

# Clinical Cancer Research



## Determination of Genes Related to Gastrointestinal Tract Origin Cancer Cells Using a cDNA Microarray

Tae Moon Kim, Ha Jin Jeong, Min Young Seo, et al.

*Clin Cancer Res* 2005;11:79-86.

**Updated version** Access the most recent version of this article at:  
<http://clincancerres.aacrjournals.org/content/11/1/79>

**Cited Articles** This article cites by 26 articles, 13 of which you can access for free at:  
<http://clincancerres.aacrjournals.org/content/11/1/79.full.html#ref-list-1>

**Citing articles** This article has been cited by 3 HighWire-hosted articles. Access the articles at:  
<http://clincancerres.aacrjournals.org/content/11/1/79.full.html#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, contact the AACR Publications Department at [permissions@aacr.org](mailto:permissions@aacr.org).

# Determination of Genes Related to Gastrointestinal Tract Origin Cancer Cells Using a cDNA Microarray

Tae Moon Kim,<sup>1</sup> Ha Jin Jeong,<sup>1,2</sup> Min Young Seo,<sup>1</sup>  
Sang Chul Kim,<sup>1,2</sup> Gabee Cho,<sup>1</sup> Chan Hee Park,<sup>1,2</sup>  
Tae Soo Kim,<sup>1</sup> Kyu Hyun Park,<sup>1</sup> Hyun Cheol  
Chung,<sup>1,2</sup> and Sun Young Rha<sup>1,2</sup>

<sup>1</sup>Cancer Metastasis Research Center, Yonsei Cancer Center and <sup>2</sup>Brain Korea 21 Project for Medical Science, Yonsei University College of Medicine, Seoul, Korea

## ABSTRACT

**Purpose:** We evaluated the genome-wide gene expression profiles of various cancer cell lines to identify the gastrointestinal tract cancer cell-related genes.

**Experimental Design:** Gene expression profilings of 27 cancer cell lines and 9 tissues using 7.5K human cDNA microarrays in indirect design with Yonsei reference RNA composed of 11 cancer cell line RNAs were done. The significant genes were selected using significant analysis of microarray in various sets of data. The selected genes were validated using real-time PCR analysis.

**Results:** After intensity-dependent, within-print-tip normalization by loess method, we observed that expression patterns of cell lines and tissues were substantially different, divided in two discrete clusters. Next, we selected 115 genes that discriminate gastrointestinal cancer cell lines from others using significant analysis of microarray. Among the expression profiles of five gastric cancer cell lines, 66 genes were identified as differentially expressed genes related to metastatic phenotype. YCC-16, which was established from the peripheral blood of one advanced gastric cancer patient, produced a unique gene expression pattern resembling the profiles of lymphoid cell lines. Quantitative real-time reverse transcription-PCR results of selected genes, including *PXN*, *KRT8*, and *ITGB5*, were correlated to microarray data and successfully discriminate the gastrointestinal tract cancer cell lines from hematologic malignant cell lines.

**Conclusions:** A gene expression database could serve as a useful source for the further investigation of cancer biology using the cell lines.

## INTRODUCTION

Established cancer cell lines actually differ from their tissues of origins, which are composed of various cell types in histologic or biological characters. However, because of the difficulty in accessing human cancer and normal tissues, cell lines are still the most useful tools for cancer research. Many efforts have been made to understand the basic biological and genetic characteristics of cell lines, which might be useful to determine the proper cell line model for the study purposes. Sixty National Cancer Institute (NCI) cancer cell lines have been used for drug screening for more than four decades, which significantly contributed to anticancer drug development (1–4). Based on their biological properties, the NCI's Developmental Therapeutics Program did a comprehensive gene expression analysis of these 60 NCI cell lines, which resulted in the identification of the genetic characteristics of each cell line (5). Additional studies suggested that drug sensitivity genes could be identified by evaluating the correlation between the gene expression profiles and the drug responsiveness in each cell line (6, 7).

Recent microarray technology has proven to be useful in cancer research, including cancer cell line characterization. Ji et al. (8) used a cDNA microarray to analyze the gene expression profile of 12 gastric cancer cell lines that were not included in the NCI studies. This study raised the possibility that RF1 and RF48 designated as gastric cancer cell lines from the American Type Culture Collection (Manassas, VA) could be misclassified. In addition, Sakakura et al. (9) reported on the differential gene expression profiles in gastric cancer cell lines established from primary tumor and malignant ascites. Furthermore, Virtanen et al. (10) showed that gene expression profiling, which is applicable to both lung cancer cell lines and lung tumors, could provide the technology to reclassify cell lines.

Although current treatments have significantly improved patient survival, cancers of the gastrointestinal tract remain the most common cause of cancer-related deaths in the world. However, the evaluation of individual molecules failed to elaborate the complex patterns of carcinogenesis and cancer progression in gastrointestinal cancer. Recent reports suggest that the phenotypic diversity of cancer might be related to corresponding diversity in their gene expression patterns (11–13).

We did a comprehensive gene expression analysis of various cancer cell lines with different tissue origins using a high-density cDNA microarray. Our goals in this study were to identify a subset of classifier genes that distinguish cancer cell lines of gastrointestinal tract origin from other cell lines and to identify genes differently expressed in five gastric cancer cell lines. In addition, because the expression pattern of known genes can show novel phenotypic aspects of the cell lines and tissues (11–13), we tried to explore the molecular characteristics of cancer cell lines with respect to their metastatic behavior to identify optimal cell lines for given biological experiments.

Received 2/17/04; revised 8/30/04; accepted 10/6/04.

**Grant support:** Korea Science and Engineering Foundation through the Cancer Metastasis Research Center at Yonsei University College of Medicine.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Requests for reprints:** Sun Young Rha, Cancer Metastasis Research Center, Yonsei Cancer Center, Yonsei University College of Medicine, 134 Shinchon-Dong, Seodaemun-Ku, Seoul, Korea 120-752. Phone: 82-2-361-7650; Fax: 82-2-362-5592; E-mail: rha7655@yumc.yonsei.ac.kr.

©2005 American Association for Cancer Research.

