INTERNATIONAL JOURNAL OF ONCOLOGY 32: 413-423, 2008

Selection of a novel drug-response predictor in esophageal cancer: A novel screening method using microarray and identification of *IFITM1* as a potent marker gene of CDDP response

SHOICHI FUMOTO[1], TATSUSHI SHIMOKUNI[1], KEIJI TANIMOTO[1], KEIKO HIYAMA[1], KEIKO OTANI[2], MEGU OHTAKI[2], JUN HIHARA[3], KAZUHIRO YOSHIDA[4], EISO HIYAMA[5], TSUYOSHI NOGUCHI[6] and MASAHIKO NISHIYAMA[1,7]

Departments of [1]Translational Cancer Research, [2]Environmetrics and Biometrics, [3]Surgical Oncology, Research Institute for Radiation Biology and Medicine; [4]Department of Oncologic Surgery, Gifu University Graduate School of Medicine; [5]Natural Science Center for Basic Research and Development, Hiroshima University, Hiroshima; [6]Department of Gastrointestinal Surgery, Faculty of Medicine, Oita University, Oita; [7]Saitama Medical University International Medical Center, Saitama, Japan

Received September 19, 2007; Accepted November 7, 2007

Abstract. Prior laboratory prediction of individual drug response is of key importance in esophageal squamous cell carcinoma (ESCC), because of the extremely narrow therapeutic index of chemotherapy. However, very few critical markers have been validated to date for ESCC. We previously demonstrated that simultaneous performance of two different types of comprehensive gene expression analysis might provide a way to identify potent marker genes for drug sensitivity from the expression-sensitivity correlation analysis alone, but the screening method appeared not to be always effective. Therefore, we attempted to identify novel potent marker genes using a new statistical analysis of oligonucleotide microarray expression data, based on a two-dimensional mixed normal model, and selected 3 and 7 novel candidates for 5-fluorouracil (5-FU) and cis-platinum (CDDP), respectively. Interferon induced transmembrane protein 1 (*IFITM1*) gene alone, being suggested as a key gene of Wnt pathway, was commonly selected in both screening methods. The transfection analyses and siRNA-mediated knock-down experiments revealed that expression of *IFITM1* closely related to cellular sensitivity to CDDP. Considering the fact that drug sensitivity is determined by multiple genes, we established the best linear model using quantified expression data of a set of all the selected marker genes including *IFITM1*, which converted the quantified expression data of ESCC cell lines into an IC_{50} value of each drug. In the same way, using the representative genes selected *in vitro*, we developed highly predictive formulae for disease-free survival (DFS) of the CDDP/5-FU combination after curative operation in esophageal cancer patients (R=0.917). A two-dimensional mixed normal model can be a powerful tool to identify novel drug-response determinants, and the *IFITM1* gene selected by the statistical method a novel critical biomarker of CDDP response in ESCC.

Correspondence to: Dr Keiko Hiyama, Department of Translational Cancer Research, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551, Japan
E-mail: khiyama@hiroshima-u.ac.jp

Key words: personalized medicine, drug sensitivity, gene expression, microarray, *IFITM1*, esophageal cancer

Introduction

Pharmacogenomic biomarkers hold great promise for the prediction of clinical outcomes of cancer chemotherapy, which would allow the selection of an optimal regimen for each individual (1-3). Extensive efforts to promote such personalized medicine have led to better predictive markers, but enormous tasks remain to be done (4,5). Emerging evidence has revealed that none of the suggested factors alone is consistently critical in drug response, and prediction of a responder for chemotherapy by 'the snapshot expression profile' of microarray is increasingly recognized to be more challenging than previously expected (6). Identification of a better prediction marker is urgently needed.

Among a variety of cancers, esophageal cancer is likely one of the most important targets of individualized chemotherapy. For esophageal cancer, chemotherapy is considered to be a most potent treatment option to improve the poor prognosis. However, the therapeutic index of chemotherapy is extremely narrow, and the optimal therapy remains unclear (7,8). Numerous patients undergo a regimen without benefit. These facts encouraged us to focus on the biomarker of individual response to chemotherapy for the disease.

The most difficult obstacle for the prediction of therapeutic efficacy is an intricate mechanism of drug sensitivity: multiple factors are involved in drug response mechanisms,

key determinants of the response significantly vary among individuals, and they intricately interact. The multifactorial mechanisms limit the prediction of individual drug response by any single marker including a 'snapshot expression profile' of microarray (6,9,10). Therefore, we have attempted to select a set of key marker genes using DNA microarray *in vitro* and developed a prediction system for clinical chemotherapeutic response through multiple regression analysis using expression data of the selected genes in several cancers, such as gastric, ovarian, and esophageal cancers (11-13). The observed predictive values of fixed formulae suggested that our attempts are likely a practical and potent approach to better prediction. The genes selected by the expression-sensitivity correlation analysis were more correlative with drug efficacy than those previously proven as drug-sensitivity determinants, and multiple regression analysis might work well to embrace the variable expressions of the selected genes and arrange them in order to predict the efficacy of the drugs. Nevertheless, in certain selected genes, the functional significance of drug sensitivity determinant *in vitro* was not fully proven indicating that there exist more significant prediction marker genes. DNA chip technology enables us to overview a huge number of gene expressions simultaneously and can provide a variety of candidates for novel prediction markers, but still there is no definitive way to determine the critical ones from such a huge number of candidates.

In this study, focusing on esophageal cancer and 2 key chemotherapeutic agents for the advanced disease, cis-platinum (CDDP) and 5-fluorouracil (5-FU), we attempted to select more powerful sensitivity markers using a new statistical method, a two-dimensional mixed normal model proposed by Ohtaki *et al* (14), and demonstrated for the first time that interferon induced transmembrane protein 1 (*IFITM1*) gene was possibly a key determinant of the CDDP sensitivity. We also found that a set of the selected genes including *IFITM1* allowed us to predict therapeutic responses to CDDP chemotherapy both *in vitro* and clinically in esophageal cancer. These findings may contribute to promoting study of individualized chemotherapy.

Materials and methods

We applied microarray analysis and cytotoxic assay data of ESCC cell lines obtained in the previous study (12) to the new statistical analysis based on a two-dimensional mixed normal model (14) to explore the gene critically responsible for the 5-FU/CDDP efficacy with the following biological evidence.

Chemicals. 5-FU was kindly provided by Kyowa Hakko Kogyo Co., Ltd. (Tokyo, Japan). CDDP was generously provided by Bristol-Myers K.K. (Tokyo, Japan). All other chemicals were of analytical grade and were purchased from Wako Pure Chemicals (Osaka, Japan) and Sigma (St. Louis, MO, USA).

Cells. The 20 KYSE human esophageal squamous cell carcinoma cell lines, (KYSE-30, -140, -150, -170, -180, -200, -220, -350, -410, -450, -510, -520, -590, -770, -850, -890, -1170, -1190, -1250, and -2270) were prepared as previously described (12). All cell lines were cultured in RPMI-1640 medium (Life Technologies, Inc., Grand Island, NY) containing 10% heat-inactivated fetal bovine serum (FBS; BioWhittaker, Verviers, Belgium) at 37˚C in a humidified atmosphere of 5% CO_2 and maintained in continuous exponential growth by passage every 3 days. For gene expression analysis, exponentially growing cultured cells were collected and stored at -80˚C until use.

