Prediction of individual response to platinum/paclitaxel combination using novel marker genes in ovarian cancers

Masaaki Komatsu,[1,4] Keiko Hiyama,[1]
Keiji Tanimoto,[1] Mayu Yunokawa,[1,4] Keiko Otani,[2]
Megu Ohtaki,[2] Eiso Hiyama,[3] Junzo Kigawa,[5]
Michitaka Ohwada,[6] Mitsuaki Suzuki,[6]
Nobutaka Nagai,[4] Yoshiki Kudo,[4]
and Masahiko Nishiyama[1]

Departments of [1]Translational Cancer Research and
[2]Environmetrics and Biometrics, Research Institute for Radiation
Biology and Medicine, [3]Natural Science Center for Basic
Research and Development, [4]Department of Obstetrics and
Gynecology, Graduate School of Biomedical Sciences, Hiroshima
University, Hiroshima, Japan; [5]Department of Obstetrics and
Gynecology, Tottori Medical School, Tottori, Japan; and
[6]Department of Obstetrics and Gynecology, Jichi Medical School,
Tochigi, Japan

Abstract

We attempted to identify potent marker genes using a new statistical analysis and developed a prediction system for individual response to platinum/paclitaxel combination chemotherapy in ovarian cancer patients based on the hypothesis that expression analysis of a set of the key drug sensitivity genes for platinum and paclitaxel could allow us to predict therapeutic response to the combination. From 10 human ovarian cancer cell lines, genes correlative in the expression levels with cytotoxicities of cisplatin (CDDP) and paclitaxel were chosen. We first selected five reliable prediction markers for the two drugs from 22 genes already known as sensitivity determinants and then identified another 8 novel genes through a two-dimensional mixed normal model using oligomicroarray expression data. Using expression data of genes quantified by real-time reverse transcription-PCR, we fixed the best linear model, which converted the quantified expression data into an IC_{50} of each drug. Multiple regression analysis of the selected genes yielded three prediction formulae for in vitro activity of CDDP and paclitaxel. In the same way, using the same genes selected in vitro, we then attempted to develop prediction formulae for progression-free survival to the platinum/paclitaxel combination. We therefore constructed possible formulae using different sets of 13 selected marker genes (5 known and 8 novel genes): Utility confirmation analyses using another nine test samples seemed to show that the formulae using a set of 8 novel marker genes alone could accurately predict progression-free survival ($r = 0.683$; $P = 0.042$). [Mol Cancer Ther 2006;5(3):767–75]

Introduction

Existing chemotherapies are largely palliative for advanced tumors, and numerous patients undergo a regimen without benefit (1, 2). These undesirable conditions and growing evidence that genetic difference is an important cause of variable individual drug response led to the creation of a novel chemotherapeutic strategy, personalized medicine: This would allow selection of an optimal regimen for each individual based on genomic makeup or gene expression profile (3–5). If personalized medicine is to be realized, however, molecular indicators of individual response to cancer chemotherapy are urgently required.

For advanced ovarian cancer, platinum/paclitaxel combination is the recommended chemotherapy at present, with complete response rate to the primary treatment as high as 75% (6). Nevertheless, >90% of patients develop recurrences and <25% survive for 5 years, and the prognosis for patients without clinical complete response or with recurrence disease developed within 6 months after primary treatment is consistently poor (7, 8). Despite intensive studies, none of the critical prediction markers of efficacy has been validated to date (9, 10).

The ingenious and intricate mechanism of drug sensitivity creates obstacles in predicting chemotherapeutic efficacy: Multiple genes are involved in the mechanisms, and genes or expression profiles for drug sensitivity vary significantly among tumors. DNA chip technology enables us to overview a huge number of gene expressions simultaneously, but gene expression profiles of drug sensitivity vary considerably even for the same drug, which shows the limited value of a static microarray expression profile as a marker aimed at individualizing patient therapy (5, 11, 12).

Selection of a set of truly significant genes and understanding of their interplay are of key importance in prediction of individual response to drug therapies.

Using cDNA array, we previously sorted out 12 potential prediction marker genes of *in vitro* activity for eight drugs, including cisplatin (CDDP) and paclitaxel. We then fixed prediction formulae, which embraced the variable expressions of the 12 genes, and arranged them to predict the efficacy of the drugs along with individual clinical responses to 5-fluorouracil (5). The potent predictive value suggested that the better markers had been selected and that evaluation of the variable expression of the 12 genes by multiple regression analysis might work well as a prediction model. Nevertheless, the final selection of marker genes was still limited to functionally proven genes. In fact, no definitive way to determine the critical marker genes from a huge number of candidates at one stroke has yet been established.

In this study, we focused on platinum/paclitaxel therapy in ovarian cancer and attempted to select more powerful sensitivity markers using a new statistical method of array expression data: a two-dimensional normal mixture model. We then did multiple regression analysis using selected genes with and without proven functional significance to drug sensitivity, seeking to develop more reliable prediction models of the *in vitro* activity of CDDP and paclitaxel and the clinical efficacy of the combination.

Materials and Methods

Chemicals

CDDP and paclitaxel were generously provided by Bristol-Myers K.K. (Tokyo, Japan). All other chemicals were analytic grade and were purchased from Wako Pure Chemicals (Osaka, Japan) and Sigma (St. Louis, MO).

Cells

The 10 human ovarian cancer cell lines used in this study were kindly provided as follows: an ovarian serous adenocarcinoma cell line, SKOV3 (N. Nagai, Hiroshima University, Hiroshima, Japan); KF28 and its CDDP/paclitaxel-resistant variant (Dr. Y. Kikuchi, National Defense Medical College, Saitama, Japan); and KF, SHIN3, and five ovarian clear cell adenocarcinoma cell lines, KK, TAYA, RMG-1, OVISE, and OVTOKO (M. Suzuki, Jichi Medical School, Tochigi, Japan). All cell lines were cultured in RPMI 1640 (Life Technologies, Inc., Grand Island, NY) containing 10% heat-inactivated fetal bovine serum (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.

