Predicting response to docetaxel neoadjuvant chemotherapy for advanced breast cancers through genome-wide gene expression profiling

HITOSHI ZEMBUTSU[1], YASUYO SUZUKI[2], AYA SASAKI[3], TATSUHIKO TSUNODA[4], MINORU OKAZAKI[5], MASATAKA YOSHIMOTO[6], TADASHI HASEGAWA[3], KOICHI HIRATA[2] and YUSUKE NAKAMURA[1]

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639; [2]First Department of Surgery, Sapporo Medical University, School of Medicine, Sapporo 060-8543; [3]Department of Surgical Pathology, Sapporo Medical University, School of Medicine, Sapporo 060-8543; [4]Laboratory for Medical Informatics, Center for Genomic Medicine, RIKEN, Yokohama 230-0045; [5]Department of Surgery, Sapporo Nyusen Geka Clinic, Sapporo 060-0006; [6]Department of Breast Surgery, The Cancer Institute Hospital of JFCR, Tokyo 135-8550, Japan

Received August 13, 2008; Accepted October 3, 2008

DOI: 10.3892/ijo_00000159

Abstract. Neoadjuvant chemotherapy with docetaxel for advanced breast cancer can improve the radicality for a subset of patients, but some patients suffer from severe adverse drug reactions without any benefit. To establish a method for predicting responses to docetaxel, we analyzed gene expression profiles of biopsy materials from 29 advanced breast cancers using a cDNA microarray consisting of 36,864 genes or ESTs, after enrichment of cancer cell population by laser microbeam microdissection. Analyzing eight PR (partial response) patients and twelve patients with SD (stable disease) or PD (progressive disease) response, we identified dozens of genes that were expressed differently between the 'responder (PR)' and 'non-responder (SD or PD)' groups. We further selected the nine 'predictive' genes showing the most significant differences and established a numerical prediction scoring system that clearly separated the responder group from the non-responder group. This system accurately predicted the drug responses of all of nine additional test cases that were reserved from the original 29 cases. Moreover, we developed a quantitative PCR-based prediction system that could be feasible for routine clinical use. Our results suggest that the sensitivity of an advanced breast cancer to the neoadjuvant chemotherapy with docetaxel could be predicted by expression patterns in this set of genes.

Correspondence to: Dr Yusuke Nakamura, Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
E-mail: yusuke@ims.u-tokyo.ac.jp

Key words: docetaxel, breast cancer, neoadjuvant chemotherapy, prediction, gene expression profile, microarray

Introduction

Neoadjuvant systemic treatment before surgery for advanced breast cancer is one of the most crucial factors in reducing mortality (1,2). However, although the estrogen-receptor status is predictive of response to hormonal treatments, there are no clinically useful markers to predict responses to chemotherapy. Therefore, all patients who are eligible for chemotherapy receive the same treatment, even though *de novo* drug resistance results in failures of the treatment in many cases. Docetaxel is one of the most commonly used anticancer agents in the treatment of breast cancer. It binds to ß-tubulin, one of the major components of microtubules. It exerts its growth inhibitory effects by stabilizing microtubules, which arrests the growth of tumor cells at the G2-M phase (3). Docetaxel has been widely used for treatment of metastatic breast cancers (4) and its application for primary breast cancers in the adjuvant and the neoadjuvant settings has also been indicated (5).

However, since no method is yet available to predict responses of individual patients to docetaxel chemotherapy, some patients suffer from adverse reactions of the drug without having any clinical benefit and often lose an opportunity for additional therapy when their physical condition deteriorates (6,7). Hence, development of a method to predict the effectiveness of a specific therapy is of critical importance for patients with advanced breast cancer. Certain factors were reported to be associated with chemosensitivity or prognosis, but application of only one or a few of such factors has failed to reliably predict individual responses and it is obvious that a larger body of information is required toward establishment of a clinically applicable method.

Profiling of gene expression patterns on genome-wide cDNA microarrays enables investigators to perform

comprehensive characterization of molecular activities in cancer cells. Systematic analysis of expression levels for thousands of genes is also a useful approach for identifying molecules related to response to anticancer drugs or radiation. We have been attempting to establish systems based on gene expression profiling that would allow accurate prediction of responses to chemotherapeutic agents in diseases such as acute lymphoblastic leukemia and bladder cancer (8,9). In the study reported here, we established a system for predicting response to docetaxel neoadjuvant chemotherapy among patients with advanced breast cancer, using genome-wide information obtained for 29 cases on a cDNA microarray consisting of 36,864 transcribed elements in combination with laser microbeam microdissection of the tumors to obtain pure populations of cancer cells for analysis. We identified nine genes that showed significantly different levels of expression between the responder and non-responder groups of breast cancer patients who were treated with a neoadjuvant docetaxel monotherapy. We suggest that such information may lead ultimately to our goal of 'personalized treatment'.

Materials and methods

Patients, tissue samples and neoadjuvant chemotherapy. Breast cancer tissue samples from core needle biopsy or surgical biopsy and corresponding clinical information were obtained from four hospitals (Sapporo Social Insurance General Hospital, The Cancer Institute Hospital of JFCR, Sapporo Breast Surgery Clinic and Sapporo Medical University) after each patient had provided informed consent. A total of 29 cancer samples that had been confirmed histologically as invasive breast cancer were selected for this study. Clinical stage of each patient was judged according to the UICC tumor-node-metastasis (TNM) classification. Participants were required to have no serious abnormality in renal, hepatic, or hematological function, with Eastern Cooperative Oncology Group performance status (PS) judged to be ≤2. A piece of cancer tissue had been taken from each patient at the time of biopsy before neoadjuvant chemotherapy. These samples were immediately embedded in TissueTek OCT compound (Sakura, Tokyo, Japan), frozen and stored at -80˚C. The frozen tissues were sliced into 8 μm sections using a cryostat (Sakura) and then stained with H&E for histological examination. Breast cancer cells were selectively enriched for our experiments using the EZ-cut system with a pulsed UV narrow beam focus laser (SL Microtest GmbH, Germany) according to the manufacturer's protocols. Patients were given at least two 28-day cycles of docetaxel neoadjuvant chemotherapy as follows: docetaxel (35 mg per square meter of body surface area) on days 1, 8 and 15.