Patients and human tissue sample. ESCC tissue specimens were collected at surgery from chemo-naïve patients with advanced esophageal cancer (TNM/UICC classification: Stage III or IV) as previously described (12). The patients received curative esophagectomy with the subsequent 5-FU/CDDP combination chemotherapy as the post-operative adjuvant chemotherapy, and their prognosis and follow-up until August 1st, 2007 are presented. The patients median age was 61, range 49-78 years) with performance status (World Health Organization, WHO) 0-2 without significant baseline-laboratory abnormalities, and life expectancy was estimated at >3 months. 5-FU was given by continuous intravenous administration at a dose of 250 mg/m^2 for 28 days or 5-day continuous-infusion of 500 mg/body/day per week for 28 days, as a combination regimen with cisplatin at an extremely low dose of 3 mg/m^2 or 10 mg/body/day. Total administered doses of 5-FU and CDDP ranged between 2,625 and 10,500 mg (median: 10,000 mg, mean: 8,912 mg), and between 26 and 200 mg (median: 200 mg, mean: 143 mg), respectively. CT (computed tomography) scanning was performed every one or two months to evaluate disease-free survival (DFS). Among the 18 tumor samples obtained from 17 patients, 14 tumors obtained early were used to yield the prediction formulae and 4 subsequently obtained tumors were used as test samples. Written informed consent was obtained from all patients, and the protocol was approved by our institutional ethics committee. The collected tumor specimens were stored at -80˚C until use.

Cytotoxic assay. Drug-induced cytotoxicity was evaluated by conventional MTT dye reduction assay as previously reported (12). Briefly, 4x10^3/well cells were seeded in 96-microwell plates (Nunclon, Nunc, Roskilde, Denmark) RPMI-1640 with 10% FBS. After 24 h of incubation, the medium was replaced and cells were exposed to the indicated drug concentrations for 72 h, after which 10 μl of 0.4% MTT reagent and 0.1 M sodium succinate were added to each well. After 2 h of incubation, 150 μl of DMSO was added to dissolve the purple formazan precipitate. The formazan dye was measured spectrophotometrically (570-650 nm) using MAXline™ microplate reader (Molecular Devices Corp., Sunnyvale, CA). The cytotoxic effect of each treatment was assessed by IC_{50} value (inhibitory drug concentration of 50% cell growth: drug concentration of 50% optical density of control).

Extraction and purification of RNA. Total RNA of cell pellets or frozen tissue samples was prepared using Qiagen RNeasy mini kit (Qiagen, Inc., Valencia, CA). The quality of the RNA was checked using Agilent Technologies 2100 bioanalyzer (Agilent, Palo Alto, CA) and tissue samples with poor quality were excluded.

Screening of candidate genes using data of comprehensive gene expression analyses. Gene expression data of 20 KYSE esophageal cancer cell lines analyzed by CodeLink expression bioarray system (GE Healthcare, Tokyo, Japan) (12) were applied to the new statistical model. The oligonucleotide microarray data were registered to the gene expression Omnibus under GE accession no. GSE 2447 (http://www.ncbi.nlm.nih.gov.geo/). On these data, a two-dimensional mixed normal model, in addition to the rank correlation analysis, was applied to select the most potent prediction marker genes from the large number of candidates. The rank correlation coefficient (Spearman's correlation coefficient) is well-known as a robust statistical index for quantifying degrees of correlation between ranks of two sets of measurements; it is useful even when data are contaminated with certain outliers.

The statistical significance was evaluated with *P*-value obtained from the Monte Carlo method by generating null distribution under the hypothesis that there was no correlation between any two sets of measurements. Two-dimensional mixed normal model is a statistical method proposed by Ohtaki *et al*, which can effectively adjust the microarray data to facilitate comparisons through eliminating systemic biases in the measured expression levels, referred to as normalization, and identify differentially expressed genes between two cells showing different biological behaviors based on the functional status of the genes (13-15).

The probability of the gene being differentially expressed between the query and the reference samples, i.e., the status of the gene is ('on', 'off') or ('off', 'on') between them, was obtained as a posterior probability. The terms 'on' and 'off' are used to express the functional status of a gene. If a gene actually expressed yielding its product (i.e. 'mRNA') as the true signal, the status is 'on'; otherwise (i.e., mRNA is not in the sample), it is 'off'. When the status of a gene is 'off', the observed measurement reflects only the amount of systematic error and measurement error.

Real-time RT-PCR (reverse transcription-polymerase chain reaction). Total RNA (2 μg) extracted from each cell line, gene-transfected cell clone, or cancer tissue was reverse-transcribed using High-Capacity cDNA Archive™ kit (Applied Biosystems), and then 1/200 aliquot of the cDNA (equivalent to 10 ng total RNA) was subjected to real-time RT-PCR using ABI Prism™ 7900HT sequence detection system (Applied Biosystems) to estimate the expression levels of the candidate genes. Primer and probe set was provided by TaqMan™ gene expression assays (Applied Biosystems) except for the internal control (Pre-Developed TaqMan assay reagents; Applied Biosystems) and *KLRC2* (Roche Universal ProbeLibrary™; Roche Diagnostics, Basel, Switzerland). Each reaction was carried out in triplicate and averaged. The relative gene expression level was calculated as a ratio to *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) gene expression level, and the genes whose expression levels statistically correlated again with sensitivity to 5-FU or CDDP were further selected as prediction marker candidates.

Construction of plasmid. The cDNAs derived from an EB virus transformed B cell line C123 (for *IFITM1*) established from a healthy donor and a fibroblast strain MJ90 (for *B4GALT5, UGCG* and *XBP1*) were used to amplify each gene. The sequence of each primer used for amplification was: *B4GALT5*: forward, 5'-GGAATTCTATGCGCGCCCGCCGGGGGCT-3', reverse, 5'-GAAGATCTCCTCTCAGTACTCGTTCACCTG-3'; *UGCG*: forward, 5'-TGAATTCTATGGCGCTGCTGGACCTGGC-3', reverse, 5'-GAAGATCTGCTGTAGTTATACATCTAGGATTTCCTC-3'; *XBP1*: forward, 5'-TGAATTCTATGGTGGTGGTGGCAGCCGC-3', reverse, 5'-GAAGATCTCGAATTAGTTCATTAATGGCTTCC-3'; *IFITM1*: forward, 5'-GGAATTCGATGCACAAGGAGGAACATGA-3', reverse, 5'-GAAGATCTATGGGCGGCTACTAGTAAC-3'.

These sequences of the forward and reverse primers contained restriction sites of *Eco*RI and *Bgl*II, respectively. PCR reaction mixture (50 μl) containing 0.2 μg of cDNA, 5 μl of 2 mM each dNTP, 0.3 μM of each primer, and 1 U KOD-Plus-DNA polymerase (Toyobo, Osaka, Japan) was subjected to initial incubation at 94˚C for 2 min followed by 38 amplification cycles, each cycle consisting of denaturation at 94˚C for 40 sec, annealing at 58˚C for XBP1 or 60˚C for *B4GALT5, UGCG, IFITM1* for 60 sec, and extension at 68˚C for 1 min for *UGCG, XBP1, IFITM1* or 1.5 min for *B4GALT5*. After digestion with *Eco*RI and *Bgl*II, the PCR products and the expression vector p3xFLAG-CMV10 (Sigma) were ligated using the Quick Ligation™ kit (New England BioLabs) according to the manufacturer's manual and transformed into *E. coli* XL1-Blue (for *IFITM1*) or DH5α (for *B4GALT5, UGCG* and *XBP1*). All constructs were confirmed by DNA sequencing.