Patients and Human Tissue Samples

Human tumor specimens were collected from 23 patients in International Federation of Gynecology and Obstetrics stages III and IV advanced-stage ovarian cancer (1 case of stage IIIb, 15 cases of stage IIIc, and 7 cases of stage IV) at initial laparotomy. All of the patients had histologically proven ovarian serous adenocarcinoma and had not received any treatment before tumor sampling. The patients were all <80 years old (median, 58; range,

42–76) with performance status (WHO) 0 to 2 without significant baseline laboratory abnormalities, and life expectancy was estimated as >3 months. All received platinum/paclitaxel chemotherapy as postoperative adjuvant chemotherapy after noncurative operations. Paclitaxel was given by continuous infusion at a dose of 175 mg/m^2 over 3 hours followed by platinum, carboplatin (area under the curve, 5.0 infused over 30 minutes; ref. 6). Cycles of this regimen were repeated every 21 days and patients received a median of six cycles of the treatment (range, 4–10), which is the current standard number of cycles for the patients with advanced-stage ovarian cancer. Because most of the cases (>75%) generally achieve complete response after primary treatment, progression-free survival (PFS) was estimated as the prime indicator of therapeutic effect in this study. PFS was measured from the start of treatment until the first documentation of progression, date of last contact, or start of subsequent antitumor therapy. Computed tomography (computed tomography scanning) was done every two or three chemotherapy cycles to estimate PFS. Written informed consent was obtained from all patients, and the protocol was approved by institutional ethics committees. The tumor specimens were divided into two groups according to collection date (14 earlier-obtained and 9 later-obtained samples). The former were used as experimental samples to develop a prediction model and the latter were used as test samples to confirm the predictive accuracy of the developed model. These samples were stored at −80°C until use.

Extraction and Purification of RNA

For gene expression analysis, exponentially grown cultured cells (2 × 10^6) were collected after two washings with PBS. The cell pellets were immediately frozen in liquid nitrogen and stored at −80°C until use. 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 Technologies, Palo Alto, CA).

Oligonucleotide Array Analysis

Codelink Expression Bioarray System (Amersham Biosciences, Tokyo, Japan) was used according to the manufacturer's protocol. Briefly, first-strand cDNA was generated from 1 μg total RNA of cell lines using reverse transcriptase and a T7 primer, and second-strand cDNA was produced using DNA polymerase mix and RNase H. cRNA was generated via an *in vitro* transcription reaction using T7 RNA polymerase and biotin-11-UTP (Perkin-Elmer, Boston, MA), which was quantified by spectrometry and checked using Agilent Technologies 2100 Bioanalyzer. cRNA (10 μg) was then fragmented and hybridized to a Codelink Uniset Human 20K I Bioarray containing 19,981 probes with positive and negative bacterial control probes. After hybridization, the arrays were rinsed and labeled with streptavidin-Cy5, scanned using Agilent DNA Microarray Scanner (Agilent Technologies), and analyzed with Codelink Expression Analysis Software. Expression levels

were normalized to the median expression value of the entire spot array. The microarray data were registered to the Gene Expression Omnibus under GE accession no. GSE3001.[7]

Real-time Reverse Transcription-PCR

Total RNA (2 μg) extracted from each cell line was reverse transcribed using High-Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA). Aliquot (1/200) of the cDNA (equivalent to 10 ng total RNA) was subjected to real-time reverse transcription-PCR (RT-PCR). Real-time RT-PCR was done using TaqMan Gene Expression Assays (Applied Biosystems) or originally designed TaqMan probe (Applied Biosystems) and primer set. Each reaction was carried out in triplicate for both cell lines and tissues using ABI Prism 7900HT Sequence Detection System (Applied Biosystems). These triplicate measurements were averaged, and relative gene expression levels were calculated as a ratio to glyceraldehyde-3-phosphate dehydrogenase expression level.

Cytotoxicity Assay

Drug-induced cytotoxicity was evaluated by conventional 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide dye reduction assay. Cells were counted with hemocytometer and seeded in 96-microwell plates (Nunclon, Nunc, Roskilde, Denmark) at a density of 4×10^3 per well in RPMI 1640 with 10% fetal bovine serum. After 24-hour incubation, the medium was replaced and cells were exposed to the indicated drug concentrations for 72 hours, after which 10 μL of 0.4% 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide reagent and 0.1 mol/L sodium succinate were added to each well. After 2-hour incubation, 150 μL 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$_{50}$ (drug concentration of 50% absorbance of control).

Rank Correlation Analysis

The rank correlation coefficient (Spearman 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 some outliers. The statistical significance was evaluated with P 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

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, called normalization, and identify differentially expressed genes between two cells showing different biological behaviors based on the functional

status of the genes (13, 14). We used this mathematical model to select the most potent prediction marker genes from the large number of candidates that were sort out as correlative genes with drug sensitivity in the first microarray screening analysis.

The mathematical model for the microarray data with each channel (denoted as $Y_i^{(1)}$ and $Y_i^{(2)}$, where $Y_i^{(1)}$ and $Y_i^{(2)}$ are the suitably transformed and normalized gene expression measurements) is

$$
\begin{cases}
Y_i^{(1)} = \tau_i^{(1)} \alpha_i + \beta_i + \varepsilon_i^{(1)}, \\
Y_i^{(2)} = \tau_i^{(2)} \alpha_i + \beta_i + \varepsilon_i^{(2)}.
\end{cases}
$$

The symbols $\tau_i^{(1)}$ and $\tau_i^{(2)}$ represent the expression status of gene i in the query and reference samples (if gene i is on, $\tau_i = 1$; if gene i is off, $\tau_i = 0$). The symbol α_i represents the expression intensity of gene i when it is on. The terms of "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. The symbol β_i denotes systematic error common to channels 1 and 2. The symbols $\varepsilon_i^{(1)}$ and $\varepsilon_i^{(2)}$ indicate random errors, which are mutually independent. The variable transformation $U = Y_i^{(1)} + Y_i^{(2)}$ and $V = Y_i^{(1)} - Y_i^{(2)}$ is introduced, and the gene is characterized by the joint sum and difference of its expression intensities in two samples. The plot (U,V) is called the S-D plot. Then, a two-dimensional mixed normal model with four components is applied (Fig. 1A). The genes of interest to us belong to the *region c* (on, off) and *region d* (off, on), where V reflects the true difference of expression intensities between the two samples. Genes belonging to the *regions a* and *b* are not informative, because V reflects only the measurement error. 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.