## MATERIALS AND METHODS

**Cell Lines and Clinical Samples.** Twenty cancer cell lines of various tissue origins (stomach: AGS; breast: MDA-MB-231, MDA-MB-435, and MCF7; colon: HCT116, COLO 205, and HT-29; liver: SK-HEP-1 and HepG2; lung: A549 and NCI-H460; lymphoid organ: HL-60, MOLT-4, and Raji; cervix: HeLa; fibrosarcoma: HT-1080; kidney: Caki-2; brain: U-87 MG; melanoma: SK-MEL-2; and pancreas: Capan-2) were obtained from the American Type Culture Collection. Seven cancer cell lines that were established at Yonsei Cancer Metastasis Research Center (CMRC, Seoul, Korea), including YCC-B1, YCC-B2 (from pleural effusion of breast cancer patients), YCC-2, YCC-3, YCC-7 (from ascites of gastric cancer patients), YCC-16 (from the blood of gastric cancer patient), and YCC-P1 (pancreas), were also studied. The cells were cultured and maintained in MEM with 10% fetal bovine serum (Life Technologies, Rockville, MD) in 100 units/mL penicillin and 0.1 mg/mL streptomycin (Life Technologies) at 37°C in a 5% CO<sub>2</sub> incubator. Nine tissue samples (three colon cancer tissues, two normal colon tissues, three liver metastatic tumor tissues from colon cancer, and a single sample of normal liver tissue) were collected from patients who had undergone surgery at the Severance Hospital, Yonsei University College of Medicine. Tissue samples were immediately frozen in liquid nitrogen and stored at -80°C until further use.

**RNA Preparation and Purification.** Total RNA was extracted from the tissues and the cell lines using Trizol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. In the case of tissue RNA, extracted RNA was purified before probe preparation using a RNeasy kit (Qiagen, Hilden, Germany) based on the supplier's manuals. The quantity and quality of RNA were evaluated using a GeneSpec III (Hitachi, Tokyo, Japan) and a Gel Documentation-Photo System (Vilber Lourmat, France), respectively.

**Probe Preparation.** Total RNA (50 µg) was directly labeled and transcribed to cDNA. We combined total RNA from the following 11 cancer cell lines of various tissue origins in equal quantities to prepare the Yonsei reference RNA (CMRC, Yonsei University College of Medicine): stomach, breast, colon, liver, lung, lymphoid and myeloid system, cervix, fibrous tissue, kidney, and brain cancer cell line. Cell lines were selected from various organs to ensure that the pooled RNA contained as many transcripts as possible (see below). Yonsei reference RNA was labeled with Cy3-dUTP (NEN Co., Boston, MA) and test samples were labeled with Cy5-dUTP (NEN).

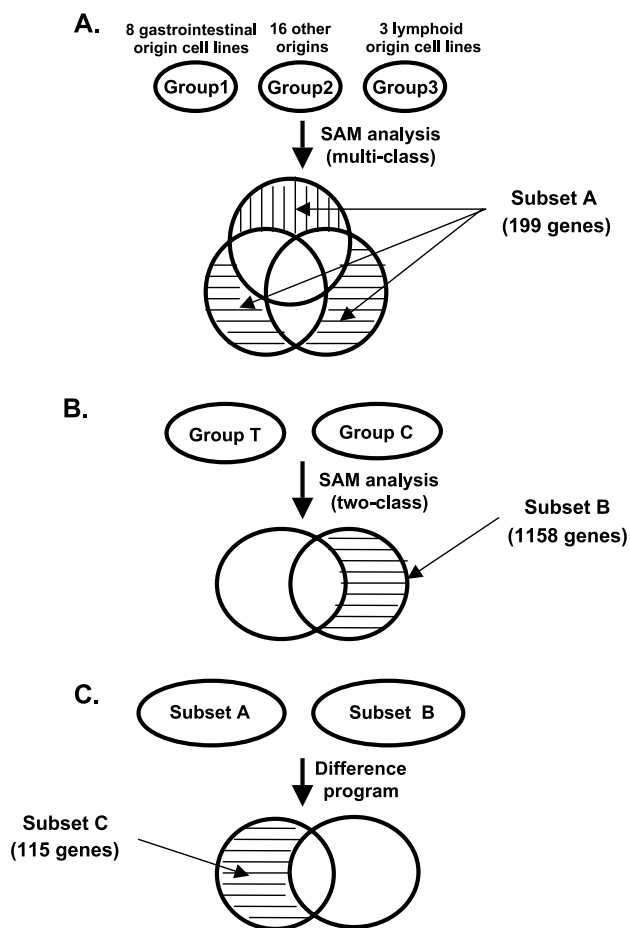
The labeling was done at 42°C for 2 hours in a total volume of 30 µL containing 400 units SuperScript II (Life Technologies); 3 µL Cy5-dUTP (or Cy-3 dUTP), 1.5 µL of each of dATP, dCTP, and dGTP; 0.6 µL dTTP; 300 mmol/L; 6 µL of 5× first-strand buffer; and 4 µg of modified oligo(dT) primer. Unincorporated nucleotide was removed by using a PCR purification kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Eluted probes were then mixed and supplemented with 20 µL of 1 µg/µL human Cot1 DNA (Life Technologies), 2 µL of 10 µg/µL polyadenylate RNA (Sigma, St. Louis, MO), and 2 µL of 10 µg/µL and 288 µL of 1 mol/L TE buffer. This probe mixture was concentrated using a Microcon-30 tube (Millipore, Bedford, MA), and the labeled mixture

(48 µL) was mixed with 10.2 µL of 20× SSC and 1.8 µL of 10% SDS for hybridization.

**Hybridization of Fluorescence-Labeled cDNA.** We used the 7.5K human cDNA microarray (GenomicTree Co., Daejeon, Korea), which contains 6,360 known genes and 152 expressed sequence tags. Before the hybridization, slides were preblocked in 10 mg/mL bovine serum albumin, 3.5× SSC, 0.1% SDS solution to prevent nonspecific hybridization, and probe mixtures were heated at 95°C for 2 minutes and centrifuged at 13,000 rpm for 2 minutes. After applying the probe to the slides, the slides were hybridized in hybridization chambers (GenomicTree) at 65°C for 16 hours. After the hybridization, slides were washed in 2× SSC for 10 minutes, transferred to 0.1× SSC and 0.1% SDS for 10 minutes, and rinsed twice with 0.1× SSC for 10 minutes. After washing, slides were spun at 600 rpm for 5 minutes (Hanil Science Industrial Co., Incheon, Korea). Hybridized slides were scanned using a GenePix 4000B (Axon Instruments, Union City, CA) and the images were analyzed using GenePix Pro3.0 (Axon Instruments).

