | -0.88 | -1.95 | 1.1 |
| 197 | AA052960 | dyskeratosis congenita 1, dyskerin | cell growth | -0.58 | -1.55 | 1.0 |
| 198 | H52119 | flavin containing monooxygenase 5 | transport | 1.62 | 3.56 | -1.9 |
| 199 | AA136983 | cadherin 11, type 2, OB-cadherin (osteoblast) | extracellular matrix | 1.77 | 0.38 | 1.4 |
| 200 | AW072500 | creatine kinase, brain | metabolism | -0.22 | 2.00 | -2.2 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 201 | AA987505 | Homo sapiens transcribed sequences | unknown | 0.24 | -1.36 | 1.6 |
| 202 | R32848 | S100 calcium binding protein P | unknown | 2.05 | 4.33 | -2.3 |
| 203 | W76339 | nuclear factor (erythroid-derived 2)-like 3 | transcription | 0.98 | -0.54 | 1.5 |
| 204 | AW004895 | EST | unknown | -0.41 | -1.72 | 1.3 |
| 205 | AI799888 | tenascin XB | cell adhesion | 1.26 | 2.69 | -1.4 |
| 206 | AI830324 | pancreatitis-associated protein | cell growth | 2.88 | 6.71 | -3.8 |
| 207 | AA491209 | EST | unknown | 1.08 | 2.54 | -1.5 |
| 208 | AA922832 | intercellular adhesion molecule 3 | cell adhesion | 0.47 | 2.90 | -2.4 |
| 209 | AI312971 | EST | unknown | 0.41 | 1.45 | -1.0 |
| 210 | AA504130 | cytoskeleton associated protein 2 | unknown | -0.53 | -1.68 | 1.1 |
| 211 | AA279980 | basic helix-loop-helix domain containing, class B, 3 | transcription | 0.07 | 1.77 | -1.7 |
| 212 | AA857804 | claspin homolog | unknown | -1.05 | -2.32 | 1.3 |
| 213 | AA457114 | tumor necrosis factor, alpha-induced protein 2 | angiogenesis. | 0.39 | -1.16 | 1.5 |
| 214 | AI000474 | hypothetical protein BC007436 | unknown | 0.36 | 1.47 | -1.1 |
| 215 | AI797648 | T-cell leukemia, homeobox 2 | transcription | 0.39 | 2.10 | -1.7 |
| 216 | W92764 | tumor necrosis factor, alpha-induced protein 6 | cell adhesion | 2.04 | 0.42 | 1.6 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 217 | R56774 | bone morphogenetic protein 1 | proteolysis | -0.17 | -1.43 | 1.3 |
| 218 | AI016051 | chemokine ligand 14 | signaling | 3.67 | 6.25 | -2.6 |
| 219 | AA464606 | pannexin 1 | transport | 0.32 | -0.70 | 1.0 |
| 220 | R48303 | dermatopontin | extracellular matrix | 1.05 | 3.09 | -2.0 |
| 221 | AI000966 | KIAA1324 protein | unknown | 2.14 | 4.88 | -2.7 |
| 222 | AI952218 | neuromedin B | signaling | -1.65 | -0.40 | -1.2 |
| 223 | AI015711 | putative ATPase | metabolism | 0.30 | 2.95 | -2.6 |
| 224 | AI628353 | KIAA0882 protein | unknown | -0.01 | 1.58 | -1.6 |
| 225 | AI253136 | ERO1-like beta | unknown | -0.42 | 1.56 | -2.0 |
| 226 | H72723 | metallothionein 1B | unknown | -1.41 | 0.62 | -2.0 |
| 227 | AA423957 | thrombospondin 4 | cell adhesion | 2.04 | 0.03 | 2.0 |
| 228 | AA630784 | thyroid hormone receptor interactor 13 | transcription | -1.26 | -2.55 | 1.3 |
| 229 | AI951084 | diacylglycerol kinase | signaling | -0.34 | 0.86 | -1.2 |
| 230 | H17883 | Kallmann syndrome 1 sequence | cell adhesion | -0.08 | -1.24 | 1.2 |
| 231 | AA970402 | cathepsin B | extracellular matrix | 1.66 | 0.67 | 1.0 |
| 232 | AA857437 | EST | unknown | 2.41 | 4.22 | -1.8 |