Multiple Regression Analysis

The multiple regression model is expressed by $Y_i = \theta_0 + \theta_1 x_{i1} + \ldots \theta_p x_{ip} + \varepsilon_i$, where ε_i is a random error and $\theta_0, \theta_1, \ldots, \theta_p$ denote parameters to be estimated. Trimmed Least Squares Regression, a reliable regression method based on an extended algorithm of Least Median Squares Regression proposed by Rousseeuw (15), was done to explore one set of effective genes whose expression levels would explain the value of drug efficacy (IC$_{50}$ or PFS), in which a value of Y_i (a logarithmic-transformed value of IC$_{50}$ of cell or PFS) is a response variable and values of $x_{i1}, x_{i2}, \ldots, x_{ip}$ (expression levels of genes) are explanatory variables (5). A set of effective genes was explored by referring to the value of Akaike's information criterion (AIC) per sample or by checking residuals graphically. When the correlation coefficient (r) becomes
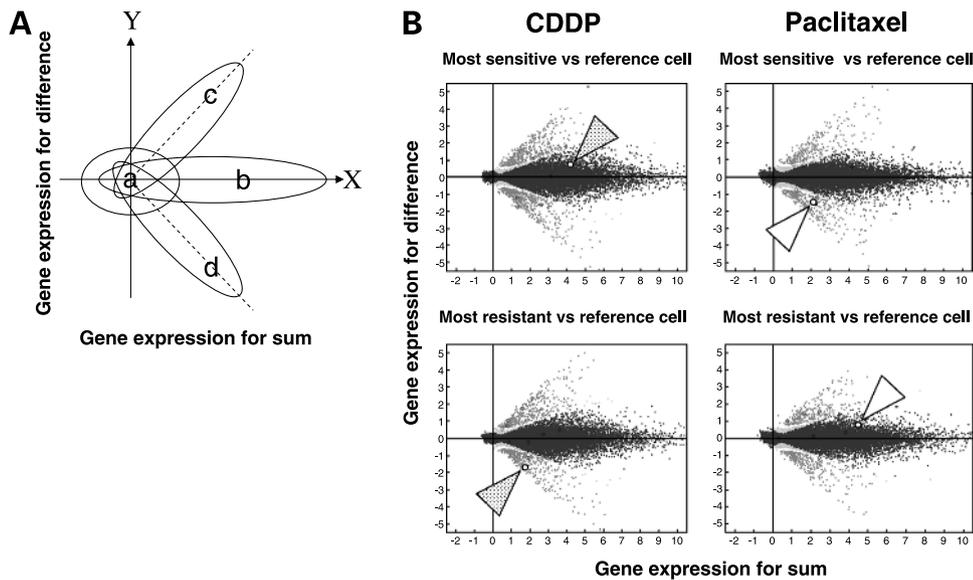
---

[7] http://www.ncbi.nlm.nih.gov.geo/

Figure 1. Selection of potent prediction marker genes using a two-dimensional mixed normal model. **A,** two-dimensional mixed normal model with four components for S-D plot shown schematically. *X axis,* sums of expression intensities of the query and reference samples; *Y axis,* differences. The expression status of genes inside regions *a* to *d* are (off, off), (on, on), (on, off), and (off, on) in the query and reference samples, respectively. The terms ''on'' and ''off'' represent the functional status of a gene. **B,** for potent drug sensitivity markers, we explored genes differentially expressed between the most drug-resistant (e.g., *dotted arrows*) or drug-sensitive cells (e.g., *open arrows*) and cells showing median IC_{50} for each drug using the model.

closer to 1 and AIC per sample grows smaller, the model becomes more predictive and confirmative. We used the software NLReg, developed by Ohtaki et al.,[8] which implemented a robust regression analysis (5). Outliers were identified by referring to the value of AIC per sample or checking residuals graphically, and a set of effective genes that satisfied the value of IC_{50} or PFS was explored. The NLReg analysis software provides estimated θ_p with P, where a lower P indicates a lower probability for the observation that θ_p could be 0 in the formula. A positive θ indicates that the corresponding explanatory variable acts as a positive factor in the formulae, whereas a negative θ indicates the inverse action of the variable. The levels of θ do not directly account for the importance of the explanatory variable when values (or levels) of the explanatory variables differ considerably.

Results

Prediction Markers Selected from Genes Known as Drug Sensitivity Determinants

A variety of genes have been shown as drug sensitivity determinants to date, so we first attempted to select prediction markers from these genes. After a search through the literature on gene-drug sensitivity (or resistance) to CDDP and paclitaxel on the National Library of Medicine's PubMed, 22 candidate genes whose functional significance have been shown in more than two reports were selected (16–28). These 22 candidates were subjected to real-time RT-PCR analysis and analyzed for correlation with the quantified expression levels of drug sensitivity in 10 ovarian cancer cell lines. The observed IC_{50}s for CDDP (302.7–9,118.1 nmol/L) and paclitaxel (4.4–2,104.4 nmol/L) significantly varied among cell lines, and expression levels of *IL6,*

BCL2, VEGF, and *ERCC2* for CDDP and those of *ABCB1, CYP2C8*, and *CYP3A4* for paclitaxel were found to correlate with the sensitivity to each drug ($P < 0.05$). Nevertheless, analysis also revealed that expressions of *CYP2C8* and *CYP3A4* were undetectable in 7 and 6 of the 10 cell lines investigated, respectively. The insecure expression of *CYP2C8* and *CYP3A4* led us to omit them from prediction markers, and we selected a set of the other 5 correlative genes as a potent prediction marker of efficacy of platinum/paclitaxel combination therapy (Table 1).