Transfection and selection of stable cell pools. The plasmid expressing each gene was linearized by a single cut with a restriction enzyme *Sca*I (for *UGCG, XBP1,* and *IFITM1*) or *Apa*LI (for *B4GALT5*), and then transfected into 2 human esophageal squamous cell carcinoma cell lines, KYSE-170 and -2270 using *Trans*IT®-LT1 reagent (Mirus Bio Corporation, USA) according to the manufacturer's manual. Transfected cells were cultured in RPMI-1640 medium with 10% FBS containing 150 and 500 μg/ml of G418 for KYSE-170 and -2270 cells, respectively, from 24 h after transfection for approximately one month. The established cell lines were maintained under G418-free conditions for at least one week before use to avoid any effects of G418. mRNA expression level of each gene in transfected cells were measured by real-time RT-PCR.

Knock-down analysis of IFITM1 using siRNA. *IFITM1*-specific siRNA (Silencer® pre-designed siRNA; sense: GAUAAUACAGGAAAAACGGtt, antisense: CCGUUUUUCCUGUAUUAUCtg) and negative control siRNA (Silencer negative control siRNA) were purchased from Ambion (USA) and transfected into *IFITM1* overexpression cells using the siPORT™ *NeoFX*™ (Ambion) according to the manufacturer's manual. A mixture of $4x10^3$ cells, 0.5 μl of *NeoFX*, 0.5 μl of 2 μM siRNA, and serially diluted chemicals were seeded in 96-microplates and incubated at 37˚C in a humidified 5% CO_2 incubator for 72 h, and then cytotoxicity was measured by MTT assay. Efficacy of siRNA-mediated knock-down of *IFITM1* mRNA was evaluated in the cells exposed to siRNA without chemicals for 24 h by real-time RT-PCR.

Table I. Selected prediction marker genes (expression-drug sensitivity correlation).

A, Marker genes previously proposed by two-array screening

Drug	Gene	*r (Rank correlation)* cDNA microarray	*r (Rank correlation)* Oligo microarray	*r (Peason correlation)* Real-time RT-PCR	sumPH
5-FU	*XBP1*	0.776[a]	0.569[a]	0.804[a]	0.01
	B4GALT5	0.632[a]	0.662[a]	0.772[a]	0
	UGCG	0.579[a]	0.578[a]	0.656[a]	0
CDDP	*IFITM1*	-0.630[a]	-0.734[a]	-0.567[a]	1.00
	SIPA1L2	-0.737[a]	-0.595[a]	-0.499[b]	0.90
	SAPS2	-0.567[a]	-0.570[a]	-0.462[b]	0
	ARFRP1	0.615[a]	0.565[a]	0.440[c]	0

B, Novel genes presently selected by the two-dimensional mixed normal model

Drug	Gene	*r (Rank correlation)* cDNA microarray	*r (Rank correlation)* Oligo microarray	*r (Peason correlation)* Real-time RT-PCR	sumPH
5-FU	*PTPN6*	N.A.	0.638[a]	0.854[a]	1.00
	MAP3K8	-0.140	0.607[a]	0.753[a]	1.00
	RSRC2	0.109	0.568[a]	0.596[a]	1.00
CDDP	*C5orf13*	-0.257	-0.607[a]	-0.567[a]	1.05
	NSBP1	N.A.	0.609[a]	0.636[a]	1.02
	IFITM1	-0.630[a]	-0.734[a]	-0.567[a]	1.00
	LRIG1	0.012	-0.585[a]	-0.459[b]	1.00
	RPP25	-0.314	-0.581[a]	-0.491[b]	1.00
	EDN1	-0.156	-0.568[a]	-0.448[b]	1.00
	CCDC3	N.A.	-0.566[a]	-0.536[b]	1.00

r in A was reported previously (12) and sumPH was evaluated in the present study. sumPH, sum of 'probability of heterogeneity'; N.A., not analyzed; [a]$P<0.01$; [b]$0.01\leq P<0.05$; [c]$0.05\leq P<0.1$.

Multiple regression analysis. The relationship between y_i (response value of *i*-th individual) and x_{i1}, ..., x_{ip} (explanatory variables) is formulated in the linear model $y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + ... + \theta_p x_{ip} + \varepsilon_i$, where θ_0 is constant and ε_i denotes error term. Trimmed least squares regression (TLSR) was performed to determine a set of effective genes that would satisfy the value of IC_{50}: (θ_0,..., θ_p) were estimated from the data [y_i; (x_{i1},...,x_{ip})] when we used gene expression levels and cellular sensitivity to drugs (IC_{50} value for each drug), respectively as the explanatory and the response variables. The TLSR is a robust regression method based on an extended algorithm of least median squares regression (LMSR) by Rousseeuw, which explores models using masked samples with large residuals (16). We used the software, NLReg, developed by Ohtaki (http://apollo.rbm.hiroshima-u.ac.jp/), which implemented the robust regression analysis. Outliers were identified by referring to the value of AIC (Akaike's information criterion) for each sample or checking residuals graphically, and a set of effective genes that satisfied the value of IC_{50} *in vitro* or DFS for clinical samples was explored.

Statistical analysis. Mathematical methods to process the microarray data and predict the drug efficacy are described above. Other statistical tests were performed using StatView® version 5.0 software (SAS Institute Inc., Cary, NC, USA), and Student's *t*-test was used to determine the *P*-value.

Results

Re-evaluation of the previously selected candidate marker genes by the novel statistical model. We previously demonstrated that simultaneous performance of two different types of comprehensive gene expression analysis (conveniently named as a two-array screening) might provide a way to identify potent marker genes for drug sensitivity from the expression-sensitivity correlation analysis alone, and a set of

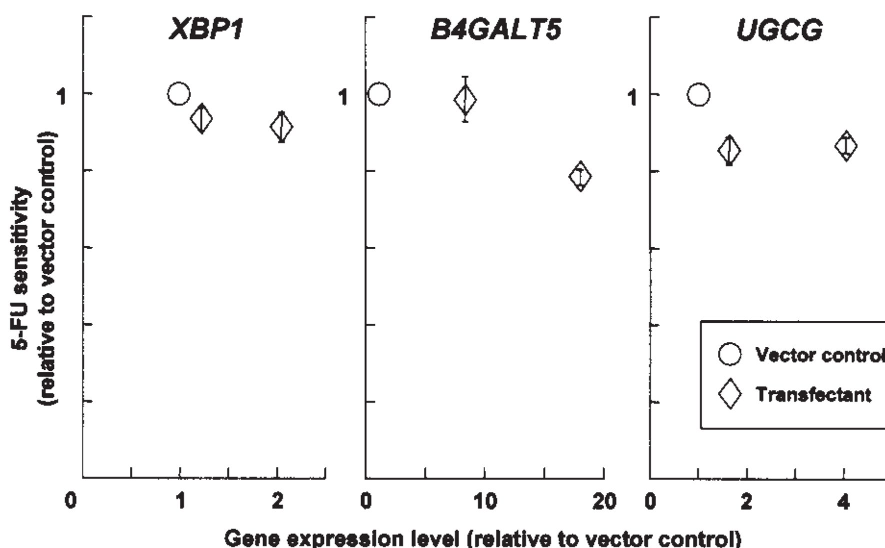

Figure 1. Effect on the 5-FU sensitivity by overexpression of the 3 previous candidate marker genes for 5-FU (*XBP1, B4GALT5,* and *UGCG*) selected by the two-array screening. No significant change in the sensitivity to 5-FU was observed in any of the *XBP1-, B4GALT5-,* or *UGCG*-transfected KYSE-170 clones, despite the significant overexpression levels compared to the vector control.