**Table 2. continued**

| No | Accession No. | Name | Biological process | Cancer | Normal | Ratio |
|---|---|---|---|---|---|---|
| 233 | H18932 | Kell blood group precursor | transport | 1.34 | 2.83 | -1.5 |
| 234 | AA455369 | sodium/hydrogen exchanger | transport | 0.60 | 2.18 | -1.6 |
| 235 | AA995128 | vascular endothelial growth factor D | angiogenesis. | -0.08 | 1.50 | -1.6 |
| 236 | AI160214 | EST | unknown | 1.14 | 2.58 | -1.4 |
| 237 | T70057 | immunoglobulin J polypeptide | Extracellular | 5.95 | 8.69 | -2.7 |
| 238 | AI651536 | protocadherin alpha 12 | cell adhesion | -1.06 | -0.08 | -1.0 |

**Table 3. Ontological information of 238 classifier genes**

| Biological process | Number of genes |
|---|---|
| angiogenesis | 2 |
| cell motility | 2 |
| apoptosis | 4 |
| cell cycle | 4 |
| proteolysis | 7 |
| cell adhesion | 11 |
| cell growth | 12 |
| transcription | 14 |
| signaling | 15 |
| transport | 15 |
| extracellular matrix | 17 |
| metabolism | 30 |
| Etc. | 37 |
| unknown | 68 |
| Total | 238 |

**4. Identification of lymph node metastasis related genes**

As the lymph node metastasis is an important clinical parameter for prognosis, we identified lymph node metastasis related genes among 238 classifiers. By comparison of mean expression ratio between patients with lymph node metastasis and without lymph node metastasis, we selected 66 lymph node metastasis related genes, which showed more than 0.5 in expression ratio. Among 66 genes, 25 genes were increased and 41 genes were decreased in patients who had lymph node metastasis (Table 4). 18 kDa antrum mucosa protein, which may protect the antral mucosa and promote healing, was the most increased gene in patients with lymph node metastasis. Among decreased genes in patients with lymph node metastasis, many genes were related to extracellur matrix components such as collagen type III, IV, VI, X, XI, fibronectin 1, and proteolysis related genes such as matrix metalloproteinase 11, tissue inhibitor of metalloproteinase 1, catepsin L, catepsin L2 and urokinase-type plasminogen activator.

**Table 4. Sixty-six Lymph node metastasis related genes among 238 classifier genes**

| Name | Lymph node (+) | Lymph node (-) | Ratio |
|---|---|---|---|
| **18 kDa antrum mucosa protein** | 2.65 | 0.64 | 2.01 |
| **pancreatitis-associated protein** | 3.53 | 1.56 | 1.97 |
| **regenerating islet-derived 1 beta** | 3.14 | 1.66 | 1.48 |
| **keratin 20** | 1.13 | -0.33 | 1.45 |
| **carbonic anhydrase II** | 2.10 | 0.73 | 1.37 |
| **secretoglobin, family 2A, member 1** | -1.38 | -2.64 | 1.26 |
| **creatine kinase, brain** | 0.20 | -1.05 | 1.25 |
| **regenerating islet-derived 1 alpha** | 2.78 | 1.72 | 1.06 |
| **Kell blood group precursor** | 1.65 | 0.74 | 0.91 |
| **chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)** | 2.07 | 1.20 | 0.87 |
| **cytochrome P450, family 3, subfamily A, polypeptide 4** | 2.07 | 1.21 | 0.87 |

Lymph node (+) denotes the mean log R/G of all patients with lymph node metastasis. Lymph node (-) denotes the mean log R/G of all patients without lymph node metastasis. Ratio denotes the expression ratio between two patients group.