Clinical response to neoadjuvant chemotherapy. The chemotherapeutic response of primary breast tumors was clinically evaluated according to the Response Evaluation Criteria in Solid Tumors (RECIST) as follows: i) complete response (CR), disappearance of all target lesions; ii) partial response (PR), at least a 30% decrease in the sum of the longest diameter (LD) of target lesions, taking as reference the baseline sum LD; iii) stable disease (SD), neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum LD since the treatment started; and iv) progressive disease (PD), at least a 20% increase in the sum of the LD of target lesions, taking as reference the smallest sum LD recorded since the treatment started or the appearance of one or more new lesions. Entire tumor burden was measured by magnetic resonance imaging, computed tomography scan and ultrasound before and within 30 days after two courses of docetaxel treatment. Among the evaluations by the three modalities, the worst response was used to define the clinical response for each individual.

RNA extraction and T7-based RNA amplification. Total RNAs were extracted from each population of microdissected cancer cells, as described previously (10). To guarantee the quality of RNAs, total RNA extracted from the residual tissue of each case were electrophoresed on a denaturing agarose gel and quality was confirmed by the presence of rRNA bands. Extraction of total RNA and T7-based RNA amplification were done, as described previously (11), except that we used RNeasy micro kits (Qiagen, Valencia, CA). After two rounds of RNA amplification, we obtained 50 to 550 μg of amplified RNA for each sample. A mixture of ductal cells in breast tissues from seven patients was prepared as a universal control and was amplified in the same way. RNA amplified by this method accurately reflects the proportions in the original RNA source, as we had confirmed earlier by semiquantitative reverse transcription-PCR (RT-PCR) experiments (10), in which data from the microarrays were consistent with results from RT-PCR regardless of whether total RNAs or amplified RNAs were used as templates.

cDNA microarray. To obtain cDNAs for spotting on the glass slides, we performed RT-PCR amplification for each gene, as described previously (10). The PCR products were spotted on Alice glass slides^{O,R} (GE Healthcare, Amersham Biosciences, Buckinghamshire, UK) with a high-density Microarray Spotter Lucidea (GE Healthcare, Amersham Biosciences); 9,216 genes were spotted in duplicate on a single slide. We prepared four different sets of slides (a total of 36,864 gene spots), on each of which the same 52 house-keeping genes and two negative control genes were spotted as well. The cDNA probes were prepared from amplified RNA using the method described previously (11). For hybridization experiments, 2.5 μg of amplified RNAs from each patient and from the control were reversely transcribed in the presence of Cy5-dCTP and Cy3-dCTP (GE Healthcare BioSciences), respectively. Hybridization, washing and detection of signals were carried out, as described previously (11).

Quantification of signals. We quantified the signal intensities of Cy3 and Cy5 from the 36,864 spots and analyzed the signals by substituting backgrounds, using ArrayVision software (Imaging Research, Inc., St. Catharines, Ontario, Canada). Subsequently, the fluorescence intensities of Cy5 (breast cancer cells) and Cy3 (control) for each target spot were adjusted so that the mean Cy5/Cy3 ratio of the 52 house-

keeping genes became one. Because the data with low signal intensities are less reliable, we defined a cutoff value for the data on each slide as described previously (12) and excluded genes from further analysis when both Cy3 and Cy5 signal-intensities were lower than the cutoff value (12). For genes for which either or both of the signal-intensities were above the cutoff value, we calculated the ratio of Cy5/Cy3 signals as a relative expression ratio using the raw data of each sample. However, if either of the Cy3 or Cy5 signal intensity was lower than the cutoff value, the Cy5/Cy3 ratio might be calculated to an extremely high or low and lead to make a significant false-influence for the further analysis. Hence, when either Cy3 or Cy5 signal intensity was less than the cutoff value, we adjusted the lower one to be half of each cutoff value plus the signal intensities and then calculated the Cy5/Cy3 ratios.

Hierarchical clustering analysis. We used web-available software ('Cluster' and 'TreeView') written by M. Eisen (http://genome-www5.stanford.edu/MicroArray/SMD/restech.html) to create a graphic representation of the microarray data and to create a dendrogram of hierarchical clustering. Before the clustering algorithm was applied, the fluorescence ratio for each spot was first log-transformed and then the data for each sample were median-centered to remove experimental biases.

Extraction of genes for predicting the chemosensitivity. We applied a random permutation test to identify genes that were expressed at a significantly different level between the two groups, that is, tumors with good response and those with poor response to the chemotherapy. Mean (μ) and standard deviation (δ) were calculated from the log transformed relative expression ratios of each gene in responder (r) and non-responder (n) cases. A discrimination score (DS) for each gene was defined as follows:

$$DS = (\mu_r - \mu_n) / (\delta_r + \delta_n)$$

We carried out permutation tests to estimate the ability of individual genes to distinguish the two groups; samples were randomly permutated between the two groups 10,000 times. Because the DS data set of each gene showed a normal distribution, we calculated a P-value for the user-defined grouping (13). For the initial analysis, we applied the expression data for original 20 cases consisting of 8 responders and 12 non-responders. As the next step, we randomly excluded one case from each group and made 12 different combinations consisting of 7 responders and 11 non-responders. We performed additional 12 permutation tests using the 12 combinations of the two groups.

Calculation of prediction score. We calculated prediction scores according to procedures described previously (13). Each gene (gi) votes for either responder or non-responder depending on whether the expression level (x_i) in the sample is closer to the mean expression level of responders or non-responders in reference samples. The magnitude of the vote (V_i) reflects the deviation of the expression level in the sample from the average of the two classes:

$$V_i = | x_i - (\mu_r + \mu_n) / 2 |$$

We summed the votes to obtain total votes for the responders (V_r) and non-responders (V_n) and calculated PS values as follows:

$$PS = [(V_r - V_n) / (V_r + V_n)] \times 100$$

reflecting the margin of victory in the direction of either responder or non-responder. PS values range from -100 to 100; a higher absolute value of PS reflects a stronger prediction.