the selected genes were likely better drug-sensitivity markers in esophageal cancer (12). To re-evaluate the potential of these genes as drug sensitivity determinants, we applied the two-dimensional mixed normal model to these previously selected 7 marker genes (*XBP1, B4GALT5,* and *UGCG* for 5-FU and *IFITM1, SIPAIL2, SAPS2,* and *ARFRP1* for CDDP) and calculated the sum of 'probability of heterogeneity' ('sumPH' in Table IA: when expression level of a gene in the cell line with median IC_{50} is between those of maximum IC_{50} and minimum IC_{50}, sumPH of the gene = [probability of heterogeneity in a cell line with maximum IC_{50}] + [probability of heterogeneity in a cell line with minimum IC_{50}]; when expression level of a gene in the cell line with median IC_{50} is not between those of maximum IC_{50} and minimum IC_{50}, sumPH of the gene = |[probability of heterogeneity in a cell line with maximum IC_{50}] - [probability of heterogeneity in a cell line with minimum IC_{50}]|) as drug sensitivity determinants. The calculated sumPH of the 5 genes selected by two-array screening except *IFITM1* and *SIPAIL2* were too low to estimate them to be a potent predictor in the new screening. Among them, 3 possible marker genes for 5-FU (*XBP1, B4GALT5,* and *UGCG*) showed extremely low probability of heterogeneity (Table 1A), and in fact revealed not to have functional significance as 5-FU-sensitivity determinants by transfection analysis in KYSE-170 cells (Fig. 1).

Novel prediction marker genes selected by a combination of rank correlation analysis and novel statistical model. A two-dimensional mixed normal model may have certain advantages in selection of potent marker genes than the previously employed selection methods. Therefore, we attempted to select more reliable prediction marker genes by combining the mathematical model with rank correlation analysis. The normalized expression level of each gene in oligonucleotide array analysis and the IC_{50} for each drug in 20 esophageal cancer cell lines were ranked and the correlation between ranks of the two sets of measurements was evaluated. The rank correlation analysis showed that 124 and 272 genes closely correlated with cellular sensitivity to 5-FU and CDDP, respectively, in their expression levels ($P<0.01$), although any genes widely known as CDDP sensitivity determinants were not included. We then applied a two-dimensional mixed normal model to explore differentially expressed genes between the most resistant and sensitive cells to 5-FU or CDDP through the comparison to the cells with median IC_{50} for each drug, and selected 5 and 12 novel candidates for 5-FU and CDDP, respectively (sumPH ≥1.0).

Real-time RT-PCR analysis on these 17 candidate genes confirmed that 10 showed significant correlation between drug sensitivity and their expression levels ($P<0.05$ in the linear regression analysis), indicating that these 10 genes are reliable potent candidates as novel prediction markers: *PTPN6* (protein tyrosine phosphatase, non-receptor type 6), *MAP3K8* (mitogen-activated protein kinase kinase kinase 8) and *RSRC2* (arginine/serine-rich coiled-coil 2) for 5-FU, and *C5orf13* (chromosome 5 open reading frame 13), *NSBP1* (nucleosomal binding protein 1), *IFITM1* (interferon induced transmembrane protein 1), *LRIG1* (leucine-rich repeats and immunoglobulin-like domains 1), *RPP25* (ribonuclease P 25 kDa subunit), *EDN1* (endothelin 1), and *CCDC3* (coiled-coil domain containing 3) for CDDP (Table IB). The 7 genes selected by microarray analysis but excluded by real-time RT-PCR were *NDUFA4L2* and *PYCARD* for 5-FU and *GSTA4, GRHL3, FOXC2, SERPINB2,* and *KLRC2* for CDDP.

IFITM1 identified as a key marker gene. Whereas all of the selected genes might be potent predictors of 5-FU- and CDDP-induced cytotoxicity, their functions and the usefulness of the employed selection method remain to be elucidated. Both the two-array screening previously employed and the two-dimensional mixed normal model in the present study commonly selected *IFITM1* to be a potent predictor of CDDP efficacy with a high correlation coefficient and a sufficient

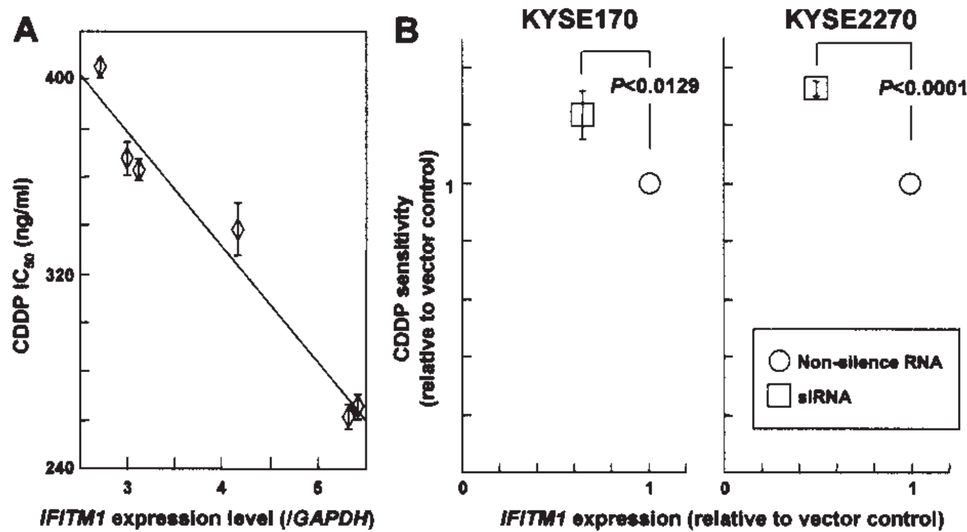

Figure 2. Effect on the CDDP sensitivity by overexpression (A) or knockdown (B) of *IFITM1*. The sensitivity to CDDP was highly correlative with the expression levels of *IFITM1* in 6 stable KYSE-170 transfectants (r=0.982, P=0.0005) (A), and was reduced by siRNA-mediated knockdown of the gene in *IFITM1* overexpressed KYSE-170 and -2270 transfectants (B).

Table II. *In vitro* prediction formulae for 5-FU and CDDP: Explanatory variables (x_{ip}) and estimated coefficients (θ_p).