**Table 4. continued**

| Name | Lymph node (+) | Lymph node (-) | Ratio |
|---|---|---|---|
| **Sapiens, clone IMAGE:4471726, mRNA** | 0.58 | -0.28 | 0.85 |
| **lipase, gastric** | 1.60 | 0.75 | 0.85 |
| **UDP glycosyltransferase 2 family, polypeptide B17** | 1.47 | 0.64 | 0.83 |
| **immunoglobulin heavy constant** | 3.13 | 2.32 | 0.82 |
| **aldo-keto reductase family 1, member B10 (aldose reductase)** | -4.21 | -5.02 | 0.82 |
| **cytochrome P450, family 3, subfamily A, polypeptide 5 pseudogene 2** | 1.62 | 0.82 | 0.81 |
| EST | 1.35 | 0.54 | 0.80 |
| **transmembrane protease, serine 2** | 2.74 | 1.95 | 0.79 |
| **Sapiens, clone IMAGE:5478062, mRNA** | 0.28 | -0.39 | 0.67 |
| **integral membrane protein 2A** | -1.19 | -1.85 | 0.66 |
| **hypothetical protein LOC143381** | -0.85 | -1.42 | 0.57 |
| **protein inhibitor of activated STAT protein PIASy** | -1.70 | -2.25 | 0.55 |
| **lactotransferrin** | 2.53 | 1.97 | 0.55 |
| **interferon regulatory factor 5** | 1.88 | 1.34 | 0.54 |

**Table 4. continued**

| Name | Lymph node (+) | Lymph node (-) | Ratio |
|---|---|---|---|
| **KIAA0882 protein** | -0.19 | 0.36 | -0.54 |
| **chemokine (C-X-C motif) ligand 14** | 1.93 | 2.48 | -0.55 |
| **cathepsin L** | -0.05 | 0.50 | -0.55 |
| **serine protease inhibitor, Kazal type 1** | 3.21 | 3.78 | -0.56 |
| **Sapiens clone FLB9440 PRO2550 mRNA** | -0.15 | 0.44 | -0.59 |
| **monoamine oxidase A** | 0.68 | 1.29 | -0.61 |
| **collagen, type V, alpha 2** | 1.28 | 1.92 | -0.64 |
| **leukemia inhibitory factor** | -1.41 | -0.70 | -0.71 |
| **plasminogen activator, urokinase** | -0.71 | 0.00 | -0.71 |
| **bone morphogenetic protein 1** | -0.41 | 0.32 | -0.73 |
| **glutathione S-transferase A4** | -0.15 | 0.58 | -0.73 |
| **collagen, type III, alpha 1** | 5.06 | 5.80 | -0.74 |
| **ribonucleotide reductase M2 polypeptide** | -1.11 | -0.37 | -0.74 |
| **carbonic anhydrase IX** | -0.05 | 0.69 | -0.74 |
| **mucin 5, subtype B, tracheobronchial** | 1.12 | 1.88 | -0.75 |

**Table 4. continued**

| Name | Lymph node (+) | Lymph node (-) | Ratio |
|---|---|---|---|
| **cytochrome P450, family 24, subfamily A, polypeptide 1** | -2.14 | -1.37 | -0.77 |
| **serine (or cysteine) proteinase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)** | 0.35 | 1.14 | -0.80 |
| **secreted protein, acidic, cysteine-rich (osteonectin)** | 1.45 | 2.25 | -0.80 |
| **fibroblast activation protein, alpha** | 1.65 | 2.46 | -0.81 |
| **matrix metalloproteinase 11** | 1.87 | 2.68 | -0.81 |
| **tissue inhibitor of metalloproteinase 1** | 0.67 | 1.51 | -0.84 |
| **collagen, type VI, alpha 3** | 0.66 | 1.51 | -0.86 |
| **chitinase 3-like 1** | 1.44 | 2.31 | -0.87 |
| **Kallmann syndrome 1 sequence** | -0.37 | 0.51 | -0.88 |
| **LOC150225** | 1.30 | 2.18 | -0.89 |
| **myosin IB** | -0.84 | 0.07 | -0.91 |
| **cathepsin L2** | -0.01 | 0.92 | -0.93 |
| **tumor necrosis factor, alpha-induced protein 6** | 1.72 | 2.68 | -0.96 |
| **cadherin 11, type 2, OB-cadherin (osteoblast)** | 1.43 | 2.45 | -1.03 |