Evaluation of the classification and leave-one-out method. We calculated the classification score (CS) using prediction scores of responders (PSr) and non-responders (PSn) in each gene set, as follows:

$$CS = [\mu(PS_r) - \mu(PS_n)] / [\delta(PS_r) + \delta(PS_n)]$$

A larger value of CS indicates better separation of the two groups by the predictive-scoring system. For the leave-one-out test, one sample is withheld, the permutation P-value and mean expression levels are calculated using remaining samples and the class of the withheld sample is subsequently evaluated by calculating its prediction score. We repeated this procedure for each of the 20 samples.

Quantitative reverse transcription-PCR. Aliquots of the same aRNA hybridized to the microarray slides from individual samples and from a mixture of pooled mRNA from normal ductal cells of 7 breast tissues were reversely transcribed using oligo$(dT)_{12-18}$ primer and SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA). Quantitative-RT-PCR were carried out using SYBR Premix Ex Taq (Takara, Ohtsu, Japan) on an ABI Prism 7700 Sequence Detection system (Applied Biosystems, Foster, CA), as described previously (14). The sequences of primers used in the real-time PCR are shown in Table II. To normalize the expression of each gene, we selected as internal controls Sin3A-associated protein, 130 kDa (*SAP130*), NADH dehydrogenase (ubiquinone) 1 ß sub-complex, 8, 19 kDa (*NDUFB8*) and Chloride intracellular channel 1 (*CLIC1*) from among the 52 housekeeping genes because they showed the smallest Cy5/Cy3 fluctuations in our microarray data. Because normalization to these three endogenous control genes (*SAP130*, *NDUFB8* and *CLIC1*) led to similar conclusions (data not shown), we subsequently recorded only the data normalized according to levels of *SAP130* expression. For generation of standard curves we used a mixture of mRNAs derived from cancer samples from 20 patients. Quantitative RT-PCR experiments were done in duplicate for all 9 'predictive' genes and relative expression ratios of each sample were calculated. The normalized gene expression values were log-transformed (on a base 2 scale), in a manner similar to the transformation of microarray-based hybridization data.

Results

Identification of genes associated with docetaxel neoadjuvant chemotherapy. We enrolled 29 patients with breast cancer

Table I. Clinicopathological features of patients with breast cancer.

Sample ID	Gender	T	N	M	Clinical stage	Menopause	ER	PgR	Response	Prediction
BCA-1	Female	4	2	1	IV	Postmenopausal	Positive	Positive	Responder (PR)	Learning
BCA-2	Female	4	1	0	IIIB	Postmenopausal	Positive	Negative	Responder (PR)	Learning
BCA-3	Female	3	1	1	IV	Postmenopausal	Negative	Negative	Responder (PR)	Learning
BCA-4	Female	4	1	0	IIIB	Postmenopausal	Positive	Negative	Responder (PR)	Learning
BCA-5	Female	4	3	0	IIIC	Postmenopausal	Positive	Positive	Responder (PR)	Learning
BCA-6	Female	4	2	0	IIIB	Postmenopausal	Positive	Positive	Responder (PR)	Learning
BCA-7	Female	4	0	0	IIIB	Postmenopausal	Positive	Positive	Responder (PR)	Learning
BCA-8	Female	2	1	0	IIB	Postmenopausal	Positive	Positive	Responder (PR)	Learning
BCA-9	Female	2	2	1	IV	Premenopausal	Positive	Positive	Nonresponder (SD)	Learning
BCA-10	Female	4	1	0	IIIB	Postmenopausal	Negative	Positive	Nonresponder (PD)	Learning
BCA-11	Female	4	0	0	IIIB	Premenopausal	Positive	Positive	Nonresponder (SD)	Learning
BCA-12	Female	1	3	1	IV	Postmenopausal	Positive	Positive	Nonresponder (PD)	Learning
BCA-13	Female	2	2	0	IIIA	Premenopausal	Negative	Negative	Nonresponder (PD)	Learning
BCA-14	Female	2	2	1	IV	Postmenopausal	Positive	Positive	Nonresponder (SD)	Learning
BCA-15	Female	3	0	0	IIB	Premenopausal	Negative	Negative	Nonresponder (PD)	Learning
BCA-16	Female	4	2	0	IIIB	Postmenopausal	Negative	Negative	Nonresponder (SD)	Learning
BCA-17	Female	2	1	0	IIB	Postmenopausal	Negative	Negative	Nonresponder (SD)	Learning
BCA-18	Female	4	2	1	IV	Postmenopausal	NE	NE	Nonresponder (PD)	Learning
BCA-19	Female	4	3	1	IV	Postmenopausal	Positive	Positive	Nonresponder (PD)	Learning
BCA-20	Female	4	1	0	IIIB	Postmenopausal	Positive	NE	Nonresponder (PD)	Learning
BCA-21	Female	2	3	0	IIIC	Postmenopausal	Positive	Negative	Responder (PR)	Test
BCA-22	Female	4	1	1	IV	Postmenopausal	Positive	Positive	Responder (PR)	Test
BCA-23	Female	4	2	0	IIIB	Premenopausal	Positive	Negative	Nonresponder (PD)	Test
BCA-24	Female	4	1	0	IIIB	Postmenopausal	Negative	Negative	Nonresponder (SD)	Test
BCA-25	Female	3	1	0	IIIA	Postmenopausal	Positive	Positive	Nonresponder (SD)	Test
BCA-26	Female	3	0	0	IIB	Postmenopausal	Negative	Negative	Nonresponder (PD)	Test
BCA-27	Female	4	0	0	IIIB	Postmenopausal	Positive	Positive	Nonresponder (SD)	Test
BCA-28	Female	1	2	1	IV	Postmenopausal	Negative	Negative	Nonresponder (SD)	Test
BCA-29	Female	4	3	1	IV	Postmenopausal	Positive	Positive	Nonresponder (SD)	Test

ER, estrogen receptor; PgR, progesterone receptor; response, response to neoadjuvant docetaxel treatment; responder, patient who achieved partial response (PR) after two courses of treatment; PR, at least a 30% decrease in the sum of the longest diameter (LD) of target lesions, taking as reference the baseline sum LD; nonresponder, patient who could not achieve complete response (CR) nor PR after two courses of treatment; SD (stable disease), neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum LD since the treatment started; PD (progressive disease), at least a 20% increase in the sum of the LD of target lesions, taking as reference the smallest sum LD recorded since the treatment started or the appearance of one or more new lesions; learning, samples used to develop the prediction system and test, samples used for test cases.

whose clinicopathological features are summarized in Table I. According to their responses to the treatment, we categorized the patients into two groups: 'responders', patients who achieved partial response (PR) after two courses of docetaxel neoadjuvant chemotherapy and 'non-responders', who revealed no shrinking of the tumors (SD or PD) after the chemotherapy.