A, Formulae using expression data of the previously proposed marker genes

5-FU			CDDP		
x_{ip}	θ_p	P	x_{ip}	θ_p	P
ln [*XBP1*]	87.833	0.0014	ln [*IFITM1*]	-56.174	0.0001
ln [*B4GALT5*]	76.933	0.0009	ln [*SIPA1L2*]	32.861	0.3138
ln [*UGCG*]	94.528	0.0003	ln [*SAPS2*]	-155.610	0.0001
			ln [*ARFRP1*]	124.470	<0.0001

B, Formulae using expression data of the genes presently selected by the two-dimensional mixed normal model

5-FU			CDDP		
x_{ip}	θ_p	P	x_{ip}	θ_p	P
ln [*PTPN6*]	120.160	<0.0001	ln [*C5orf13*]	-30.619	0.0031
ln [*MAP3K8*]	88.292	<0.0001	ln [*NSBP1*]	25.605	0.0001
ln [*RSRC2*]	75.948	0.0382	ln [*IFITM1*]	-78.540	<0.0001
			ln [*LRIG1*]	-33.566	0.0072
			ln [*RPP25*]	39.964	0.0144
			ln [*EDN1*]	-48.052	0.0001
			ln [*CCDC3*]	-3.597	0.4424

[ ], expression level of indicated gene.

probability of heterogeneity (Table I), which led us to focus on this gene as the most plausible key marker gene of CDDP response in esophageal cancer. The transfection of *IFITM1* into human esophageal squamous cell carcinoma cell lines KYSE-170 and -2270 revealed that expression of *IFITM1* closely related to the cellular sensitivity to CDDP: the expression levels in 6 stable KYSE-170 transfectants were inversely correlated with the IC_{50} values for CDDP (P<0.01) (Fig. 2A). This inverse correlation was also observed in KYSE-2270 (data not shown). We further confirmed that repression of *IFITM1* by siRNA significantly increased cellular resistance to CDDP in KYSE-170 and -2270 (Fig. 2B): The

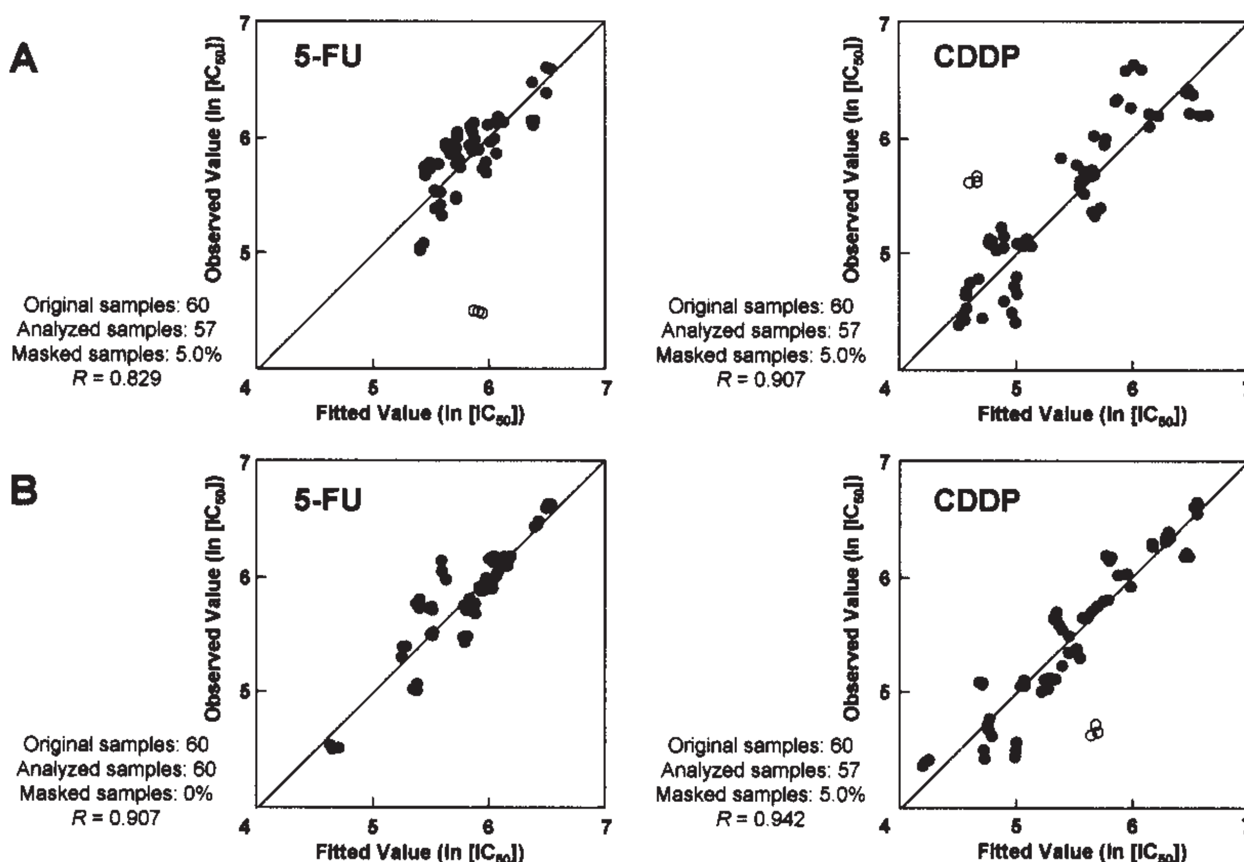

Figure 3. Prediction of *in vitro* sensitivity to 5-FU and CDDP by fixed prediction formula using expression data of a set of all the possible marker genes for each drug. Prediction formula for the IC_{50} of each drug was fixed using the variable expression data of the possible marker genes selected by previous two-array screening method (A) and the present two-dimensional mixed normal model (B). A total of 60 independent data sets, expression levels of the selected genes and IC_{50} values for 20 KYSE cell lines, were used (●, analyzed sample data; ○, a masked outlier). The vertical and horizontal axes show observed value and fitted value, respectively.

reduced expression levels (36 and 50% in KYSE-170 and -2270, respectively) closely correlated with the observed increase of IC_{50} values for CDDP (20 and 33% in KYSE-170 and -2270, respectively).

Prediction of in vitro sensitivities to 5-FU and CDDP using each set of the selected genes. *IFITM1* appeared to be one of the key determinants of CDDP sensitivity, thus a powerful predictor of CDDP sensitivity, which suggested that the two-dimensional mixed normal model likely worked well to identify novel marker genes from numerous candidates. The selection method also suggested that 5-FU and CDDP might have plural sensitivity marker genes other than *IFITM1*. Selection of a set of truly significant genes for sensitivities to drugs would allow us to predict the therapeutic response to the agents more accurately, at which point we could understand their interplay in the expression. Therefore, we performed multiple regression analysis to compose such prediction models for the *in vitro* activity of 5-FU and CDDP using expression data quantified by real-time RT-PCR of the 2 sets of selected genes (novel and previously suggested genes), and compared the potential for the prediction. The attempts provided 2 prediction formulae each for 5-FU and CDDP, to show the highest fitness, and the observed correlation coefficient (*R*, 0.907 vs. 0.829 for 5-FU; 0.942 vs. 0.907 for CDDP for novel and previously suggested genes, respectively) indicated potent predictive values of the fixed formulae (Table II and Fig. 3). The P-value of *IFITM1* was lowest among those of the other selected genes in the novel prediction formulae for CDDP sensitivity ($P<0.0001$), suggesting the significant role of *IFITM1* in the prediction.