**Table 4. continued**

| Name | Lymph node (+) | Lymph node (-) | Ratio |
|---|---|---|---|
| **thrombospondin 2** | 2.42 | 3.46 | -1.04 |
| **dermatopontin** | 0.70 | 1.75 | -1.05 |
| **collagen, type IV, alpha 1** | -0.20 | 0.87 | -1.07 |
| **inhibin, beta A (activin A, activin AB alpha polypeptide)** | 0.93 | 2.13 | -1.20 |
| **sulfatase 1** | 5.38 | 6.65 | -1.27 |
| **Homo sapiens transcribed sequences** | 1.64 | 3.13 | -1.49 |
| **collagen, type X, alpha 1** | 2.43 | 4.06 | -1.63 |
| **collagen, type XI, alpha 1** | 2.42 | 4.22 | -1.80 |
| **Homo sapiens transcribed sequence with weak similarity to protein ref:NP_060265.1 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]** | -0.73 | 1.34 | -2.07 |
| **secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)** | -2.82 | -0.69 | -2.14 |
| **melanoma cell adhesion molecule** | -1.56 | 0.86 | -2.41 |
| **fibronectin 1** | -1.47 | 1.12 | -2.59 |

# IV. DISCUSSION

Gastric cancer is a multi-step disease with multiple genetic alterations, and shows various individual properties such as different stage, tumor size, depth of invasion, growth pattern and metastasis. However, these clinical parameters do not accurately reflect the prognosis of gastric cancer patients. Although recent advanced therapeutic technologies have provided a better prognosis, gastric cancer remains one of the main causes of cancer-related death in Asia. Regardless of several reports upon genetic alterations in gastric cancer, the molecular mechanisms underlying cancer are not well understood due to lack of sufficient genetic information. Therefore, the necessity of genome wide analysis has been concerned to identify molecular signatures of gastric cancer. Microarray, an effective tool for the analysis of thousands of genes simultaneously, contributed for sub-classification based on pathological, histological difference, clinical features and molecular signatures in various

cancer with sufficient sensitivity and specificity[30-32]. Hence, we undertook to comprehensively investigate the gastric cancer genome using a high-density cDNA microarray.

After normalization and filtering, we compared expression profiles of all the cancer and normal tissues using unsupervised two-way hierarchical clustering to identify the molecular portrait of the gastric cancer. Although there was a trend of discrimination between cancer from normal tissue, whole genome wide expressions were not sufficient to identify cancer specific signature probably due to contained noises from too much information (data not shown). To reduce the noises from genome wide expression, the optimal statistic and systemic method for the significant gene selection were considered as the most important issue.

To obtain verified classifiers, an adequate experiment scheme and systemic data analysis are necessary. To start with optimal training, we performed training and cross-validation with 11paired samples. The optimal training was

done by cross-validation error rate (data not shown).

We used PAM for gene selection method, which have advantage in removal of the noisy genes and could process systemic training, cross-validation and class prediction. After training the training set using 12,856 genes, we considered some parameters for choosing the amount of shrinkage to select classifiers. First, the number of genes should be manageable. Thousands of genes contain many noisy genes, which is lack of specificity and insufficient for further data analysis. Second, classifier genes should have the maximum cross-validation probability in the training set and the hierarchical clustering pattern should clearly discriminated cancer and normal. Although the hierarchical clustering has different algorithm with PAM, it could support different classes. We used 11 pairs as a training set, which might be considered as insufficient number of samples for cancer identification. By maximizing resolving power and combining two different algorithms, we tried to overcome the small number of training samples.

After the gene selection, we verified the specificity of 238 classifiers by performing class prediction with test samples. For strict verification, it is important that the test samples should be independent to training samples. Although previous studies[11,19,23] reported significant genes, the test samples were not independent to training samples. For class prediction, we used the Gaussian linear discriminant analysis. Briefly, the method computes a standardized centroid for each class. The class of new sample is determined by comparing the expression profile of new samples with each of class centroid. As shown in Figure 8, the class prediction result proved our gene selection was in the acceptable range.

The 238 classifiers showed various biological processes. It is remarkable that many genes are related to known biological behavior of cancer and showed reproducible expression level with many previous reports [33-40].