Screening of Novel Candidate Genes Using Oligonucleotide Array

Even so, none of the genes selected are always critical in drug sensitivity mechanisms. To seek more powerful marker genes, we therefore did oligonucleotide array analysis using Codelink Uniset Human 20K I Bioarray containing 19,981 probes. The normalized expression level of each gene and the IC_{50} for each drug in 10 cell lines were ranked and the correlation between ranks of the two sets of measurements was evaluated. The rank correlation analysis showed that 114 and 303 genes correlated with cellular resistance to CDDP and paclitaxel, respectively, in their expression levels ($P < 0.01$; supplement data_rank correlation.xls).[9] To select the most potent prediction marker genes from such a large number of candidates, we further explored differentially expressed genes between the most drug-resistant or drug-sensitive cells and cell lines with median IC_{50} for each drug, using a novel statistical method, a two-dimensional normal model. This method provided 8 and 12 genes for paclitaxel and CDDP, respectively, as highly reliable candidates for prediction markers (>0.8 of probability of heterogeneity; Fig. 1).

---

[8] http://apollo.rbm.hiroshima-u.ac.jp/

These 20 selected genes were subjected to real-time RT-PCR analysis to confirm correlation with drug sensitivity in the quantified expression levels. To select more potent prediction marker genes, the selection criterion was determined as $P < 0.01$ and $R > 0.5$ in the linear regression analysis. From the correlative genes, we then omitted *DBC1* whose expression was undetectable in 6 of the 10 cell lines, and finally selected a total of 8 genes as novel prediction markers: *MYO5C, SPINK1, ARMCX3, PLEK2,* and *PRSS11* for CDDP, and *TNFSF13B, IFIT3* and *BTN3A2* for paclitaxel (Table 1).

Prediction Model of Sensitivity to CDDP and Paclitaxel *In vitro*

CDDP and paclitaxel seemed to have plural sensitivity marker genes, except when we selected such markers for paclitaxel from 22 genes already known as drug sensitivity determinants. Using expression data of these three sets of selected genes quantified by real-time RT-PCR, we did multiple regression analysis to compose prediction models for the *in vitro* activity of CDDP and paclitaxel. As results, we fixed a total of three prediction formulae of *in vitro* drug sensitivity, two prediction formulae for CDDP and one for paclitaxel, to show the highest fitness (Table 2; Fig. 2). The observed correlation coefficient and AIC per sample indicated potent predictive values of all the fixed formulae, and the model using novel marker genes appeared among the better in the prediction (Fig. 2). In the prediction formulae for CDDP sensitivity using expression data of four known genes, the estimated P for each θ_p of *BCL2* and *VEGF* was lower than those of the others (Table 2). These findings could be interpreted that these genes played a more important role in the prediction than the other two genes, because the lower P indicates a lower probability of obtaining the observation that θ_P could be 0 in the formula.

Likewise, *MYO5C, SPINK1,* and *PRSS11* for CDDP and *IFIT3* for paclitaxel showed greater sensitivity to the corresponding drug than the other genes selected as novel prediction markers.

Prediction Model of Clinical Response to Platinum/Paclitaxel Chemotherapy

Because potent predictive value was shown in the prediction models of *in vitro* drug sensitivity, we attempted to construct a prediction model of clinical response (i.e., PFS) to platinum/paclitaxel combination chemotherapy in the same way using the same genes selected in *in vitro*. To construct a prediction model, we first used 14 tumor specimens from 23 collected specimens. The samples were subjected to real-time RT-PCR analysis to quantify the expression levels of 13 selected marker genes, and we developed clinical prediction models using two sets of the selected marker genes, 5 genes known as drug sensitivity determinants and 8 novel marker genes.

Because the 13 selected genes were powerful indicators of *in vitro* drug sensitivity even when used alone, we investigated the correlation between expression of each gene in tumor samples and clinical response to platinum/paclitaxel chemotherapy for a start. However, none of the selected genes alone could predict PFS of platinum/paclitaxel chemotherapy. Expression data of the set of 5 genes known as sensitivity determinants and the other set of 8 novel marker genes were then subjected to multiple regression analysis. In contrast to the findings in the analysis using each of the selected 13 genes or each set of genes related to sensitivity to either CDDP or paclitaxel alone, analysis using 14 data sets of gene expression and clinical response provided two prediction formulae for PFS that showed the highest fitness for each set of prediction marker genes (Table 3). We also attempted to fix other

Table 1. Selected prediction marker genes

Drug	Gene	Accession no.	Microarray			Real-time RT-PCR	
			Probability of heterogeneity	r	P	r	P
Genes known as sensitivity determinants							
CDDP	*IL6*	NM_000600	—	—	—	0.806	<0.0001
	BCL2	NM_000633	—	—	—	0.782	<0.0001
	VEGF	NM_003376	—	—	—	0.401	0.028
	ERCC2	NM_000400	—	—	—	0.403	0.027
Paclitaxel	*ABCB1*	NM_000927	—	—	—	0.985	<0.0001
Genes selected as novel markers							
CDDP	*MYO5C*	NM_018728	1.00	0.964	<0.01	0.617	0.0003
	SPINK1	NM_003122	0.98	0.865	<0.01	0.612	0.0003
	ARMCX3	NM_016607	1.00	0.861	<0.01	0.734	<0.0001
	PLEK2	NM_016445	1.00	0.810	<0.01	0.503	0.004
	PRSS11	NM_002775	1.00	−0.794	<0.01	−0.517	0.004
Paclitaxel	*TNFSF13B*	NM_006573	0.98	0.844	<0.01	0.663	<0.0001
	IFIT3	NM_001549	0.94	0.806	<0.01	0.769	<0.0001
	BTN3A2	NM_007047	0.92	0.773	<0.01	0.568	0.001

NOTE: We selected 8 novel genes as the most potent markers using a rank correlation analysis ($P < 0.01$) and a two-dimensional mixed normal model (PH > 0.8). All the selected prediction marker genes were confirmed in their correlation of expression level with drug sensitivity through quantitative real-time RT-PCR. —, not analyzed.