**Data Analysis.** All the array data were normalized by intensity-dependent, within-print-tip normalization with lowest fit (14, 15). We then selected genes with <20% missing values among the 27 experiments, leaving 3,658 genes for further analysis. Of the remaining genes, we determined a subset of classifier genes to distinguish gastrointestinal origin from other cell lines using three steps (Fig. 1). First, we chose a subset of genes to classify such origin of cell lines using multiclass significant analysis of microarray (SAM; refs. 16, 17) algorithm (origin classifier genes; Fig. 1A). For group 1 of gastrointestinal tract origin cell lines, we used four gastric cancer (AGS, YCC-2, YCC-3, and YCC-7) and three colorectal cancer (HCT116, COLO 205, and HT-29) cell lines. YCC-16 cell line was not used in any of gene selection procedure, as the cell line was established from the blood of the gastric cancer patient. Group 3 was the hematologic cells, including Raji, HL-60, and MOLT-4, and the group 2 was consisted of the residual 16 cell lines. Next, we selected a subset of classifier genes specific to cell lines by using two-class SAM (genes related to the cell line establishment; Fig. 1B). Thereafter, by using our Difference Program (CMRC, Yonsei College of Medicine, Korea), we removed a subset of genes specific for cell lines from a subset of classifier genes specific for tissue origin (Fig. 1C). The Difference Program based on Python (<http://www.python.org>) was developed by CMRC to easily and rapidly obtain the complement set of genes, which only belong to the specific group of the samples. Using Pearson correlation, we did two-way hierarchical clustering of selected genes and samples. TreeView was used to visualize the results using the GeneSpring program (Silicon Genetics, Inc., Redwood City, CA). Functional annotation is based on Stanford Web site (<http://genome-www.stanford.edu/cgi-bin/source/sourceSearch>).

**Real-time Reverse Transcription-PCR Analysis.** For validation of microarray data, we did real-time reverse transcription-PCR (RT-PCR) using 16 cell line RNAs, which were used for microarray experiment. Sixteen cell lines tested were YCC-B2, MDA-MB-231, MCF7, HCT116, COLO 205, HT-29, SK-HEP-1, HepG2, HL-60, MOLT-4, Raji, AGS,



**Fig. 1** Scheme of finding a subset of differentially expressed genes in gastrointestinal cancer cell lines from other cell lines derived from the other tissue origins. *A*, selection of a subset of origin classifier genes unique to groups 1, 2, and 3. (group 1, cell lines of gastrointestinal tract origin, including AGS, YCC-2, YCC-3, YCC-7, HCT116, COLO 205, and HT-29; group 2, other 16 cell lines, except groups 1 and 3 among 27 cell lines; group 3, hematologic cells, including Raji, HL-60, and MOLT-4). YCC-16 cell line was not used in any of gene selection procedure, as the cell line was established from the blood of the gastric cancer patient. *B*, selection of a subset of genes unique to cell lines (*group T*, three colon cancer tissues; *group C*, colon cancer cell lines HCT116, COLO205, and HT-29). *C*, selection of gastrointestinal tract origin cell line classifier genes, a subset of genes that distinguish gastrointestinal cancer cell lines from other cancer cell lines with different origins. Difference program was developed at the CMRC, Yonsei University College of Medicine.

YCC-2, YCC-3, YCC-7, and YCC-16. Six genes were randomly chosen from the selected gene set: dystroglycan 1 (*DAG1*, AA496691), integrin  $\beta_5$  (*ITGB5*, AA434397), keratin 8 (*KRT8*, AA598517), neuregulin 1 (*NRG1*, R72075), paxillin (*PXN*, AA430574), and suppression of tumorigenicity 14 (*STI4*, AA489246). The primer sequences of six genes and  $\beta$ -actin were as follows as in forward/reverse sequences: *DAG1* (268 bp) AGTATCCACACAAAACCAG/CTTCACCTCAAAGTAGGTGC, *ITGB5* (244 bp) CTGTCCATGAAGGATGACTT/TGTCCACTCTGTCTGTGAGA, *KRT8* (259 bp) ACAAGTTTGCCTCCTTCATA/CGCTTATTGATCTCATCCTC, *NRG1* (275 bp) AGTATCCACAGAAGGAG-

CAA/ATTCAACATGATGCAACAAA, *PXN* (159 bp) CGCTCTGTTTATAGTGACCC/AATCACAGGAATTGAAATGG, *STI4* (145 bp) CTCCACTGAGTTTGTAAAGCC/AGACCAGTAGTAGGCGATGA, and  $\beta$ -actin (98 bp) GGGAATTCAAACCTGGAACGGTGAAGG/GGAAGCTTATCAAAGTCCTCGGCCACA.

After the cDNA production from 5  $\mu$ g of total RNA in reaction with oligo(dT) primer and 40 IU Moloney murine leukemia virus reverse transcriptase (MBI Fermentas Hanover, MD), 2  $\mu$ L of cDNA of each cell line were used for the real-time PCR assay. The total volume of the reaction mixture was 20  $\mu$ L, which contained HotstarTaq DNA polymerase, QuantiTect SYBR Green PCR buffer, deoxynucleotide triphosphate mix, including dUTP, SYBR Green, 10  $\mu$ L of QuantiTect SYBR Green PCR kit including 2.5 mmol/L  $MgCl_2$  (Qiagen, Valencia, CA), 2  $\mu$ L of the cDNA, and 20 pmol of each primer in distilled water. PCR was done at 95°C for 15 minutes to activate the HotstarTaq DNA polymerase and then for 35 cycles of amplification at 95°C for 20 seconds, 50°C for 30 seconds, 72°C for 45 seconds on a Rotor Gene 2072D real-time PCR machine (Corbett Research, New South Wales, Sydney, Australia). The amplified fluorescence signal in each specimen was measured at the late extension step of each cycle.