The lymph node metastasis related genes, which were selected from 238 classifiers, were mostly related to extracellur matrix such as composition,

proteolysis and interactions. As related to metastasis process, these genes may contribute to the pro-metastastic environment. Briefly, fibronectin 1 is involved in cell adhesion and migration processes including embryogenesis, wound healing, blood coagulation, host defense and metastasis[41]. Dermatopontin is an extracellular matrix protein with possible functions in cell-matrix interactions and matrix assembly[42].

Moreover, we compared our result with previous microarray studies in gastric cancer. Hippo *et al*[16], Inoue *et al* [19], and Hasegawa *et al*[23]. reported various numbers of differentially expressed genes in gastric cancer using microarray. Compare to those genes, we found that 18 genes were also significantly expressed in our data with 12 up-regulated and 6 down-regulated genes (Table 5). These genes were mostly selected as significant genes in the previous studies and current study, despite of the different number of genes on array, number of samples and the gene selection methods used. As we can assume that these common genes might be more valuable genes than other

selected genes, we verified these 18 genes by performing class prediction in 36 samples. As shown in Figure 9, we had one misclassified cancer sample among 36 samples, which represent 2.77% class prediction error rate with these common 18 gastric cancer specific genes. Considering the biological function of 18 overlapping genes, these genes are mostly related to cancer invasion and metastasis. Cathepsin B and urokinase-type plasminogen activator, cysteine and serine protease respectively, showed increased expression levels in early gastric cancer and both are thought to play important roles in cancer invasion, progression, metastasis and prognosis[43]. Trefoil factor 1 and 2 are members of trefoil factor family constitutively expressed in the gastric mucosa, where they play a role in intestinal mucosal defense and repair. They have been suggested as potential targets for therapeutic intervention in gastric cancer[44].

**Table 5. Commonly identified differentially expressed genes between previous reports and current 238 classifiers**

| | Gene name | Changes of expression |
|---|---|---|
| Hippo et al.(16) | Collagen type V alpah 12 | Up |
| | Collagen type I alpah 2 | Up |
| | Collagen type III alpha 1 | Up |
| | Collagen type IV alpha 1 | Up |
| | Collagen type VI alpah 2 | Up |
| | Matrix metalloproteinase 11 | Up |
| | Mesothelin | Up |
| | Plasminogen activiator, urokinase | Up |
| | Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3 | Up |
| Inoue et al.(19) | Cathepsin B | Up |
| | Collagen type IV alpha 1 | Up |
| | Collagen type I alpha 2 | Up |
| | Collagen type III alpha 1 | Up |
| Hasegawa et al.(23) | ATPase, Ca++ transporting | Down |
| | Carbonic anhydrase II | Down |
| | Ectodermal-neural cortex | Up |
| | Metallothionein 1E | Down |
| | Progastricsin | Down |
| | Solute carrier family 16 | Up |
| | Trefoil factor 1 | Down |
| | Trefoil factor 2 | Down |