To extract genes that were differentially expressed between the two groups, we first analyzed initial 20 samples (8 responders and 12 non-responders) by comparing expression levels of 36,864 transcripts. We carried out a random permutation test to identify genes that showed significantly different expression levels between the two groups and identified dozens of genes whose permutation P-values were $<1x10^{-4}$ and whose signal intensities were higher than the cutoff level in >60% of samples in at least one group (%presence >60). To further select the genes that more effectively distinguish the two groups, we performed 12 additional random permutation tests with twelve different combinations of 7 responders and 11 non-responders that were selected randomly, because the number of samples for this analysis was limited. As shown in Table III, *LOC286109* and *CENPT* were selected by 11 tests; of course, the genes with smaller P-values in the primary random permutation test had a tendency to be selected frequently through these 12 additional tests. The additional random permutation tests further defined a set of 51 genes, which were commonly selected by more than four additional permutation tests and isolated by the same criteria as described above ($P<1x10^{-4}$, %presence >60; Table III). A supervised hierarchical clustering analysis using this set of genes with Cluster and

Table II. List of primer sets.

Accession no.	Symbol	Forward primer	Reverse primer
Internal controls			
NM_024545.2	*SAP130*	GATGCATCAGTGTCCACCAG	GCCTGCAGGAATCCACTAGA
U93205.1	*CLIC1*	ACCATGGCTGAAGAACAACC	CCCTTGAGCCACAGTACCAT
AI096694.1	*NDUFB8*	TATGCAGCTTTTCGGTTTCC	CTGGTTCTTTGGAGGGATCA
Predictive genes			
AK092172.1	*LOC286109*	GCCCTTTAGCTTCGTGTCTG	TGCAACGAGTCTGCTTAGGA
NM_025082.3	*CENPT*	GCCCTTACACGATGGAGTTG	TGTGTCCCTCAGCCTCTTCT
AK026524.1	*POLR3H*	CCACCACTTCCAGTGAGGAG	CTGGTCCACCAGGAGAGAAG
AK098202.1	*HDAC5*	TCACGAGGTCAAGAGTTGGA	GGAGTAGTGCAGTGGTGCAA
BC030535.1	*KIAA1430*	ATCTGATTCGTCTCCGTCATC	TGAAGGCTGTGTTTCTGTCG
NM_000773.3	*CYP2E1*	CCTACATGGATGCTGTGGTG	TGGGGATGAGGTATCCTCTG
NM_181714.1	*LCA5*	GCATTGGTCTTTTGGCTGTT	CTCCTTGGGCTGGTGGTAT
NM_001009184.1	*GRINA*	CGCTGAACCTGTACACAGACATC	GGCTCATGCAACAATGAGTAGAC
AA885242.1	*KIFC2*	CTCCTTTCCAGACAGATGAGAGA	ATGCCTGTTTTTCCTACACTCAG

Treeview software (http://rana.lbl.gov/EisenSoftware.htm) yielded good separation of the two groups with regard to the response to docetaxel (Fig. 1).

Establishment of predictive scoring system for clinical response to docetaxel neoadjuvant chemotherapy. Using the 51-gene set that seemed to effectively distinguish the two groups, we calculated the prediction score of each sample by the weighted-vote method (13). Then we rank-ordered these candidates on the basis of the number of selected times by twelve permutation tests (Table III) and calculated prediction scores by the leave-one-out test for cross-validation. For the leave-one-out test, we withheld one sample and calculated the permutation P-values and mean expression levels using the remaining samples to identify genes that were the most powerful for separating the two groups.

We calculated the classification score (CS) using the prediction scores of eight responders and twelve non-responders in each gene set and obtained the best separation of the two groups by using the 9 genes that were ranked highest in our candidate gene list (Fig. 2A and Table III). A hierarchical clustering analysis using this set of genes with Cluster and Treeview software (http://rana.lbl.gov/Eisen Software.htm) yielded good separation of the two groups with regard to sensitivity to docetaxel neoadjuvant chemotherapy (Fig. 2B).

Finally, to verify the prediction scoring system based on expression data for this set of 9 genes, we examined 9 'test' cases (2 responders and 7 non-responders; Fig. 2A). We investigated gene expression profiles in each of the nine test cases and then calculated a prediction score for each sample. As shown in Fig. 2A, all of nine test-cases (two responders and seven non-responders) were correctly predicted their prediction scores that reflected their clinical response. Our data suggest that expression levels of these nine genes or a part of them might play important roles in cellular responses induced by the docetaxel neoadjuvant chemotherapy.

Establishment of a quantitative reverse transcription-PCR-based predictive scoring system. To further validate the results of cDNA microarray analysis, we carried out real-time quantitative RT-PCR for the 9 predictive genes and three quantitative control genes, *SAP130*, *NDUFB8* and *CLIC1*, using the 29 cases (20 learning and 9 test cases). We observed significant concordance between the results from the cDNA microarray and those of the quantitative RT-PCR experiments. As shown in Table IV, Pearson and Spearman rank correlations were positive and statistically significant for all of them.

Hence, we attempted to adapt our prediction system on the basis of quantitative RT-PCR for an easy clinical test. We performed quantitative real-time RT-PCR of the 9 predictive genes for 20 learning and 9 test cases (10 in responders and 19 in non-responders) and calculated the prediction score for each case. When we estimated these scores by the leave-one-out cross validation test, all cases were placed correctly according to their response to docetaxel neoadjuvant chemotherapy (Fig. 3).

Discussion

cDNA microarrays are now widely used to analyze expression of thousands of genes simultaneously in cancer tissues. However, in our view, adequate attention has not been paid to the quality of the materials and experiments. For example, clinical samples (surgically resected tissue or biopsy materials) usually consist of various cellular components and the proportions of cancer cells in a given tissue can vary enormously from one tumor to another. Hence, most microarray data published previously are likely to be influenced significantly by heterogeneity of cell components in tumor tissues (15). To obtain precise expression data of cancer cells, we applied a laser microbeam microdissection system to purify as much as possible the populations of cancer cells from biopsy specimens of 29 advanced breast cancers, with a view to establish a scoring system to predict response to docetaxel treatment.