To quantify the level of gene expression, we evaluated the  $\beta$ -actin gene expression in 10-fold serially diluted human genomic DNA (Promega, Madison, WI). The standard curve was drawn by plotting the measured threshold cycle versus the arbitrary unit of the copies/reaction according to the  $\beta$ -actin gene expression of diluted genomic DNA. The threshold cycle (*C<sub>t</sub>*) value was determined as the cycle number at which the fluorescence exceeded the threshold value. In the negative control, there was no fluorescent signal when the cycle number was increased to 35. To compare with microarray result, the calculated copies of each gene were divided with the copies of  $\beta$ -actin and then were log converted. Pearson correlation coefficient was evaluated using S-PLUS 2000 (MathSoft, Inc., Seattle, WA) and the plotting was done using SigmaPlot 8.0 (SPSS, Inc., Chicago, IL).

## RESULTS

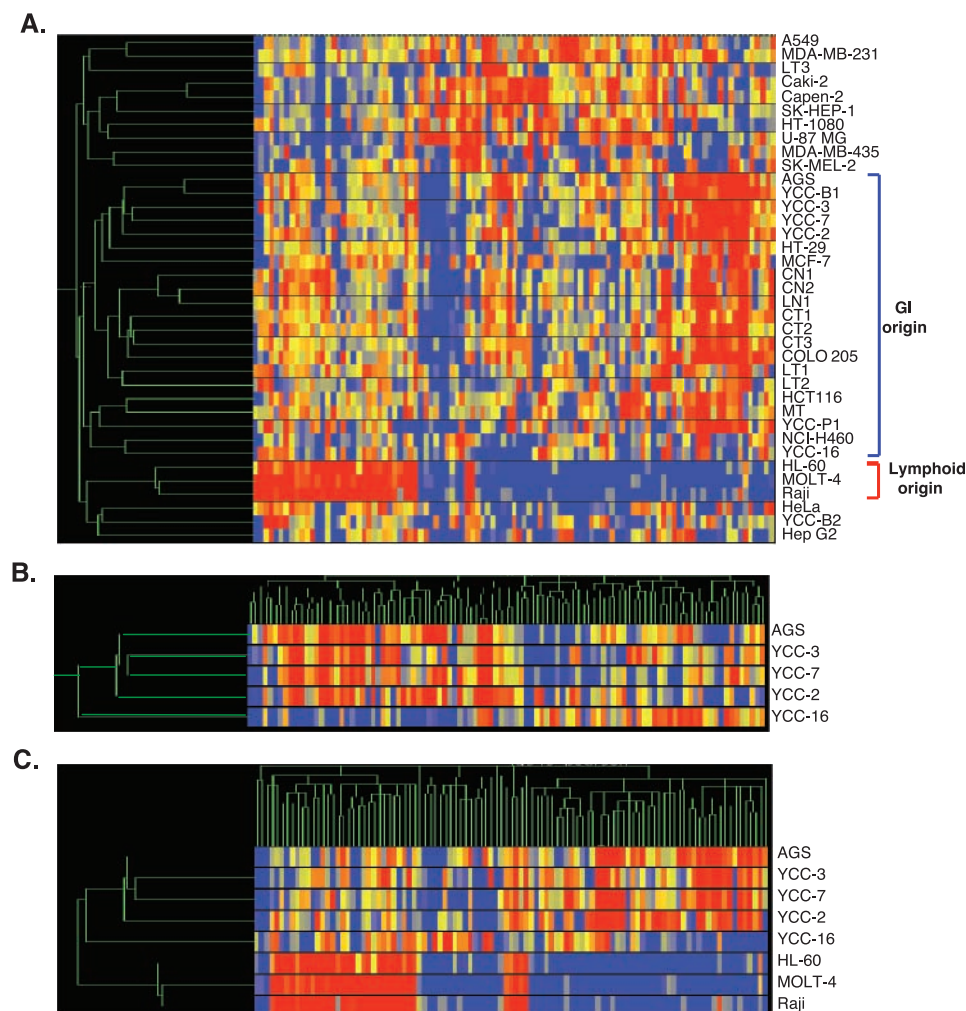
**Analysis of Global Gene Expression Patterns of Cancer Cell Lines.** We did cDNA microarray testing with 7,500 cDNA spots, representing ~6,512 genes to investigate the comprehensive gene expression patterns within 27 cancer cell lines and 9 tissues. First, we used a hierarchical clustering algorithm to group genes as well as cell lines and tissue samples according to similarities in their expression patterns by simply filtered 7,128 spots. Hierarchical clustering showed that the samples were grouped into two major branches, tissue samples and cell lines (data not shown). We observed the distinct clusters including cell lines with common tissue origin clustered together: lymphoid organs, the gastrointestinal tract, and remaining cell lines with only a few exceptions. In addition, the interesting features were that the differences of gene expression patterns among the 27 cell lines were much greater than those between normal and tumor tissues. However, normal tissues were separated from the tumor tissues.

**Selection of a Subset of Genes That Discriminate Cancer Cell Lines of a Gastrointestinal Origin from Other Cancer Cell Lines of a Different Origin.** Following the scheme of the gene selection in Fig. 1, we selected the significant genes with filtered 3,658 genes. Firstly, we used multiclass SAM to choose a subset of 199 cell origin classifier genes, which were assumed to differentiate cancer cell lines according to their origins, such as gastrointestinal tract, lymphoid organs, and remaining cell lines (false discovery rate 1%,  $\delta = 0.5751$ ; Fig. 1A). Next, by using two-class SAM, we selected a subset of 1,158 genes that expressed only in cell lines compared with tissues (false discovery rate 10%,  $\delta = 0.9765$ ; Fig. 1B). Then, to identify the subset of genes specific for gastrointestinal tract cell lines, we filtered off a subset of genes related to cell line establishment from 199 origin classifier genes using our Difference Program. Therefore, we could select 115 genes that might discriminate gastrointestinal cancer cell lines, including colon and stomach cancer cell lines, from others (Fig. 1C).