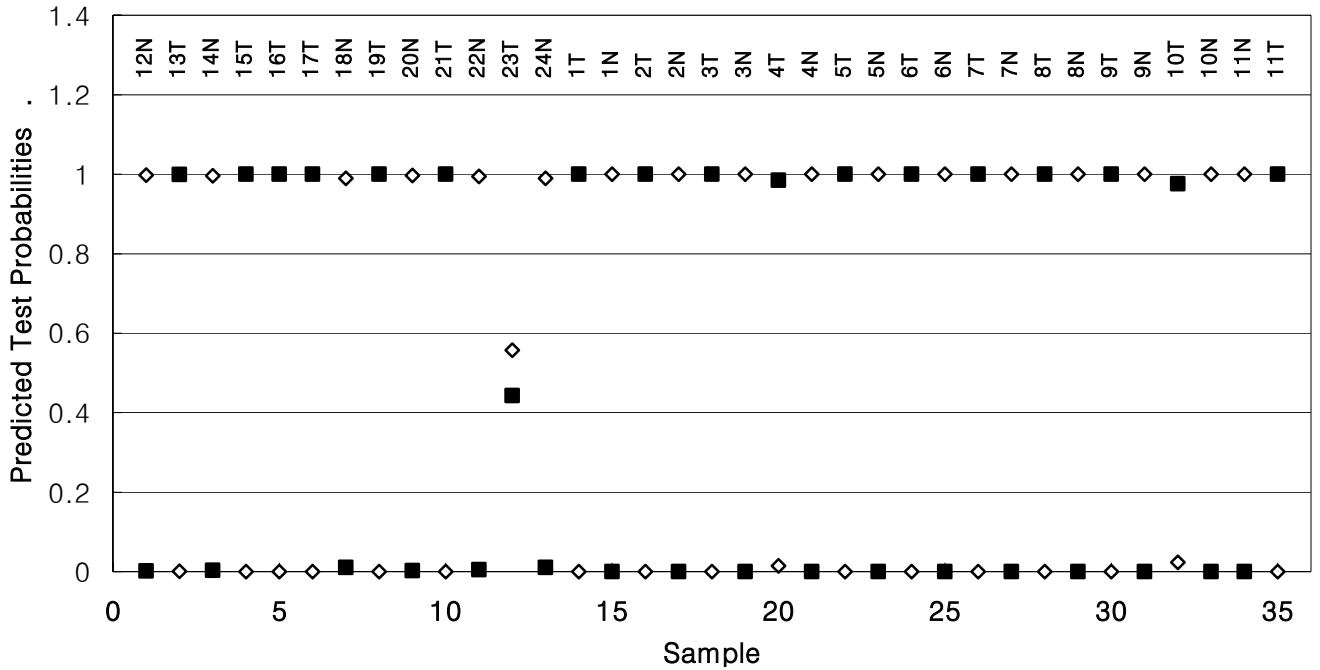


**Figure 9. Verified probabilities of common 18 genes.** The class predictability with total 35 samples using 18 commonly expressed genes was 97.2%. Normal score (white) and tumor score (black) were voted for each sample. The higher score decided the class of samples and the probability represented the distance between two scores. N: normal, T: tumor.

# V. CONCLUSION

In conclusion, we identified and verified 238 classifiers, which discriminate gastric cancer and normal tissues. The classifiers contained 66 lymph node metastasis related genes and 18 commonly expressed genes compare to previous studies. Our results suggest that extracellur matrix status is an important parameter to understand gastric cancer progression, lymph node metastasis and may supply valuable information to understand gastric cancer pathophysiology and candidate for biomarker.

**REFERECE**

1. Neugut AI, Hayek M., and Howe G. Epidemiology of gastric cancer. Semin Oncol 1996;23:281-291.

2. Roukos DH. Current status and future perspectives in gastric cancer management. Cancer Treat Rev 2000;26:243-255.

3. Yokozaki H, Yasui W, Tahara E. Genetic and epigenetic changes in stomach cancer. Int Rev cytol 2001;204:49-95.

4. Park WS, Oh RR, Park JY, Lee SH, Shin MS, Kim YS et al. Frequent somatic mutations of the β-catenin gene in intestinal-type gastric cancer. Cancer Res 1999;59:4227-4260.

5. Berx G, Becker KF, Hofler H, van Roy F. Mutations of the human E-cadherin (CDH1) gene. Hum Mutat 1998;12:226-237.

6. Park WS, Oh RR, Park JY, Lee JH, Shin MS, Kim HS et al. Somatic mutations of the trefoil factor family 1 gene in gastric cancer. Gastroenterology 2000;119:691-698.

7. Lee JH, Han SU, Cho H, Jennings B, Gerrard B, Dean M et al. A novel germ line juxtamemberane Met mutation in human gastric cancer. Oncogene 2000;19:4947-4953.

8. van Berkum NL, Holstege FC. DNA microarrays: raising the profile. Curr Opin Biotechnol 2001;12:48-52.

9. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403;503-511.

10. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, Frederiksen CM, Laiho P, Aaltonen LA et al. Gene expression in colorectal cancer. Cancer Res 2002;62:4352-4363.

11. Tay ST, Leong SH, Yu K, Aggarwal A, Tan SY, Lee CH et al. A combined comparative genomic hybridization and expression microarray analysis of gastric cancer reveals novel molecular subtypes. Cancer Res 2003;63:3309-3316.