Table 2. Estimated coefficients (θ_p) in *in vitro* prediction formulae for IC$_{50}$ (ln[IC$_{50}$] = $x_{i1}\theta_1 + x_{i2}\theta_2 + \ldots + x_{ip}\theta_p + \varepsilon_i$)

	CDDP			Paclitaxel	
x_{ip}	θ_p	P	x_{ip}	θ_p	P
Formulae using expression data of genes known as sensitivity determinants					
ln[*IL6*]	−0.069	0.133	ln[*ABCB1*]	—	—
ln[*BCL2*]	−0.393	0.0002			
ln[*VEGF*]	0.834	<0.0001			
ln[*ERCC2*]	−0.320	0.111			
ε_i	5.482			—	
Formulae using expression data of genes selected as novel markers					
ln[*MYO5C*]	0.423	<0.0001	ln[*TNFSF13B*]	0.087	0.649
ln[*SPINK1*]	0.141	<0.0001	ln[*IFIT3*]	0.676	0.0001
ln[*ARMCX3*]	0.133	0.011	ln[*BTN3A2*]	0.323	0.180
ln[*PLEK2*]	−0.124	0.105			
ln[*PRSS11*]	−0.126	0.0001			
ε_i	6.262		ε_i	3.169	

NOTE: Gene names inside brackets indicate the expression level of indicated gene.

prediction formulae using several different sets of marker genes, including the sets of genes related to sensitivity to either CDDP or paclitaxel alone, but none of them precisely predicted PFS of platinum/paclitaxel chemotherapy. The observed correlation coefficient and AIC per sample in the fixed formulae indicated that prediction for PFS was more precise when using a set of 8 novel genes than a set of 5 known genes.

To confirm the predictive accuracy of the fixed formulae, we examined additional 9 tumor samples: The expression levels of the 13 selected genes were quantified by real-time RT-PCR and then PFS was predicted by the developed formulae using their expression data. Despite the limited number of samples, the results showed that PFS was reliably predictable only when using the set of 8 novel marker genes (Fig. 3).

Discussion

In this study, starting with the hypothesis that expression analysis of a set of key drug sensitivity genes for platinum and paclitaxel could allow us to predict therapeutic
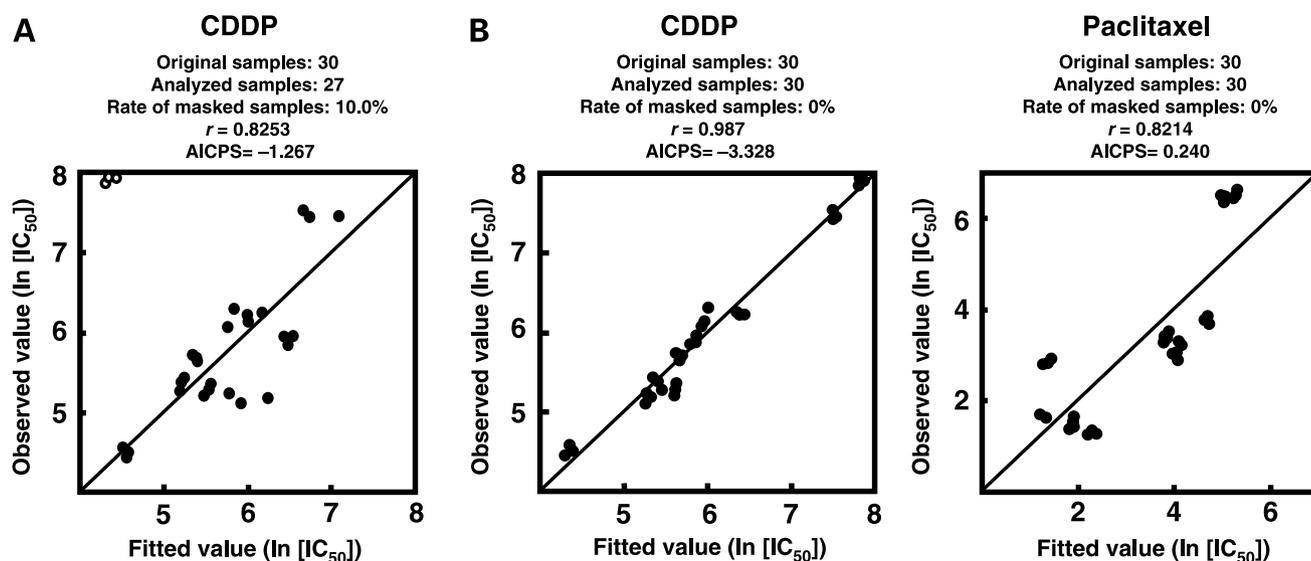

Figure 2. Predictive accuracy of the fixed formulae for *in vitro* sensitivity to paclitaxel and CDDP. Using multiple regression analysis, we developed formulae to predict the IC$_{50}$ of each drug using the variable expression data of selected marker genes. **A,** prediction for CDDP-induced cytotoxicity using a set of 4 genes selected from genes known as drug sensitivity determinants. **B,** prediction of drug sensitivity using a set of 5 and 3 genes selected as novel markers for CDDP and paclitaxel, respectively. In this analysis, 30 independent data sets, expression levels of selected genes, and IC$_{50}$s for 10 ovarian cancer cell lines were used. The most suitable model was developed by eliminating the outliers using the value of AIC per sample (*AICPS*) and checking residuals graphically. ●, analyzed sample; ○, a masked outlier.