**Gene Expression Patterns Related to Tissue Origins in Various Cancer Cell Lines.** Hierarchical clustering analysis with the selected 115 genes related to gastrointestinal tract tissue origins showed that the cell lines were generally classified based

on tissue origin with a few exceptions, such as MCF7, YCC-B1, and NCI-H460, which clustered in the gastrointestinal branch (Fig. 2A). We observed that the gastric cancer cell lines YCC-2, YCC-3, and YCC-7 established at Yonsei CMRC were grouped with other gastrointestinal tract cancer cells, such as AGS, HT-29, COLO 205, and HCT116 obtained from American Type Culture Collection (Fig. 2A). In addition, when the selected subset of genes was used for the clustering analysis, we observed that tissues of gastrointestinal origin completely integrated into gastrointestinal cell lines (Fig. 2A). However, in case of colon cancer cell lines, including COLO 205, HT-29, and HCT116, each cell line showed a unique expression pattern, suggesting different biological phenotype in each cell line.

**Gene Expression Patterns According to Metastatic Pattern in Gastric Cancer.** By comparing gene expression profiles of five gastric cancer cell lines with different metastatic patterns (18), we discovered a subset of differentially regulated 66 genes in YCC-16, a cell line isolated from the blood of a gastric cancer patient (Tables 1 and 2), suggesting the subset of genes with more hematogenous metastatic potentials. Gastric cancer cell lines were divided into independent branches according to metastatic character: AGS (primary), YCC-2,



*Fig. 2* Two-way hierarchical clustering with selected 115 genes related to gastrointestinal tract cancer cell lines. *A*, clustering with all samples showing that the 115 genes distinguish gastrointestinal samples from other cell lines of different origins; *B*, clustering among the gastric cancer cell lines; *C*, clustering among the gastric and lymphoid cancer cell lines. *CT*, colon tumor tissue; *CN*, colon normal tissue; *LT*, liver tumor tissue; *LN*, liver normal tissue; *GI*, gastrointestinal. *Red*, high relative levels of expression; *blue*, low relative levels of expression.

YCC-3, and YCC-7 in one branch (peritoneal) and YCC-16 in another (hematogenous), as shown in TreeView (Fig. 2B). Among 66 genes, 40 were down-regulated (Table 1) and 26 were up-regulated (Table 2) in blood-borne YCC-16 cell line compared with the other four gastric cancer cell lines established from the peritoneal ascites of advanced gastric cancer patients. In addition, based on Pearson correlation, we observed that cancer cell lines of a lymphoid origin were closely clustered with YCC-16, which was originated in blood-borne cancer cells of the gastric cancer patient (Fig. 2C). Among 66 genes, 26 showed the similar expression with hematogenously metastasized YCC-16 gastric cancer cell and three lymphoid origin cells with 22 down-regulated and 4 up-regulated genes (Tables 1 and 2).

**Comparison of Gene Expression between Real-time RT-PCR and Microarray Analysis.** Among the 66 selected genes, we randomly selected 6 genes for validation using real-time RT-PCR, including *PXN*, *ITGB5*, *DAG1*, *KRT8*, *ST14*, and *NRG1*. Sixteen cell lines of breast, gastric, colorectal, and liver cancer and lymphoid origin cell lines were chosen. First, we evaluated the correlation between real-time RT-PCR and microarray. Pearson correlation coefficient of each gene was *PXN* 0.86, *KRT8* 0.94, *ITGB5* 0.65, *DAG1* 0.45, *ST14* 0.5, and *NRG1* -0.13. The comparison of *PXN* and *KRT8* RNA expression levels between two methods is displayed in Fig. 3A and B with good correlation (*PXN*  $r^2 = 0.73$ , *KRT8*  $r^2 = 0.84$ ). Then, we clustered five gastric, three colorectal cancer, and three lymphoid origin cells according to real-time RT-PCR result with two genes of higher correlation between two methods, *PXN* and *KRT8* (Fig. 3C). It showed significant discrimination of three lymphoid origin cell lines from gastrointestinal tract cell lines. Especially, YCC-16, which resembles the lymphoid cell line characters in microarray data (Fig. 2B), was closely related to lymphoid cells (Fig. 3C). Next, to evaluate the performance of these gene expression results from real-time RT-PCR in

Table 2 Among 66 genes differently expressed among five gastric cell lines based on their metastatic potential, 26 up-regulated genes in blood-borne YCC-16 cell line

| Name   | Genbank accession no. |
|--|-----------------------|
| <i>peptidase β*</i>                            | N29844                |
| <i>lysylhydrogenase 2</i>                      | AA136707              |
| <i>microtubule-associated protein τ</i>        | AA199717              |
| <i>sequestosome 1</i>                          | AW074995              |
| <i>glutamic-oxaloacetic transaminase A</i>     | H22856                |
| <i>trinucleotide repeat containing 3</i>       | N57754                |
| <i>NRG1</i>                                    | R72075                |
| <i>protein kinase MUK*</i>                     | AA053674              |
| <i>casein kinase A2</i>                        | AA054996              |
| <i>polymerase RNA III (DNA directed)</i>       | AA282063              |
| <i>cyclin-dependent kinase inhibitor 3</i>     | AA284072              |
| <i>CD151 antigen</i>                           | AA443118              |
| <i>paraoxonase 2</i>                           | AA446028              |
| <i>zinc finger protein 131</i>                 | AA448919              |
| <i>lactate dehydrogenase C</i>                 | AA453467              |
| <i>PTD017 protein*</i>                         | AA464612              |
| <i>cartilage-associated protein</i>            | AA486278              |
| <i>thyroid receptor interacting protein 15</i> | AA625651              |
| <i>BCL2-associated athanogene*</i>             | AI017240              |
| <i>cyclin-dependent kinase inhibitor 1A</i>    | AI952615              |
| <i>acetyl-CoA acyltransferase 2</i>            | H07926                |
| <i>leptin receptor gene-related protein</i>    | H51066                |
| <i>replication factor C (activator 1) 3</i>    | N39611                |
| <i>integrin β-like 1</i>                       | N52533                |
| <i>TNFR1</i>                                   | W02761                |
| <i>integrin β<sub>1</sub></i>                  | W67174                |