12. Lee S, Baek M, Yang H, Bang Y, Kim WH, Ha JH et al. Identification of genes differentially expressed between gastric cancers and normal gastric mucosa with DNA microarrays. Cancer Lett 2002;184:197-206.

13. Liu LX, Liu ZH, Jiang HC, Qu X, Zhang WH, Wu LF et al. Profiling of differentially expressed genes in human gastric carcinoma by cDNA expression array. World J Gastroenterol 2002;8:580-585.

14. Merireles SI, Carvalho AF, Hirata R, Montagnini AL, Martins WK, Runza FB et al. Differentially expressed genes in gastric tumors identified by cDNA array. Cancer Lett 2003;190:199-211.

15. El-Rifai W, Frierson HF Jr, Harper JC, Powell SM, Knuutila S. Expression profiling of gastric adenocarcinoma using cDNA array. Int J cancer 2001;92:832-838.

16. Hippo Y, Taniguchi H, Tsutsumi S, Machida N, Chong JM, Fukayama M et al. Global gene expression analysis of gastric cancer by oligonucleotide microarrays. Cancer Res 2002;62:233-240.

17. Boussioutas A, Li H, Liu J, Waring P, Lade S, Holloway AJ et al. Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. Cancer Res 2003;63:2569-2577.

18. Sakakura C, Hagiwara A, Nakanishi M, Shimomura K, Takagi T, Yasuoka R et al. Differential gene expression profiles of gastric cancer cells established from primary tumor and malignant ascites. Br J cancer 2002;87:1153-1162.

19. Inoue H, Matsuyama A, Mimori K, Ueo H, Mori M. Prognostic score of gastric cancer determined by cDNA microarray. Clin Cancer Res 2002;8 3475-3479.

20. Wang J, Chen S. Screening and identification of gstric adenocarcinoma metastasis-related genes using cDNA microarray coupled to FDD-PCR. J Cancer Res Clin Oncol 2002;128:547-553.

21. Suganuma K, Kubota T, Saikawa Y, Abe S, Otani Y, Furukawa T et al. Possible chemoresistance-related genes for gastric cancer detected by cDNA microarray. Cancer Sci 2003;94:355-359.

22. Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D et al. Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene 2003;22:1872-1879.

23. Hasegawa S, Furukawa Y, Li M, Satoh S, Kato T, Watanabe T et al. Genome-wide analysis of gene expression in intestinal-type gastric cancers using a complementary DNA micrarray representing 23,040 genes. Cancer Res 2002;62:7012-7017.

24. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees C et al. Molecular portraits of human breast tumors. Nature 2000;406:747-752.

25. Leung SY, Chen X, Chu KM, Yusen ST, Mathy J, Ji J et al. Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. Proc Natl Acad Sci USA 2002;99:16203-16208.

26. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J et al. Normalization for cDNA microarray data: a robust composite method addressing single and

multiple slide systematic variation. Nucleic Acids Res 2002;30: e15.

27. Troyanskaya O, Cantor M, Sherlock G., Brown P, Hastie T, Tibshirani R et al. Missing value estimation methods for DNA microarrays. Bioinformatics 2001;17:520-525.

28. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci USA 2002;99:6567-6572.

29. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide patterns. Proc Natl Acad Sci USA 1998;95:14863-14868.

30. Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression pfofiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 2001;61:3124-3130.

31. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with

clinical implications. Proc Natl Acad Sci USA 2001;98:10869-10874.

32. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, Mclaughlin ME et al. Prediction of central nervous system embryonal tumor outcome based on gene expression. Nature 2002;415:436-442.

33. Shanbhag SA, Sheth AR, Nanivadekar SA, Sheth NA. Immunoreactive inhibin-like material in serum and gastric juice of patients with benign and malignant diseases of the stomach. Br J Cancer 1985;51:877-882.

34. Aitkenhead M, Wang SJ, Nakatsu MN, Mestas J, Heard C, Hughes CC. Identification of endothelial cell genes expressed in an in vitro model of angiogenesis: induction of ESM-1, (beta) IG-h3, and NrCAM. Microvasc Res 2002;63:159-171.