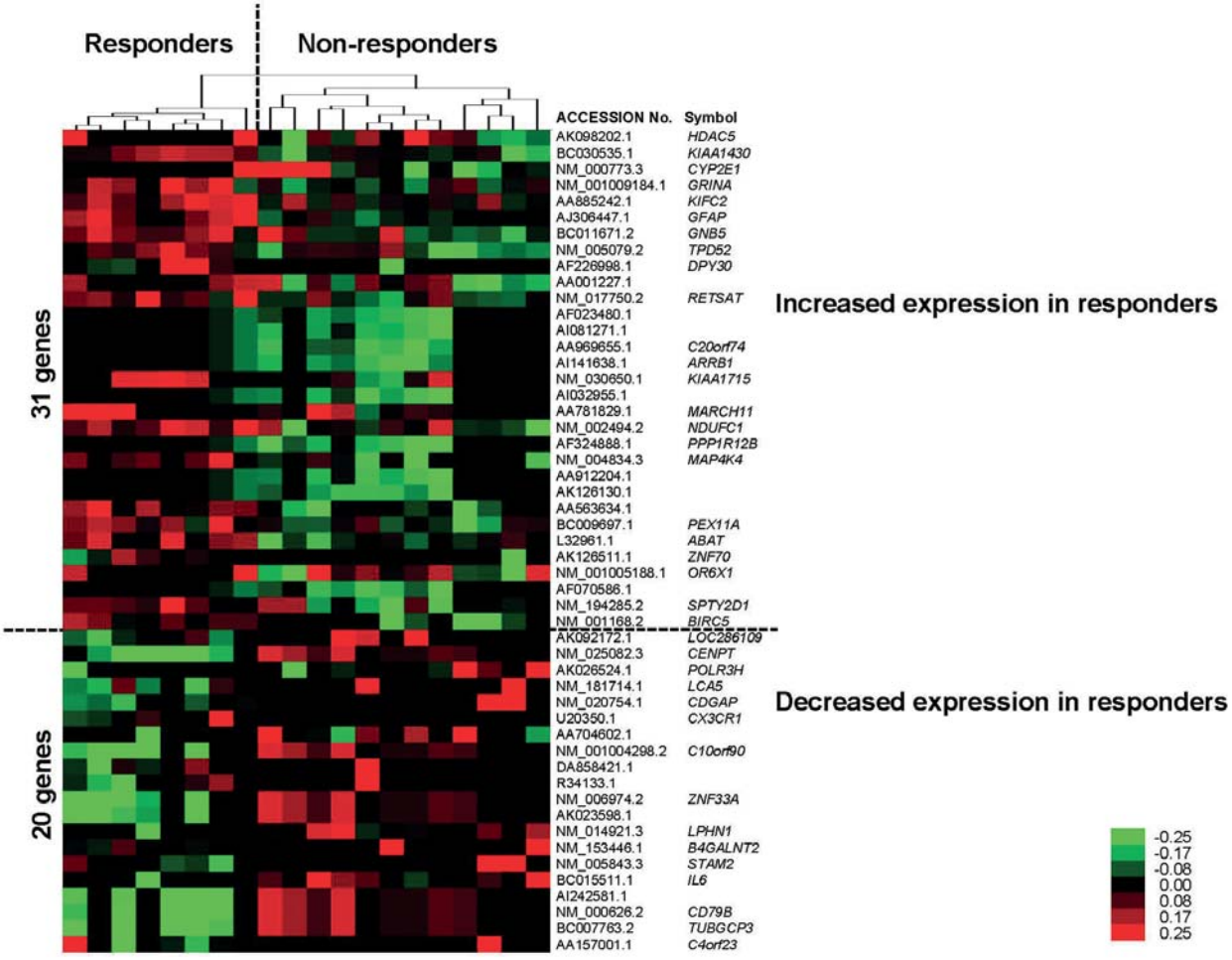

Figure 1. Expression patterns of the 51 genes that discriminated eight responders from twelve non-responders of breast cancer patients treated with docetaxel. Horizontal rows represent individual genes; vertical columns represent individual samples. Each cell in the matrix represents the expression level of a single transcript in single sample, with red and green indicating transcript levels, respectively, above and below the median for that gene across all samples. Black represents unchanged expression or slight expression (intensities of both Cy3 and Cy5 under the cutoff value). Color saturation is proportional to the magnitude of the difference.

Despite recent advances, 64-77% of patients with advanced breast cancer who received docetaxel chemotherapy showed no or very poor response in terms of staging and a large proportion of them suffer from adverse events such as alopecia, myelosuppression and/or gastrointestinal toxicity (16-18). Although certain factors were reported to be associated with chemosensitivity or prognosis of breast cancer patients (17-19), characterization of tumor natures using only one or a few of these factors has failed thus far to reliably predict individual responses, indicating a need for a more accurate method for predicting responses to anticancer drugs.

This study was designed to develop a prediction system for docetaxel neoadjuvant chemotherapy on the basis of gene expression profiles of purified populations of breast cancer cells. We identified 51 genes whose expression was significantly different between the responders and non-responders and further ranked them by the number of selected times in twelve different random permutation tests that reflected a power to discriminate the two groups ($P<0.0001$; Table III). Then we further selected nine genes and established the numerical scoring system. We further tested the scoring system by the leave-one-out cross validation method and found it to provide the best separation of the responders from the non-responders. Furthermore, our scoring system was able to predict accurately the response of all of nine additional test cases to the docetaxel neoadjuvant chemotherapy (Fig. 2A).

The list of 51 genes that showed significant differences between the two groups might provide insight into the biological mechanism(s) underlying sensitivity or resistance to docetaxel chemotherapy. Among those 51 genes, we found that cytochrome P450, family 2, subfamily E, polypeptide 1 (*CYP2E1*) was up-regulated in the responder group (Fig. 1, Table III). *CYP2E1* has been identified as a source of reactive oxygen species in *CYP2E1*-dependent monooxygenation reactions. Moreover, *CYP2E1* is the most active isozyme in the initiation of NADPH-dependent lipid peroxidation (20). Oxidation of docetaxel by *CYP2E1* was thought to influence the effects of docetaxel (21). CYP2E1 up-regulation might increase the cytotoxicity of docetaxel, possibly due to increased production of lipid peroxides and oxygen radicals (22). Thus, the level of *CYP2E1* expression in a tumor could influence the extent of oxidation to reactive metabolites and in turn modify cytotoxic effects. Hence, up-regulation of *CYP2E1* might contribute to response to docetaxel neoadjuvant chemotherapy among patients with breast cancer.