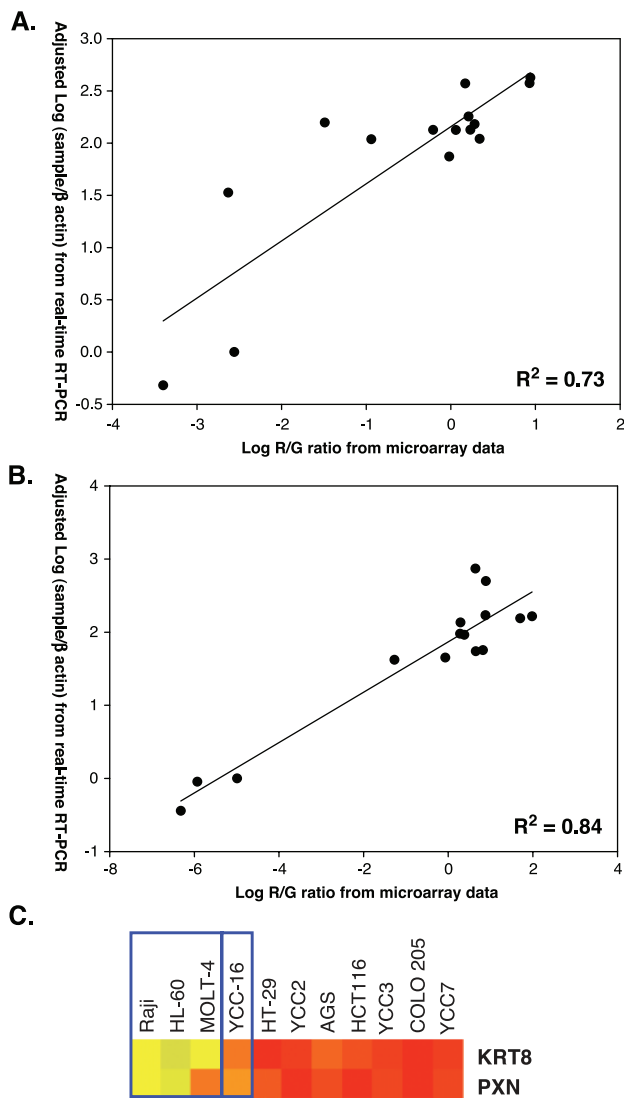
\*Indicates four genes in which YCC-16 and lymphoid cancer cell lines show the similar gene expression patterns.

discrimination of gastrointestinal cancer types, we grouped the cell lines with the same origin of tumors of colorectal, gastric, and hematologic malignant cells. When we compared the mean expression level of six genes based on the detection method, we observed that the expression patterns of gastrointestinal tract cell

Table 1 Among 66 genes differently expressed among five gastric cell lines based on their metastatic potential, 40 down-regulated genes in blood-borne YCC-16 cell line

| Name  | Genbank accession no. | Name  | Genbank accession no. |
|---|-----------------------|---|-----------------------|
| <i>myosin IC*</i>                             | AA029956              | <i>myosin VI</i>                                | AA625890              |
| <i>zinc finger protein 38</i>                 | AA088434              | <i>LCA homologue*</i>                           | AA598513              |
| <i>vaccinia-related kinase-1</i>              | AA112979              | <i>pleiomorphic adenoma gene-like 2</i>         | AA704187              |
| <i>major vault protein*</i>                   | AA158991              | <i>serine/threonine kinase 18</i>               | AA732873              |
| <i>KDEL receptor*</i>                         | AA181085              | <i>KIAA1594 protein</i>                         | AA876039              |
| <i>butyrate response factor 1*</i>            | AA424743              | <i>myocilin*</i>                                | AI971049              |
| <i>destrin (actin-depolymerizing factor)*</i> | AA424824              | <i>keratin 14</i>                               | H44051                |
| <i>PXN*</i>                                   | AA430574              | <i>hypothetical protein MGA 20576*</i>          | H51645                |
| <i>ankyrin repeat-containing protein</i>      | AA434117              | <i>crystallin ζ (quinone reductase)*</i>        | R40946                |
| <i>ITGB5</i>                                  | AA434397              | <i>neuronal protein 4.1*</i>                    | R71689                |
| <i>syntaxin 3A*</i>                           | AA436871              | <i>topoisomerase DNA IIβ (180 kDa)</i>          | T59934                |
| <i>E74-like factor 2</i>                      | AA447783              | <i>splicing factor, arginine/serine-rich, 4</i> | W87714                |
| <i>ribosome binding protein 1</i>             | AA447804              | <i>tetraspan 3*</i>                             | AA284492              |
| <i>coronin, actin-binding protein 1C*</i>     | AA456063              | <i>malate dehydrogenase 1, NAD*</i>             | AA403295              |
| <i>HTOM34P*</i>                               | AA457118              | <i>hydroxysteroid dehydrogenase 4</i>           | AA487914              |
| <i>cyclin D1 (PRAD1)*</i>                     | AA487486              | <i>copine III*</i>                              | AA505111              |
| <i>ST14</i>                                   | AA489246              | <i>ribosomal protein S9</i>                     | AW074994              |
| <i>DAG1*</i>                                  | AA496691              | <i>EST*</i>                                     | H57136                |
| <i>heat shock protein 75*</i>                 | AA497020              | <i>arfaptin 1</i>                               | T52363                |
| <i>KRT8*</i>                                  | AA598517              | <i>keratin 13</i>                               | W60057                |

\*Indicates 22 genes in which YCC-16 and lymphoid cancer cell lines show the similar gene expression patterns.



**Fig. 3** Comparison of microarray data with quantitative real-time RT-PCR. *A*, a plot comparing the expression of *PXN* transcript in 16 cell lines using microarray and real-time RT-PCR. *B*, a plot of *KRT8*. *X* axis, log R/G ratio from microarray; *Y* axis, adjusted log (test sample copies/ $\beta$ -actin copies). *C*, clustering of five gastric, three colorectal cancer, and three lymphoid cell lines with *PXN* and *KRT8*.