Prediction of clinical response to 5-FU/CDDP combination therapy using each set of the representative genes. Expression analysis of a set of the key drug sensitivity genes for 5-FU and CDDP allowed the prediction of therapeutic response to the combination. The suggested potential in the prediction models of *in vitro* drug sensitivity encouraged us to construct a prediction model of clinical response, i.e., disease-free survival (DFS), to 5-FU/CDDP combination chemotherapy, in a similar manner using the same genes selected *in vitro* for both drugs. We used 14 tumor specimens from 18 collected specimens, and developed clinical prediction models. Since the number of samples was too small to make a model formula using all selected genes, we chose only 3 genes each including *IFITM1* from the 7 genes selected by previous two-array screening and from the 10 genes selected by the present two-dimensional mixed normal model, so that the best prediction models can be obtained using the limited number of genes: *B4GALT5*, *IFITM1*, and *ARFRP1* from the former

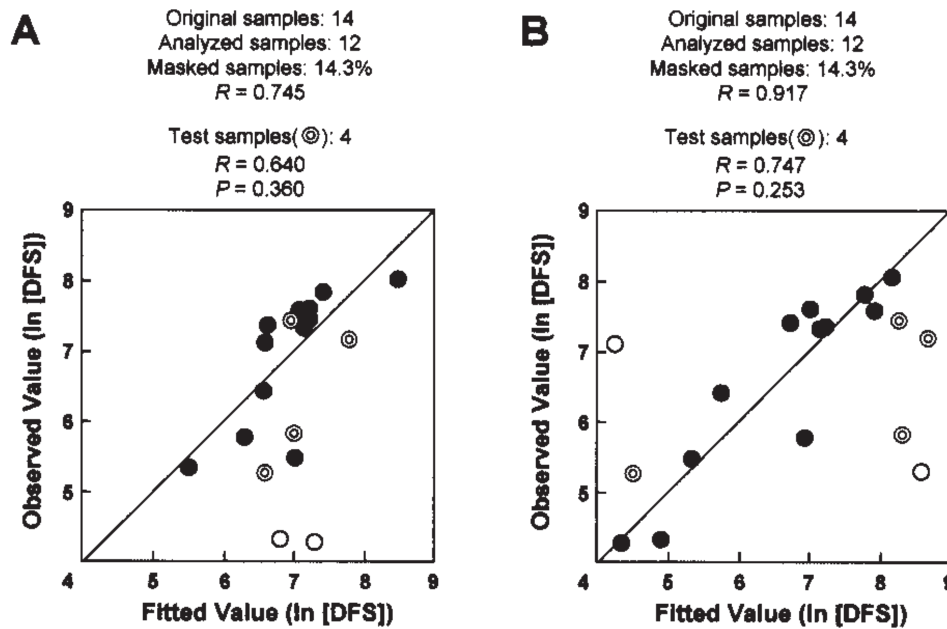

Figure 4. Prediction of disease-free survival (DFS) after surgery followed by 5-FU/CDDP combination therapy in patients with ESCC, by fixed prediction formula using expression data of a set of 3 possible marker genes for 5-FU or CDDP. Formula was fixed using the variable expression data of the possible marker genes selected by the previous two-array screening method (A) and the present two-dimensional mixed normal model (B). A total of 14 independent data sets, expression levels of the selected genes in tumor specimen and clinical response (DFS, day) data in 14 patients, were used (●, analyzed sample data; ○, a masked outlier), and another 4 sets of data were used to confirm the predictive value (circle within a circle). The vertical and horizontal axes show observed value and fitted value, respectively.

Table III. Prediction formulae for disease-free survival of 5-FU/CDDP combination: Explanatory variables (x_{ip}) and estimated coefficients (θ_p).

A, Formulae using expression data of the previously proposed marker genes

x_{ip}	θ_p	P
ln [*B4GALT5*]	-0.759	0.1363
ln [*IFITM1*]	-0.625	0.0404
ln [*ARFRP1*]	1.267	0.0278

B, Formulae using expression data of the genes presently selected by the two-dimensional mixed normal model

x_{ip}	θ_p	P
ln [*RSRC2*]	3.750	0.0003
ln [*C5orf13*]	-1.664	0.0031
ln [*IFITM1*]	0.390	0.1352

[ ], expression level of indicated gene.

and *RSRC2, C5orf13,* and *IFITM1* from the latter (Table III). Multiple regression analysis using 14 data sets of gene expression quantified by real-time RT-PCR and clinical response provided 2 prediction formulae for DFS that showed the highest fitness for each set of prediction marker genes. To confirm the predictive accuracy of the fixed formulae, we examined an additional 4 tumor samples and predicted the DFS (fitted value) by the developed formulae using their quantified expression data. Despite the limited number of samples, the DFS was more reliably predictable in the latter model using the presently selected genes (*R* of the model: 0.917, test sample: 0.747, Fig. 4). However, none of the selected genes alone could predict the DFS. We also attempted to establish other prediction formulae using several different sets of marker genes, including the sets of genes related to sensitivity to either 5-FU or CDDP alone, but the DFS was not precisely predicted by any other formulae.

Discussion

Pharmacogenomics is a large-scale systematic approach using genomic technologies such as gene sequencing, statistical genetics, and comprehensive gene expression analysis to discover drug response determinants. Nevertheless, very few critical prediction markers of drug response have been validated (1-6,9-13). New technologies have created a massive increase in the amount of genomic information, but most of the genomic information are uncharacterized and further no definitive way to exploit the full power of a global perspective, a way to identify drug response determinants from a huge number of candidates, has yet been established (11-13).

In the present study, we attempted to identify more potent marker genes using a new statistical analysis based on oligonucleotide microarray expression data, a two-dimensional

mixed normal model, and selected 3 (*PTPN6, MAP3K8* and *RSRC2*) and 7 novel candidates (*C5orf13, NSBP1, IFITM1, LRIG1, RPP25, EDN1,* and *CCDC3*) respectively for 5-FU and CDDP, as novel sensitivity marker genes in ESCC. We had previously suggested several possible marker genes selected by a two-array screening method using two different types of comprehensive gene expression analysis in esophageal cancer (12). We chose *IFITM1* as the most potent candidate for a CDDP biomarker because the gene was commonly selected in both previously and newly performed screening methods, and demonstrated its functional significance as a CDDP-sensitivity determinant through the transfection analyses and siRNA-mediated knock-down experiments.

Our additional attempt to predict *in vitro* response of 5-FU and CDDP using expression data of a set of all selected marker genes including *IFITM1* revealed that expression of *IFITM1* was of key importance also in the prediction formulae for CDDP. Furthermore, the prediction formula for clinical response (DFS) of 5-FU/CDDP combination demonstrated that a set of key drug sensitivity genes for 5-FU and CDDP allowed the prediction of therapeutic response to the combination therapy. Utility-confirmation analyses using other test samples appeared to show that the formulae using a set of representative 3 novel marker genes could predict DFS, despite the limited number of samples to obtain significant *P*-value. These results indicate that our two-dimensional mixed normal model may be effective in identifying novel drug-response determinants, and *IFITM1* selected by the new screening method can be one of the powerful biomarkers of CDDP activity in ESCC.

Our first application of a two-dimensional mixed normal model to the selection of drug response marker was for TXL/CDDP therapy in ovarian cancer patients, and the attempt suggested its significant potential (13): the differences in expression levels indicated by the mathematical model were highly confirmative in subsequent real-time RT-PCR analysis, the selected 8 novel genes were more correlative with corresponding drug sensitivity than the 5 known genes in quantified expression levels, and a combination of the 8 genes alone could work well in the prediction of clinical response to platinum/TXL combination chemotherapy.