35. Lee JH, Koh JT, Shin BA, Ahn KY, Roh JH, Kim YJ, Kim KK. Comparative study of angiostatic and anti-invasive gene expressions as prognostic factors in gastric cancer. Int J Oncol 2001;18:355-361.

36. Yoshikawa Y, Mukai H, Hino F, Asada K, Kato I. Isolation of two novel

genes, down-regulated in gastric cancer. Jpn J Cancer Res 2000;91:459-463.

37. Majima T, Ichikura T, Takayama E, Chochi K, Mochizuki H. Detecting circulating cancer cells using reverse transcriptase-polymerase chain reaction for cytokeratin mRNA in peripheral blood from patients with gastric cancer. Jpn J Clin Oncol 2000;30:499-503.

38. Papotti M, Cassoni P, Volante M, Deghenghi R, Muccioli G., Ghigo E. Ghrelin-producing endocrine tumors of the stomach and intestine. J Clin Endocrinol Metab 2001;86:5052-5059.

39. Tartaglia A, Bianchini S, Vezzadini P. Biochemical diagnosis of gastroenteropancreatic endocrine tumors. Minerva Med 2003;94:1-7.

40. Saku T, Sakai H, Tsuda N, Okabe H, Kato Y, Yamamoto K. Cathepsins D and E in normal, metaplastic, dysplastic, and carcinomatous gastric tissue: an immunohistochemical study. Gut 1990;31:1250-1255.

41. David L, Nesland JM, Holm R, Sobinho-Simoes M. Expression of laminin, collagen IV, fibronectin, and type IV collagenase in gastric carcinoma. An

immunohistochemical study of 87 patients. Cancer 1994;73:518-527.

42. Forbes EG., Cronshaw AD, MacBeath, JR, Hulmes DJ. Tyrosin-rich acidic matrix protein (TRAMP) is a tyrosine-sulphated and widely distributed protein of the extracellular matrix. FEBS Lett 1994;351:433-436.

43. Farinati F, Herszenyi L, Plebani M, Carraro P, De Paoli M et al. Increased levels of cathepsin B and L, urokinase-type plasminogen activator and its inhibitor type-1 as an early event in gastric carcinogenesis. Carcinogenesis 1996;17:2581-2587.

44. Dhar DK, Wang TC, Maruyama R, Udagawa J, Kubota H, Fuji T et al. Expression of cytoplasmic TFF2 is a marker of tumor metastasis and negative prognostic factor in gastric cancer. Lab Invest 2003;83:1343-1352.

**Abstract (in korean)**

## 고밀도 마이크로어레이를 이용한 위암에서 유의한 유전자의 발견


**<지도교수 라 선 영>**

**연세대학교 대학원 의과학과**

**박 세 원**



본 연구에서는 위암에서 분자적 신호를 밝히기 위하여 17,000개의 유전자가 점삭된 cDNA microarray로 18경우의 위암조직과 17경우의 정상 위조직간의 expression profile을 비교하였다. 실험에 사용된 35경우의 조직들은 Normalization과 filtering후, 짝이 있는 11경우는 training군으로, 짝이 없는 7경우의 위암조직과 6경우의 정상 조직은 test군으로 분류 되었다. 유의한 유전자는 training군에서 선택 되었고, 선택된 유전자의 유의도는 test군에서 평가되었으며 유의한 유전자는 'nearest shrunken centroid' 방법을 통하여 선별되었다. 위암조직과 정상조직을 구분하는 238개의 classifier는 training군에서 가장 높은

cross-validation probability와 명확한 hierarchical clustering 결과를 보였고, 독립적인 test군에서 높은 class prediction probability를 보였다. 선택된 238개의 classifier들은 대부분 암의 생물학적 현상와 관련되어 있으며, 28%의 유전자는 아직 그 역할이 명확하게 밝혀지지 않고 있다. 본 연구를 통하여 우리는 위암에서 genome-wide한 유전자 발현 정보를 얻을 수 있었다. 이러한 정보는 위암의 병태생리를 이해함에 있어서, 또한 위암에서의 생물학적 표적을 발견하기에 초석이 되는 유전정보를 제공 할 것이다.

Key words: 위암, cDNA microarray, expression profile, classifier genes