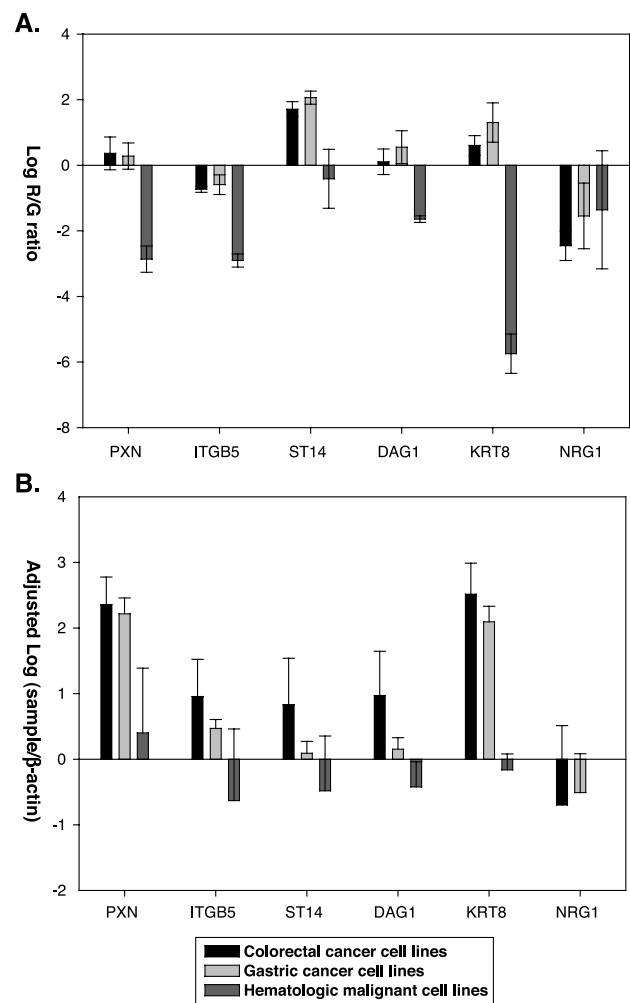
lines were different from hematologic malignant cells in both microarray (Fig. 4*A*) and real-time RT-PCR (Fig. 4*B*).

## DISCUSSION

cDNA microarray technologies allow us to comprehensive investigate cancer in various aspects at the RNA level (4–7). To understand cancer biology, the gene expression patterns measured in cell lines using microarray can help to distinguish them at a molecular level in histologically complex cancer specimens (19). Although the laser microdissection has made it possible to analyze the expression patterns of only cancer cells from clinical tumor tissues (20), tumor heterogeneity and significant role of stromal tissues are not negligible to draw the conclusion. Hence, to identify the “molecular portrait” of the

tissues from which the cell lines were derived, we evaluated the genome-wide expression profilings with 9 tissues and 27 cell lines. First, we confirmed that the tissues and the cell lines were clustered separately regardless of tissue origin (data not shown). Many reports suggested that a given cell line differs from its original tissue both biologically and genetically (5, 10, 21). Possible reasons are (a) the cell line is established from a selected clone among various cells in the tissue; (b) tissues are composed of a variety of cell origins, such as fibroblasts, blood vessel cells, immune system cells, and fat cells; and (c) *in vitro* system cells continuously change to adjust to the artificial environment.

We specifically focused on genes significantly expressed in cell lines derived from gastrointestinal tract because gastric and colon cancers are the leading cause of cancer death in Asia. Many suggestions have been made concerning the selection of specific gene sets from a variety of data sets. In our study, to determine the subset of genes related to the gastrointestinal tract cell lines, we first used direct two-class SAM comparing



**Fig. 4** Different expression of six selected genes in various tumor types based on different detection method. Mean expression level of six genes in three gastric cancer cells, three colorectal cancer cells, and three lymphoid cells were. *A*, microarray results; *B*, real-time RT-PCR results.

gastrointestinal cells with other cell lines (with or without lymphoid cells). However, the selected genes (93 genes with lymphoid cells and 45 genes without lymphoid cells, respectively) failed to distinguish the gastrointestinal tract cancer cells from others (data not shown). Therefore, we introduced our current indirect data analysis scheme (Fig. 1) and found that 115 genes successfully categorized the cells based on their origins, especially in the case of gastrointestinal tract cells (Fig. 2). Using 115 selected genes, we observed the MDA-MB-231, YCC-B1 breast cancer, and NCI-H460 lung cancer cells lay inside the gastrointestinal tract cluster. This may be because those cancer cell lines are more heterogeneous than others (5, 10), and the 115 genes were selected to differentiate primarily the gastrointestinal tract cancer cells. Even if there are still some arguments about the best false discovery rate,  $\delta$ , optimal gene numbers, and grouping for gene selection, we showed the specific gene sets selected by using SAM might be useful for understanding and categorizing the cancer cell lines based on their origins.

This gene set might be applied for the identification of the origin of cancer cells when presented with cells of an ambiguous origin, especially when the gastrointestinal tract origin is suspicious. Although it may not be conclusive for the exact diagnosis, it could be helpful as complementary information. Gene expression profiles of cancer cell lines could reflect novel aspects of the phenotype (11–13). Based on previous biological studies (18), including the cell doubling time, tumorigenicity in soft agar, or the xenograft model and motility assay, YCC-16 was the most aggressive cell line among the five gastric cancer cell lines used in the present study. This explains why YCC-16 is not clustered with the other gastric cancer cells. In addition, this cell line was established from the bloodstream in contrast to other cell lines established from primary (AGS) and ascites (YCC series), which might explain why its gene expression is tightly clustered with blood cells. In other words, if the tumor cells show the same expression pattern as YCC-16 with this subset of genes, one might presume that those tumors may survive in the bloodstream and can metastasize to distant organs. Our data may be consistent with the idea that some primary tumors may be preprogrammed to spread to other organs and that this propensity could be detectable at the time of initial diagnosis and therefore be used as a prognostic indicator (22, 23).

It is well known that the alteration in adhesion molecules and apoptosis-related genes facilitate cell escape from the primary site, induce the resistance to cell death, and finally lead to metastasis (24–30). Among selected genes, *PXN*, *ITGB5*, *coronin*, *actin-binding protein 1C*, *BAG family molecular chaperone regulator-1*, *cyclin-dependent kinase inhibitor 1A (p21)*, and fibronectin receptor (*integrin  $\beta_1$* ) are involved in cell adhesion or apoptosis. Members of coronin family are involved in apoptosis, whereas *BAG family molecular chaperone regulator-1* has antiapoptotic activity and increases the anti-cell death function of bcl-2 (<http://genome-www5.stanford.edu/cgi-bin/source/sourceSearch>). In addition, p21 was known to inhibit apoptosis (25, 26). In YCC-16, the expression of coronin was down-regulated, whereas *BAG family molecular chaperone regulator-1* and p21 were up-regulated. The decreased apoptosis could provide the opportunity for cancer cells to spread to distant organs with high survival rate. Consistent with our data, increased

expression of *integrin  $\beta_1$*  was reported to correlate with increased invasion and metastasis in some cancers (24, 29, 30).