Our data in the present study support the idea that the statistical method may identify differentially expressed genes between two cell samples with different biological behaviors based on the functional status of the genes. We found that 5 of 7 genes selected by the previous methods showed extremely low probability of heterogeneity and were not estimated to be potent predictors in the novel selection method, and the transfection analyses of the 3 possible marker genes for 5-FU sensitivity (*XBP1, B4GALT5,* and *UGCG*) revealed that none of them acted at all as the drug sensitivity determinant. Recently, serious mistakes and misunderstandings in published microarray studies to develop classifiers for tailoring individualized treatments have been pointed out (6,17).

Taken together the fact that most of our novel genes selected by the new method were not evaluated as correlative genes by previous two-array screening, cDNA- and oligonucleotide-microarray screening, the unsettled expression data of the relatively earlier developed technology might confuse the selection. Nevertheless, all of the observed data in this study leads to the proposal that a two-dimensional mixed normal model would provide certain advantages in the selection of significant genes.

The biological functions of the novel 10 selected genes are only slightly known, but several reports suggest their possible roles in drug sensitivity: *IFITM1* encodes interferon-induced transmembrane protein 1, which is now known as a key factor of Wnt pathway. The protein plays an important role in the antiproliferative activity of interferons and recent reports have suggested that the gene expression may relate to the tumor response to several anticancer therapies (18-23).

The number of reports for *IFITM1* is gradually increasing, suggesting the importance of the gene in variable drug-response and supporting our findings. *PTPN6* encodes a member of the protein tyrosine phosphatase (PTP) family, and was suggested to participate in hematopoietic differentiation. The loss of protein was also shown to enhance JAK3/STAT3 signaling and decrease proteosome degradation of JAK3 and NPM-ALK in ALK^+ anaplastic large-cell lymphoma (24,25). *MAP3K8* encodes a member of the serine/threonine protein kinase family, which can activate both the MAP kinase and JNK kinase pathways, and there are several reports that show its functional role in cancer and action on IκB kinases, thus the nuclear production of NF-κB (26-28).

For *RSRC2*, recent studies have suggested that its product possibly act as a tumor suppressor and a prognostic factor in ESCC (29). The *C5orf13* product was shown to play a role in regulation of glioma cell migration and TGFB activation (30,31); and the *NSBP1* gene, which encodes a nucleosomal binding and transcriptional activating protein, is related to the HMG-14/-17 chromosomal proteins (32,33).

The *LRIG1* product is known to maintain epidermal stem cells in a quiescent non-dividing state since its down-regulation can trigger cell proliferation. The potential as a prognostic predictor in several cancers was also documented (34-36). *RPP25* encodes protein subunit of human RNase MRP and RNase P endoribonucleases, belonging to the Alba superfamily of nucleic acid binding proteins (37).

The *EDN1* product is known as a vasoconstrictor peptide produced by vascular endothelial cells and relates to a variety of cancers (38-40), although the function of CCDC3 product is still unknown. All of these genes may play important roles in the drug-induced cytotoxicity, and possibly be a more potent predictor of CDDP-induced antitumor activity than *IFITM1*.

The detailed molecular mechanisms responsible for CDDP and 5-FU action are now under our investigation: since biological behavior and the molecular basis of cancer differ significantly according to its origin, we are first studying them focusing on ESCC, and then will expand the research area to other types of cancer. These studies may clarify the reason why CDDP marker genes in ovarian cancer selected in the first study using a two-dimensional mixed normal model significantly differed from those in ESCC selected in this second study.

In summary, we attempted to identify more potent marker genes of drug response in ESCC using a new statistical analysis of the oligonucleotide microarray expression data based on the two-dimensional mixed normal model, and

provided 3 and 7 candidates respectively for 5-FU and CDDP, as novel sensitivity marker genes. Among them, we demonstrated that *IFITM1* was of key importance in the prediction for CDDP through its transfection analysis and siRNA experiments. Since the multifactorial mechanisms limit the prediction of individual drug response by any single marker, we established the best linear model both *in vitro* and *in vivo* (prediction for clinical DFS) using quantified expression data of a set of the selected marker genes including *IFITM1*, and confirmed their potent predictive values. The work may contribute to promoting personalized medicine with novel proposal that a two-dimensional mixed normal model and/or the combined usage with other screening methods possibly improves the heretofore limited utility of microarray analysis in the selection of significant genes, in which *IFITM1* can be one of the critical prediction markers of CDDP response in ESCC.

Acknowledgements

We thank Dr Yutaka Shimada (Kyoto University) for his kind supply of 20 KYSE cell lines and helpful advice on the culture. Part of this study was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Ministry of Health, Labour and Welfare of Japan and Grant-in-Aid for University and Industry Collaboration. Part of this work was carried out at the Analysis Center of Life Science, Hiroshima University.