Table III. List of 51 discriminating genes.

Accession no.	Symbol	Gene name	P-value[a]	Group[b]	No. of selected times[c]
AK092172.1	*LOC286109*	Hypothetical protein LOC286109	$3.35x10^{-8}$	-	11
NM_025082.3	*CENPT*	Centromere protein T	$1.57x10^{-5}$	-	11
AK026524.1	*POLR3H*	Polymerase (RNA) III (DNA directed) polypeptide H (22.9 kDa)	$1.39x10^{-7}$	-	10
AK098202.1	*HDAC5*	Histone deacetylase 5	$7.04x10^{-7}$	+	10
BC030535.1	*KIAA1430*	KIAA1430	$6.13x10^{-6}$	+	10
NM_000773.3	*CYP2E1*	Cytochrome P450, family 2, subfamily E, polypeptide 1	$6.34x10^{-5}$	+	10
NM_181714.1	*LCA5*	Leber congenital amaurosis 5	$4.07x10^{-6}$	-	9
NM_001009184.1	*GRINA*	Glutamate receptor, ionotropic, N-methyl D-asparate-associated protein 1 (glutamate binding)	$1.12x10^{-5}$	+	9
AA885242.1	*KIFC2*	Kinesin family member C2	$5.06x10^{-5}$	+	9
NM_020754.1	*CDGAP*	Cdc42 GTPase-activating protein	$2.63x10^{-5}$	-	8
AJ306447.1	*GFAP*	Glial fibrillary acidic protein	$2.72x10^{-5}$	+	8
BC011671.2	*GNB5*	Guanine nucleotide binding protein (G protein), ß 5	$2.95x10^{-5}$	+	7
NM_005079.2	*TPD52*	Tumor protein D52	$8.26x10^{-5}$	+	7
AF226998.1	*DPY30*	Dpy-30 homolog (C. elegans)	$8.36x10^{-5}$	+	7
U20350.1	*CX3CR1*	Chemokine (C-X3-C motif) receptor 1	$3.19x10^{-6}$	-	6
AA001227.1		EST	$3.29x10^{-6}$	+	6
NM_017750.2	*RETSAT*	Retinol saturase (all-trans-retinol 13,14-reductase)	$9.21x10^{-5}$	+	6
AA704602.1		EST	$1.23x10^{-8}$	-	5
NM_001004298.2	*C10orf90*	Chromosome 10 open reading frame 90	$2.61x10^{-7}$	-	5
DA858421.1		Transcribed locus	$9.07x10^{-7}$	-	5
R34133.1		Transcribed locus	$1.46x10^{-6}$	-	5
AF023480.1		Transcribed locus	$2.26x10^{-6}$	+	5
AI081271.1		Transcribed locus	$2.69x10^{-6}$	+	5
AA969655.1	*C20orf74*	Chromosome 20 open reading frame 74	$3.67x10^{-6}$	+	5
NM_006974.2	*ZNF33A*	Zinc finger protein 33A	$5.28x10^{-6}$	-	5
AI141638.1	*ARRB1*	Arrestin, ß 1	$6.85x10^{-6}$	+	5
NM_030650.1	*KIAA1715*	KIAA1715	$8.00x10^{-6}$	+	5
AI032955.1		MRNA; cDNA DKFZp779G2222 (from clone DKFZp779G2222)	$1.24x10^{-5}$	+	5
AA781829.1	*MARCH11*	Membrane-associated ring finger (C3HC4) 11	$1.29x10^{-5}$	+	5
NM_002494.2	*NDUFC1*	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6 kDa	$1.52x10^{-5}$	+	5
AK023598.1		CDNA: FLJ22522 fis, clone HRC12491	$2.58x10^{-5}$	-	5
AF324888.1	*PPP1R12B*	Protein phosphatase 1, regulatory (inhibitor) subunit 12B	$4.32x10^{-5}$	+	5
NM_004834.3	*MAP4K4*	Mitogen-activated protein kinase kinase kinase kinase 4	$4.45x10^{-5}$	+	5
AA912204.1		EST	$8.12x10^{-5}$	+	5
NM_014921.3	*LPHN1*	Latrophilin 1	$5.77x10^{-21}$	-	4
NM_153446.1	*B4GALNT2*	ß-1,4-N-acetyl-galactosaminyl transferase 2	$1.22x10^{-10}$	-	4
NM_005843.3	*STAM2*	Signal transducing adaptor molecule (SH3 domain and ITAM motif) 2	$2.33x10^{-8}$	-	4
BC015511.1	*IL6*	Interleukin 6 (interferon, ß 2)	$1.17x10^{-7}$	-	4
AI242581.1		Transcribed locus	$1.24x10^{-7}$	-	4
NM_000626.2	*CD79B*	CD79b molecule, immunoglobulin-associated ß	$3.02x10^{-7}$	-	4
BC007763.2	*TUBGCP3*	Tubulin, γ complex associated protein 3	$6.09x10^{-7}$	-	4
AK126130.1		CDNA FLJ44142 fis, clone THYMU2016523	$1.52x10^{-6}$	+	4
AA563634.1		Transcribed locus	$1.77x10^{-6}$	+	4
BC009697.1	*PEX11A*	Peroxisomal biogenesis factor 11A	$2.39x10^{-6}$	+	4
L32961.1	*ABAT*	4-aminobutyrate aminotransferase	$2.93x10^{-6}$	+	4

Table III. Continued.