Usually, the microarray data results had required the validation of gene expression using other methods such as RT-PCR or real-time PCR. It started when the quality of microarray was not satisfactory from clone preparation, chip production, variations from hybridization, or inadequate data analysis. However, with the improved chip quality and hybridization technique, recent microarray data showed good correlation with RT-PCR method with reliable biological meanings (17, 21, 22). We did real-time PCR analysis for evaluating not only the technical validation of microarray result but also the potential biological validation of the selected genes. In this study, we confirmed that the gene expression levels of two methods were well correlated, especially with *PAX* and *KRT8*, and relatively well with *ITGB5*, *DAG1*, and *ST14*. However, *NRG1* showed the significantly low correlation, which might be related that the several cell lines did not express *NRG1* in real-time RT-PCR. As the basic difference in the technology between microarray and real-time RT-PCR is the simultaneous competitive hybridization of microarray. Moreover, we used the reference RNA with 11 cancer cell lines, resulting in any relative expression data of each sample in microarray. Meanwhile, the real-time RT-PCR detects the expression level of sample itself. Until now, direct integration of two data sets is not settled down yet. Regardless of various correlation between microarray and real-time RT-PCR analysis, we observed the expression patterns based on RT-PCR of selected six genes successfully discriminate gastrointestinal tract from lymphoid origin cells, suggesting that the genes were properly selected.

In conclusion, evaluation of genome-wide gene expression profiles of various cancer cell lines provides the significant genetic information to understand the cancer biology, characteristics of tumor types, and specific characters of each cell line. It may be helpful for identifying and categorizing optimal cell lines for a given experimental purpose.

## REFERENCES

1. Monks A, Scudiero DA, Johnson GS, Paull KD, Sausville EA. The NCI-cancer drug screen: a smart screen to identify effectors of novel target. *Anticancer Drug Des* 1997;12:533–41.
2. Weinstein JN, Kohn KW, Grever MR, et al. Neural computing in cancer drug development: predicting mechanism of action. *Science* 1992;258:447–51.
3. Stinson SF, Alley MC, Fiebig HH, et al. Morphological and immunocytochemical characteristics of human tumor cell lines for use in a disease-oriented anticancer drug screen. *Anticancer Res* 1992; 12:1035–53.
4. Paull KD, Shoemaker RH, Hodes L, et al. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J Natl Cancer Inst* 1989;81:1088–92.
5. Lee JC, Lashkari D, Shalon D, et al. Systemic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24: 227–35.
6. Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24: 236–44.
7. Staunton JE, Slonim DK, Coller HA, et al. Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci U S A* 2001;98:10787–92.



8. Ji J, Chen X, Leung SY, et al. Comprehensive analysis of the gene expression profiles in human cancer cell lines. *Oncogene* 2002;21:6549–56.
9. Sakakura C, Hagiwara A, Nakanishi M, et al. Differential gene expression profiles of gastric cancer cells established from primary tumor and malignant ascites. *Br J Cancer* 2002;87:1153–61.
10. Virtanen C, Ishikawa Y, Honjoh D, et al. Integrated classification of lung tumors and cell lines by expression profiling. *Proc Natl Acad Sci U S A* 2002;99:12357–62.
11. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997;278:680–6.
12. Iyer VR, Eisen MB, Ross DT, et al. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999;283:83–7.
13. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet* 1999;21 Suppl:33–7.
14. Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002;32:496–501.
15. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple side systemic variation. *Nucleic Acid Res* 2002;30:e15.
16. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
17. Xu Y, Selaru FM, Yin J, et al. Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. *Cancer Res* 2002;62:3493–7.
18. Rha SY, Noh SH, Kwak HJ, et al. Comparison of biological phenotypes according to midkine expression in gastric cancer cells and their autocrine activities could be modulated by pentosan polysulfate. *Cancer Lett* 1997;118:37–46.
19. Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JR, Elkahouh AG. *In vivo* gene expression profile analysis of human breast cancer progression. *Cancer Res* 1999;59:5656–61.
20. Perou CM, Jeffery SS, Rijn MV, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 1999;96:9212–7.
21. Welsh JB, Zarrinkar PP, Sapinoso LM, et al. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc Natl Acad Sci U S A* 2001;98:1176–81.
22. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003;31:1–6.
23. Van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009.
24. Arboleda MJ, Lyons JF, Kabbinavar FF, et al. Overexpression of AKT2/protein kinase B $\beta$  leads to up-regulation of  $\beta_1$  integrins, increased invasion, and metastasis of human breast and ovarian cancer cells. *Cancer Res* 2003;63:196–206.
25. Fang M, Liu B, Schmidt M, Lu Y, Mendelsohn J, Fan Z. Involvement of p21<sup>Waf1</sup> in mediating inhibition of paclitaxel-induced apoptosis by epidermal growth factor in MDA-MB-486 human breast cancer cells. *Anticancer Res* 2000;20:101–3.
26. Xu SQ, EL-Diery WS. p21(WAF1/CIP1) inhibits caspase cleavage by TRAIL death receptor DR4. *Biochem Biophys Res Commun* 2000;269:179–90.
27. Yawata A, Adachi M, Okuda H, et al. Prolonged cell survival enhances peritoneal dissemination of gastric cancer cells. *Oncogene* 1998;16:2681–6.
28. Huerta S, Harris DM, Jazirehi A, et al. Gene expression profile of metastatic colon cancer cells resistant to cisplatin-induced apoptosis. *Int J Oncol* 2003;663–70.
29. Fujita S, Watanabe M, Kubota T, Teramoto T, Kitajima M. Alteration of expression in integrin 1-subunit correlates with invasion and metastasis on colorectal cancer. *Cancer Lett* 1995;91:145–9.
30. Morini M, Mottolese M, Ferrari N, et al. The  $\alpha_3\beta_1$  integrin is associated with mammary carcinoma cell metastasis, invasion, and gelatinase B (MMP-9) activity. *Int J Cancer* 2000;87:336–42.