Table 3. Estimated coefficients (θ_p) in prediction formulae for clinical response to platinum/paclitaxel chemotherapy (ln[PFS] = $x_{i1}\theta_1 + x_{i2}\theta_2 + \ldots + x_{ip}\theta_p + \varepsilon_i$)

x_{ip}	θ_p	P
Formulae using expression data of genes known as sensitivity determinants		
ln[*IL6*]	0.203	0.056
ln[*BCL2*]	−0.081	0.496
ln[*VEGF*]	0.664	0.001
ln[*ERCC2*]	−1.479	<0.0001
ln[*ABCB1*]	0.337	0.023
ε_i	6.124	
Formulae using expression data of genes selected as novel markers		
ln[*MYO5C*]	0.300	0.224
ln[*SPINK1*]	−0.170	0.114
ln[*ARMCX3*]	−0.818	0.115
ln[*PLEK2*]	−0.059	0.843
ln[*PRSS11*]	0.495	0.087
ln[*TNFSF13B*]	−0.613	0.034
ln[*IFIT3*]	0.843	0.096
ln[*BTN3A2*]	−0.340	0.290
ε_i	6.340	

response to the combination therapy, we selected 5 better marker genes known to be drug sensitivity determinants (4 for CDDP and 1 for paclitaxel) and identified another 8 genes as novel potent markers (5 for CDDP and 3 for paclitaxel). We used statistical analysis based on oligomicroarray expression data, a two-dimensional mixed normal model, and subsequent real-time RT-PCR. Although the functional significance of the 8 novel genes in drug sensitivity was poorly understood, their expression levels were shown to be more correlative with cellular sensitivities to the two drugs *in vitro* than those of the 5 drug sensitivity genes. We then determined expression data of the selected genes quantified by real-time RT-PCR as probable predictors and fixed the best linear model, which examined the variable expressions of the component genes and arranged them to predict the efficacy of the drugs, using multiple regression analysis. Expression data of the three sets of genes (4 known genes for CDDP and 5 and 3 novel genes for CDDP and paclitaxel, respectively) provided prediction formulae for the *in vitro* activity of CDDP and paclitaxel and showed the highest fitness. In the same way, using the same genes selected *in vitro*, we attempted to develop prediction formulae for individual clinical response (i.e., PFS) to the platinum/paclitaxel combination. We constructed several possible formulae using different sets of 13 selected marker genes and finally found that the formulae using a set of all of the 8 novel marker genes alone could accurately predict PFS based on confirmation analyses of their utilities using additional nine tumor testing samples.

To develop a precise efficacy prediction model for chemotherapy, there were two major obstacles (5): Identifying the key marker genes from a large number of candidates and figuring out their interplay. We believe we overcome the obstacles by developing a potent model

system that can predict the *in vitro* activity of CDDP and paclitaxel and clinical response to the combination therapy with a numerical value of PFS. The potent predictive value of the fixed formulae indicates that we probably succeeded in selecting the better prediction marker genes and precisely estimating their interaction in the expression levels.
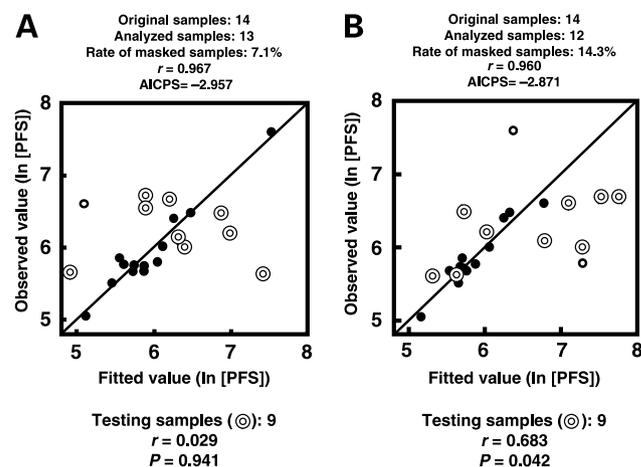

Figure 3. Predictive accuracy of the fixed formulae for clinical response (i.e., PFS) to platinum/paclitaxel combination chemotherapy. We developed a formula to predict PFS of platinum/paclitaxel combination chemotherapy in the same way using the same genes selected *in vitro*. To construct the formulae, we used 14 tumor specimens as modeling samples, and another 9 samples were used as testing samples to confirm the predictive values. The best-fitted prediction model using each set of 5 known drug sensitivity genes (**A**) and 8 novel genes (**B**) is shown together with prediction results in nine testing samples (*double open circle*). Among several possible prediction formulae, the formula using the set of all of the 8 novel marker genes alone showed the most potent predictive value.

Our work provided two sets of potent prediction marker genes: 5 genes known to be drug sensitivity determinants and another novel 8 genes. All of the 5 known genes (*IL6, BCL2, VEGF*, and *ERCC2* for CDDP and *ABCB1* for paclitaxel) were widely recognized as being of key importance among a variety of drug sensitivity genes for the two drugs even when used alone (17–21). Nevertheless, the 8 novel genes were more correlative with corresponding drug sensitivity than the 5 known genes in expression levels, and a combination of the 8 genes alone could work well in the prediction of clinical response to platinum/paclitaxel combination chemotherapy. The 8 genes might play more important roles in the drug-induced cytotoxicity than the 5 known genes. Their functions remain little known, but various results suggest their possible roles in drug sensitivity: *TNFSF13B* encodes a cytokine that belongs to the tumor necrosis factor ligand family and acts as a potent B-cell activator in proliferation and differentiation (29); the function of *IFIT3*, an IFN-induced gene, is unknown to date (30); *BTN3A2* is a butyrophilin subfamily associated with T-cell proliferation and cytokine production (31); the product of *MYO5C* is likely to act on actin-based membrane trafficking (32); *SPINK1* encodes pancreatic secretory trypsin inhibitor (33, 34). Furthermore, *ARMCX3* encodes a protein of the ALEX family, which may play a role in tumor suppression (35); *PLEK2* encodes a membrane-associating protein that may help orchestrate cytoskeletal arrangement (36, 37); *PRSS11* (*HTRA1*) encodes HtrA1, a member of the trypsin family of serine proteases, a possible regulator of the insulin-like growth factors that have been suggested as being associated with ovarian carcinogenesis and progression (38). *PRSS11* (*HTRA1*) has also been suggested as a candidate tumor suppressor gene involved in promoting serine protease–mediated cell death, and down-regulation of *HTRA1* may contribute to malignant phenotypes in various cancers, including ovarian cancer (39, 40).