Accession no.	Symbol	Gene name	P-value[a]	Group[b]	No. of selected times[c]
AK126511.1	*ZNF70*	Zinc finger protein 70	$3.52x10^{-6}$	+	4
NM_001005188.1	*OR6X1*	Olfactory receptor, family 6, subfamily X, member 1	$7.56x10^{-6}$	+	4
AF070586.1		Clone 24528 mRNA sequence	$1.51x10^{-5}$	+	4
AA157001.1	*C4orf23*	Chromosome 4 open reading frame 23	$3.09x10^{-5}$	-	4
NM_194285.2	*SPTY2D1*	SPT2, Suppressor of Ty, domain containing 1 (*S. cerevisiae*)	$6.37x10^{-5}$	+	4
NM_001168.2	*BIRC5*	Baculoviral IAP repeat-containing 5 (survivin)	$7.53x10^{-5}$	+	4

[a]P-values were calculated by random permutation tests with responders (n=8) and non-responders (n=12). [b]Plus indicate the genes up-regulated in responders and minus indicate the genes up-regulated in non-responders. [c]Number of selected times as significantly discriminating genes in twelve different random permutation tests. Information was retrieved from Unigene database in National Center for Biotechnology information.

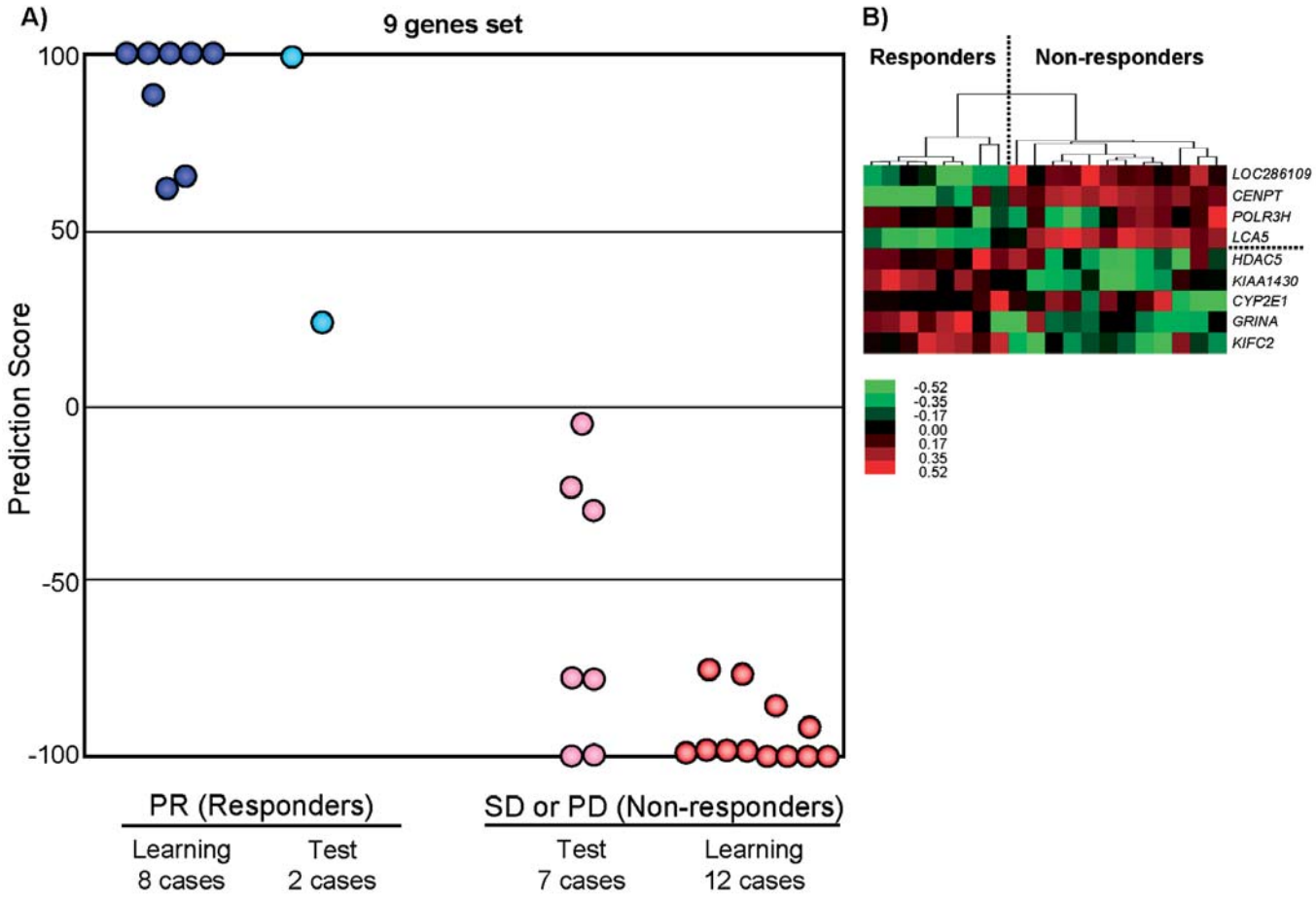

Figure 2. Distribution of prediction scores for 29 patients. (A) Blue and red circles represent scores for cross-validation cases of patients whose expression data were used to select discriminating genes (learning cases). Light blue and light red circles represent scores for nine additional cases (test cases). (B) Clustering analysis using nine predictive genes. All samples are classified into one of two 'trees' according to their responces to docetaxel neoadjuvant chemotherapy.

On the other hand, in our experiments genes encoding chemokine (C-X3-C motif) receptor 1 (*CX3CR1*), Interleukin 6 (interferon, ß 2) (*IL6*) and Signal transducing adaptor molecule (SH3 domain and ITAM motif) 2 (STAM2) were up-regulated in non-responders (Fig. 1, Table III). CX3CR1 mediates the PI3-kinase/Akt pathway and activates ERK1 and ERK2 that are involved in cell survival and promote the anti-apoptosis activity (23-25). It has been known that activation the PI3-

Table IV. Correlation of microarray expression data with quantitative-PCR-derived values.