References

1. Evans WE and Relling MV: Pharmacogenomics: translating functional genomics into rational therapeutics. Science 286: 487-491, 1999.
2. Watters JW and McLeod HL: Cancer pharmacogenomics: current and future applications. Biochim Biophys Acta 1603: 99-111, 2003.
3. Robert J: Pharmacogenetics and pharmacogenomics as new tools to optimise cancer chemotherapy. J Chemother 16 (Suppl 4): 22-24, 2004.
4. Sadée W and Dai Z: Pharmacogenetics/genomics and personalized medicine. Hum Mol Genet 14 (Review Issue 2): R207-R214, 2005.
5. Marsh S: Impact of pharmacogenomics on clinical practice in oncology. Mol Diagn Ther 11: 79-82, 2007.
6. Sun Z and Yang P: Gene expression profiling on lung cancer outcome prediction: present clinical value and future premise. Cancer Epidemiol Biomarkers Prev 15: 2063-2068, 2006.
7. Mariette C, Piessen G and Triboulet JP: Therapeutic strategies in oesophageal carcinoma: role of surgery and other modalities. Lancet Oncol 8: 545-553, 2007.
8. Martin MG, Govindan R and Morgensztern D: Second-line therapy for esophageal cancer. Expert Rev Anticancer Ther 7: 871-876, 2007.
9. McLeod HL and Evans WE: Pharmacogenomics: unlocking the human genome for better drug therapy. Annu Rev Pharmacol Toxicol 41: 101-121, 2001.
10. Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, Park J, Scherf U, Lee JK, Reinhold WO, Weinstein JN, Mesirov JP, Lander ES and Golub TR: Chemosensitivity prediction by transcriptional profiling. Proc Natl Acad Sci USA 98: 10787-10792, 2001.
11. Tanaka T, Tanimoto K, Otani K, Satoh K, Ohtaki M, Yoshida K, Toge T, Yahata H, Tanaka S, Chayama K, Okazaki Y, Hayashizaki Y, Hiyama K and Nishiyama M: Concise prediction models of anticancer efficacy of 8 drugs using expression data from 12 selected genes. Int J Cancer 111: 617-626, 2004.
12. Shimokuni T, Tanimoto K, Hiyama K, Otani K, Ohtaki M, Hihara J, Yoshida K, Noguchi T, Kawahara K, Natsugoe S, Aikou T, Okazaki Y, Hayashizaki Y, Sato Y, Todo S, Hiyama E and Nishiyama M: Chemosensitivity prediction in esophageal squamous cell carcinoma: novel marker genes and efficacy-prediction formulae using their expression data. Int J Oncol 28: 1153-1162, 2006.
13. Komatsu M, Hiyama K, Tanimoto K, Yunokawa M, Otani K, Ohtaki M, Hiyama E, Kigawa J, Ohwada M, Suzuki M, Nagai N, Kudo Y and Nishiyama M: Prediction of individual response to platinum/paclitaxel combination using novel marker genes in ovarian cancers. Mol Cancer Ther 5: 767-775, 2006.
14. Ohtaki M, Otani K, Satoh K, Kawamura T, Hiyama K and Nishiyama M: Model-based analysis of microarray data: Exploration of differentially expressed genes between two cell types based on a two-dimensional mixed normal model. Jpn J Biometrics 26: 31-48, 2005.
15. Hiyama K, Otani K, Ohtaki M, Satoh K, Kumazaki T, Takahashi T, Mitsui Y, Okazaki Y, Hayashizaki Y, Omatsu H, Noguchi T, Tanimoto K and Nishiyama M: Differentially expressed genes throughout the cellular immortalization processes are quite different between normal human fibroblasts and endothelial cells. Int J Oncol 27: 87-95, 2005.
16. Rousseeuw PJ: Least median of squares regression. J Am Statist Assoc 79: 871-880, 1984.
17. Dupuy A and Simon RM: Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 99: 147-157, 2007.
18. Andreu P, Colnot S, Godard C, Laurent-Puig P, Lamarque D, Kahn A, Perret C and Romagnolo B: Identification of the *IFITM* family as a new molecular marker in human colorectal tumors. Cancer Res 66: 1949-1955, 2006.
19. Yang G, Xu Y, Chen X and Hu G: IFITM1 plays an essential role in the antiproliferative action of interferon-γ. Oncogene 26: 594-603, 2007.
20. Lickert H, Cox B, Wehrle C, Taketo MM, Kemler R and Rossant J: Dissecting Wnt/ß-catenin signaling during gastrulation using RNA interference in mouse embryos. Development 132: 2599-2609, 2005.
21. Tsai MH, Cook JA, Chandramouli GV, DeGraff W, Yan H, Zhao S, Coleman CN, Mitchell JB and Chuang EY: Gene expression profiling of breast, prostate, and glioma cells following single versus fractionated doses of radiation. Cancer Res 67: 3845-3852, 2007.
22. Akyerli CB, Beksac M, Holko M, Frevel M, Dalva K, Ozbek U, Soydan E, Ozcan M, Ozet G, Ilhan O, Gurman G, Akan H, Williams BR and Ozcelik T: Expression of *IFITM1* in chronic myeloid leukemia patients. Leuk Res 29: 283-286, 2005.
23. Bani MR, Nicoletti MI, Alkharouf NW, Ghilardi C, Petersen D, Erba E, Sausville EA, Liu ET and Giavazzi R: Gene expression correlating with response to paclitaxel in ovarian carcinoma xenografts. Mol Cancer Ther 3: 111-121, 2004.
24. Leung KC, Brce J, Doyle N, Lee HJ, Leong GM, Sjögren K and Ho KK: Regulation of growth hormone signaling by selective estrogen receptor modulators occurs through suppression of protein tyrosine phosphatases. Endocrinology 148: 2417-2423, 2007.
25. Han Y, Amin HM, Franko B, Frantz C, Shi X and Lai R: Loss of SHP1 enhances JAK3/STAT3 signaling and decreases proteosome degradation of JAK3 and NPM-ALK in ALK^+ anaplastic large-cell lymphoma. Blood 108: 2796-2803, 2006.
26. Das S, Cho J, Lambertz I, Kelliher MA, Eliopoulos AG, Du K and Tsichlis PN: Tpl2/cot signals activate ERK, JNK, and NF-κB in a cell-type and stimulus-specific manner. J Biol Chem 280: 23748-23757, 2005.
27. Aparecida Alves C, Silva ID, Villanova FE, Nicolau SM, Custódio MA, Bortoletto C and Goncalves WJ: Differential gene expression profile reveals overexpression of *MAP3K8* in invasive endometrioid carcinoma. Eur J Gynaecol Oncol 27: 589-593, 2006.
28. Clark AM, Reynolds SH, Anderson M and Wiest JS: Mutational activation of the *MAP3K8* protooncogene in lung cancer. Genes Chromosomes Cancer 41: 99-108, 2004.
29. Kurehara H, Ishiguro H, Kimura M, Mitsui A, Ando T, Sugito N, Mori R, Takashima N, Ogawa R, Fujii Y and Kuwabara Y: A novel gene, RSRC2, inhibits cell proliferation and affects survival in esophageal cancer patients. Int J Oncol 30: 421-428, 2007.
30. McDonough WS, Tran NL and Berens ME: Regulation of glioma cell migration by serine-phosphorylated P311. Neoplasia 7: 862-872, 2005.

31. Mariani L, McDonough WS, Hoelzinger DB, Beaudry C, Kaczmarek E, Coons SW, Giese A, Moghaddam M, Seiler RW and Berens ME: Identification and validation of *P311* as a glioblastoma invasion gene using laser capture microdissection. Cancer Res 61: 4190-4196, 2001.
32. King LM and Francomano CA: Characterization of a human gene encoding nucleosomal binding protein NSBP1. Genomics 71: 163-173, 2001.
33. Dumon-Jones V, Frappart PO, Tong WM, Sajithlal G, Hulla W, Schmid G, Herceg Z, Digweed M and Wang ZQ: *Nbn* heterozygosity renders mice susceptible to tumor formation and ionizing radiation-induced tumorigenesis. Cancer Res 63: 7263-7269, 2003.
34. Jensen KB and Watt FM: Single-cell expression profiling of human epidermal stem and transit-amplifying cells: Lrig1 is a regulator of stem cell quiescence. Proc Natl Acad Sci USA 103: 11958-11963, 2006.
35. Ljuslinder I, Malmer B, Golovleva I, Thomasson M, Grankvist K, Höckenstrom T, Emdin S, Jonsson Y, Hedman H and Henriksson R: Increased copy number at 3p14 in breast cancer. Breast Cancer Res 7: R719-R727, 2005.
36. Tanemura A, Nagasawa T, Inui S and Itami S: LRIG-1 provides a novel prognostic predictor in squamous cell carcinoma of the skin: immunohistochemical analysis for 38 cases. Dermatol Surg 31: 423-430, 2005.
37. Welting TJ, Peters FM, Hensen SM, van Doorn NL, Kikkert BJ, Raats JM, van Venrooij WJ and Pruijn GJ: Heterodimerization regulates RNase MRP/RNase P association, localization, and expression of Rpp20 and Rpp25. RNA 13: 65-75, 2007.
38. Inoue A, Yanagisawa M, Kimura S, Kasuya Y, Miyauchi T, Goto K and Masaki T: The human endothelin family: three structurally and pharmacologically distinct isopeptides predicted by three separate genes. Proc Natl Acad Sci USA 86: 2863-2867, 1989.
39. Jiao WJ, Xu J, Pan H, Wang TY and Shen Y: Effect of endothelin-1 in esophageal squamous cell carcinoma invasion and its correlation with cathepsin B. World J Gastroenterol 13: 4002-4005, 2007.
40. Wilson JL, Burchell J and Grimshaw MJ: Endothelins induce CCR7 expression by breast tumor cells via endothelin receptor A and hypoxia-inducible factor-1. Cancer Res 66: 11802-11807, 2006.