This is the first attempt to use a two-dimensional mixed normal model in the selection of drug sensitivity determinant genes. The suggested significance of selected genes as prediction markers and the indicated differences in expression levels were highly confirmative in subsequent real-time RT-PCR analysis. These findings can be interpreted to mean that this statistical method will likely work well to identify novel marker genes from numerous candidates. Even so, the multifactorial mechanisms of drug sensitivity did not allow us to predict response to a drug by expression of any single gene. However, as shown in our previous studies, multiple regression analysis was promising in developing a sensitivity prediction model of a chemotherapy based on understanding the interaction of a set of key genes in their expression levels (5). Recently, much attention has been focused on microarray as a tool of the molecular classification of disease, individual response to drugs, and survival. There have been several hopeful results in the classification of disease and patient survival prognosis in ovarian cancer, but the approach to personalizing therapy has not yet shown any significant effect (12, 41). Our developed models may have some advantages in the prediction of drug response, and we believe that this developmental approach can contribute to a more accurate prediction model. The combined use of rank correlation analysis and a two-dimensional mixed normal model will, we think, contribute to improving the heretofore limited utility of microarray analysis in the selection of significant genes.

Nevertheless, we used a set of genes selected as sensitivity markers to CDDP and paclitaxel to predict the efficacy of the two drugs in combination in this work, but the platinum used in our clinical study was not CDDP but carboplatin. Although the action mechanism of carboplatin is considered to be the same as that of CDDP (42), there may be some differences in the sensitivity determinants between CDDP and carboplatin. In parallel with the functional roles of the selected 8 genes in drug sensitivity, the selection of potent marker genes for efficacy of carboplatin is now in progress. Expression-sensitivity correlation analyses also need to be done in the combination setting, although the potential of an *in vitro* sensitivity-evaluation model that precisely reflects clinical response to combination therapy is limited. Furthermore, in the clinical study, the number and the cancer type (serous adenocarcinoma) of patients were limited. Although most of the cases (>75%) generally achieve complete response after primary treatment, prediction of tumor response (regression rate) is also our interests. The set of novel 8 genes showed the advantage in prediction of PFS for the platinum/paclitaxel combination, but it is obvious that practical usefulness needs to be evaluated by a larger prospective study, including patients with clear cell carcinoma and prediction formulae for tumor regression are eagerly awaited. We are now planning such a prospective clinical study along with continuing our search for the functional roles of the selected 8 genes in drug sensitivity and more powerful predictive marker genes for drug sensitivity.

References

1. Gibbs JB. Mechanism-based target identification and drug discovery in cancer research. Science 2000;287:1969–73.

2. Nishiyama M. Cancer pharmacogenomics: progress and obstacles toward the goal. Int J Clin Oncol 2005;10:3–4.

3. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science 1999;286:487–91.

4. Marsh S, McLeod HL. Cancer pharmacogenetics. Br J Cancer 2004;90:8–11.

5. Tanaka T, Tanimoto K, Otani K, et al. Concise prediction models of anticancer efficacy of 8 drugs using expression data from 12 selected genes. Int J Cancer 2004;111:617–26.

6. Ozols RF. Update on Gynecologic Oncology Group (GOG) trials in ovarian cancer. Cancer Invest 2004;22(Suppl 2):11–20.

7. McGuire WP, Hoskins WJ, Brady MF, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. N Engl J Med 1996;334:1–6.

8. Ozols RF, Bundy BN, Greer BE, et al. Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study. J Clin Oncol 2003;21:3194–200.

9. Hartmann LC, Lu KH, Linette GP, et al. Gene expression profiles predict early relapse in ovarian cancer after platinum-paclitaxel chemotherapy. Clin Cancer Res 2005;11:2149–55.

10. Berchuck A, Iversen ES, Lancaster JM, et al. Patterns of gene

expression that characterize long-term survival in advanced stage serous ovarian cancers. Clin Cancer Res 2005;11:3686 – 96.

11. Staunton JE, Slonim DK, Coller HA, et al. Chemosensitivity prediction by transcriptional profiling. Proc Natl Acad Sci U S A 2001; 98:10787 – 92.

12. McLeod HL, Evans WE. Pharmacogenomics: unlocking the human genome for better drug therapy. Annu Rev Pharmacol Toxicol 2001; 41:101 – 21.

13. Ohtaki M, Otani K, Satoh K, Kawamura T, Hiyama K, 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 Biomet 2005;26:31 – 48.

14. Hiyama K, Otani K, Ohtaki M, et al. Differentially expressed genes throughout the cellular immortalization processes are quite different between normal human fibroblasts and endothelial cells. Int J Oncol 2005;27:87 – 95.

15. Rousseeuw PJ. Least median of squares regression. J Am Stat Assoc 1984;79:871 – 80.

16. Efferth T, Volm M. Pharmacogenetics for individualized cancer chemotherapy. Pharmacol Ther 2005;107:155 – 76.

17. Borsellino N, Belldegrun A, Bonavida B. Endogenous interleukin 6 is a resistance factor for *cis*-diamminedichloroplatinum and etoposide-mediated cytotoxicity of human carcinoma cell lines. Cancer Res 1995; 55:4633 – 9.

18. Murata T, Haisa M, Uetsuka H, et al. Molecular mechanism of chemoresistance to cisplatin in ovarian cancer cell lines. Int J Mol Med 2004;13:865 – 8.

19. Manenti L, Riccardi E, Marchini S, et al. Circulating plasma vascular endothelial growth factor in mice bearing human ovarian carcinoma xenograft correlates with tumor progression and response to therapy. Mol Cancer Ther 2005;4:715 – 25.

20. Gurubhagavatula S, Liu G, Park S, et al. *XPD* and *XRCC1* genetic polymorphisms are prognostic factors in advanced non-small-cell lung cancer patients treated with platinum chemotherapy. J Clin Oncol 2004; 22:2594 – 601.

21. Duan Z, Brakora KA, Seiden MV. Inhibition of *ABCB1* (*MDR1*) and *ABCB4* (*MDR3*) expression by small interfering RNA and reversal of paclitaxel resistance in human ovarian cancer cells. Mol Cancer Ther 2004;3:833 – 8.

22. Taniguchi R, Kumai T, Matsumoto N, et al. Utilization of human liver microsomes to explain individual differences in paclitaxel metabolism by CYP2C8 and CYP3A4. J Pharmacol Sci 2005;97:83 – 90.

23. Kelley SL, Basu A, Teicher BA, Hacker MP, Hamer DH, Lazo JS. Overexpression of metallothionein confers resistance to anticancer drugs. Science 1988;241:1813 – 5.

24. Hengstler JG, Lange J, Kett A, et al. Contribution of c-erbB-2 and topoisomerase IIα to chemoresistance in ovarian cancer. Cancer Res 1999;59:3206 – 14.

25. Bani MR, Nicoletti MI, Alkharouf NW, et al. Gene expression correlating with response to paclitaxel in ovarian carcinoma xenografts. Mol Cancer Ther 2004;3:111 – 21.

26. Huang J, Mohanty S, Basu A. Cisplatin resistance is associated with deregulation in protein kinase C-δ. Biochem Biophys Res Commun 2004; 316:1002 – 8.

27. Papouli E, Cejka P, Jiricny J. Dependence of the cytotoxicity of DNA-damaging agents on the mismatch repair status of human cells. Cancer Res 2004;64:3391 – 4.

28. Xu ZY, Loignon M, Han FY, Panasci L, Aloys R. Xrcc3 induces cisplatin resistance by stimulation of rad51-related recombinational repair, S-phase checkpoint activation, and reduced apoptosis. J Pharmacol Exp Ther 2005;314:495 – 505.

29. Novak AJ, Grote DM, Stenson M, et al. Expression of BLyS and its receptors in B-cell non-Hodgkin lymphoma: correlation with disease activity and patient outcome. Blood 2004;104:2247 – 53.

30. de Veer MJ, Sim H, Whisstock JC, Devenish RJ, Ralph SJ. IFI60/ISG60/IFIT4, a new member of the human IFI54/IFIT2 family of interferon-stimulated genes. Genomics 1998;54:267 – 77.

31. Rhodes DA, Stammers M, Malcherek G, Beck S, Trowsdale J. The cluster of *BTN* genes in the extended major histocompatibility complex. Genomics 2001;71:351 – 62.

32. Rodriguez OC, Cheney RE. Human myosin-Vc is a novel class V myosin expressed in epithelial cells. J Cell Sci 2002;115:991 – 1004.

33. Le Marechal C, Chen JM, Le Gall C, et al. Two novel severe mutations in the pancreatic secretory trypsin inhibitor gene (*SPINK1*) cause familial and/or hereditary pancreatitis. Hum Mutat 2004;23:205.

34. Weiss FU, Simon P, Witt H, et al. *SPINK1* mutations and phenotypic expression in patients with pancreatitis associated with trypsinogen mutations. J Med Genet 2003;40:e40.

35. Kurochkin IV, Yonemitsu N, Funahashi SI, Nomura H. ALEX1, a novel human armadillo repeat protein that is expressed differentially in normal tissues and carcinomas. Biochem Biophys Res Commun 2001;280:340 – 7.

36. Inazu T, Yamada K, Miyamoto K. Cloning and expression of pleckstrin 2, a novel member of the pleckstrin family. Biochem Biophys Res Commun 1999;265:87 – 93.

37. Hu MH, Bauman EM, Roll RL, Yeilding N, Abrams CS. Pleckstrin 2, a widely expressed paralog of pleckstrin involved in actin rearrangement. J Biol Chem 1999;274:21515 – 8.

38. Sayer RA, Lancaster JM, Pittman J, et al. High insulin-like growth factor-2 (IGF-2) gene expression is an independent predictor of poor survival for patients with advanced stage serous epithelial ovarian cancer. Gynecol Oncol 2005;96:355 – 61.

39. Baldi A, De Luca A, Morini M, et al. The HtrA1 serine protease is down-regulated during human melanoma progression and represses growth of metastatic melanoma cells. Oncogene 2002;21:6684 – 8.

40. Chien J, Staub J, Hu SI, et al. A candidate tumor suppressor HtrA1 is downregulated in ovarian cancer. Oncogene 2004;23:1636 – 44.

41. Berchuck A, Lancaster JM, Iversen ES, et al. Molecular profiling of patients with ovarian cancer and the potential for individualized therapy. In: Perry MC, editor. American Society of Clinical Oncology Educational Book 2005. Alexandria: American Society of Clinical Oncology; 2005. p. 416 – 20.

42. Taylor AE, Wiltshaw E, Gore ME, et al. Long-term follow-up of the first randomized study of cisplatin versus carboplatin for advanced epithelial ovarian cancer. J Clin Oncol 1994;12:2066 – 70.

Molecular Cancer Therapeutics

Prediction of individual response to platinum/paclitaxel combination using novel marker genes in ovarian cancers

Masaaki Komatsu, Keiko Hiyama, Keiji Tanimoto, et al.

Updated version	Access the most recent version of this article at: http://mct.aacrjournals.org/content/5/3/767
Supplementary Material	Access the most recent supplemental material at: http://mct.aacrjournals.org/content/suppl/2007/04/20/5.3.767.DC1

Cited articles	This article cites 39 articles, 20 of which you can access for free at: http://mct.aacrjournals.org/content/5/3/767.full#ref-list-1

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.
Permissions	To request permission to re-use all or part of this article, use this link http://mct.aacrjournals.org/content/5/3/767. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.