Accession no.	Symbol	Pearson correlation coefficient	P-value	Spearman rank correlation	P-value
AK092172.1	*LOC286109*	0.90	0.00079	0.93	0.00024
NM_025082.3	*CENPT*	0.79	0.045	0.77	0.07
AK026524.1	*POLR3H*	0.93	0.00031	0.97	0.00031
AK098202.1	*HDAC5*	0.79	0.000043	0.82	0.001
BC030535.1	*KIAA1430*	0.68	0.00099	0.65	0.0021
NM_000773.3	*CYP2E1*	0.83	0.00000037	0.78	0.000007
NM_181714.1	*LCA5*	0.74	0.022	0.72	0.03
NM_001009184.1	*GRINA*	0.83	0.0029	0.9	0.00034
AA885242.1	*KIFC2*	0.66	0.0021	0.59	0.008

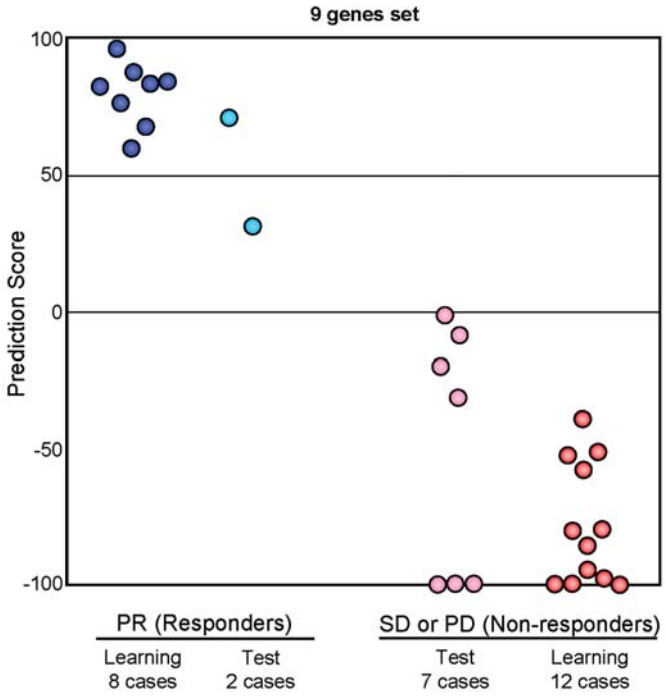

Figure 3. Quantitative reverse transcription-PCR-based prediction scoring system. Prediction scores for 29 cases using values derived from quantitative RT-PCR experiments of nine predictive genes. Blue and red circles indicate scores for selecting discriminating genes (learning cases). Light blue and light red circles represent scores of nine additional (test) cases.

kinase/Akt pathway increases resistance to chemotherapy in cancer cells (26-28). Hence, up-regulation of CX3CR1 might influence resistance to docetaxel neoadjuvant chemotherapy through activation of this cell-survival pathway. IL-6 is considered as a positive growth factor in cancer cells (29). It was reported that autocrine production of IL-6 in breast cancer cells could confer multi-drug resistance *in vitro* through induction of multidrug resistance gene-1 transcription with subsequent overexpression of P-glycoprotein (PGP) (30). Hence, up-regulation of IL-6 might influence resistance to docetaxel neoadjuvant chemotherapy. STAM2 has a unique structure containing a Src homology 3 (SH3) domain and a tyrosine cluster region including an immunoreceptor tyrosine-based activation motif (31). Furthermore, STAM2, which is associated with Jak2 and Jak3 tyrosine kinases, is implicated in playing a key role in the cell survival (32). Thus, up-regulated expression of STAM2 might contribute to resistance to docetaxel neoadjuvant chemotherapy through Jak/STAM/c-myc pathway (33).

Previously, other groups have predicted prognosis or chemosensitivity of tumors based on quantitative RT-PCR results for expression of genes selected through microarray analysis (34,35). To confirm the reliability of microarray data and establish more convenient prediction strategies for routine clinical use, we also performed quantitative RT-PCR experiments for 29 cases (20 learning cases and 9 test cases) of breast cancer using the nine selected predictive genes. We confirmed significant correlation between the data obtained by the microarray and those by the quantitative RT-PCR (Table IV). Moreover, we verified that our quantitative RT-PCR-based prediction system could also correctly classify all of our subsequent test cases with regard to their drug responses (Fig. 3). In any case, it should also be noted that the functions of 4 of our 9 predictive genes are still not well known. Therefore, further investigations will be needed to clarify their biological mechanisms associated with the response to docetaxel treatment.

In conclusion, we imply with some confidence that our prediction system for sensitivity of breast cancers to docetaxel therapy, on the basis of either the microarray-derived expression profiles or the quantitative RT-PCR results, should provide opportunities for achieving better prognosis and improved quality of life for patients, although a larger scale study is certainly warranted. Our data suggest that the goal of 'personalized medicine', prescribing the correct treatment regimen for each patient, may be achievable by selecting specific sets of genes for their predictive values according to the approach shown here.

Acknowledgements

We thank Risa Goswammi for cDNA microarray experiments; Noriko Sudo and Miwako Ando for fabricating the cDNA microarray; Fuyuki Miya and Takashi Morizono for analysis of data. We also thank all surgeons (Sapporo Social Insurance General Hospital, The Cancer Institute Hospital of JFCR, Sapporo Breast Surgery Clinic and Sapporo Medical University) for providing patient samples.

References

1. Kim R, Osaki A, Tanabe K and Toge T: Neoadjuvant chemotherapy for local advanced breast cancer with stage IIIB. Oncol Rep 11: 1265-1272, 2004.
2. Smith IC, Heys SD, Hutcheon AW, *et al*: Neoadjuvant chemotherapy in breast cancer: significantly enhanced response with docetaxel. J Clin Oncol 20: 1456-1466, 2002.
3. Ringel I and Horwitz SB: Studies with RP 56976 (taxotere): a semisynthetic analogue of taxol. J Natl Cancer Inst 83: 288-291, 1991.
4. Chan S, Friedrichs K, Noel D, *et al*: Prospective randomized trial of docetaxel versus doxorubicin in patients with metastatic breast cancer. J Clin Oncol 17: 2341-2354, 1999.
5. Tham YL, Gomez LF, Mohsin S, *et al*: Clinical response to neoadjuvant docetaxel predicts improved outcome in patients with large locally advanced breast cancers. Breast Cancer Res Treat 94: 279-284, 2005.
6. Joensuu H, Kellokumpu-Lehtinen PL, Bono P, *et al*: Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. N Engl J Med 354: 809-820, 2006.
7. Posner MR, Hershock DM, Blajman CR, *et al*: Cisplatin and fluorouracil alone or with docetaxel in head and neck cancer. N Engl J Med 357: 1705-1715, 2007.
8. Zembutsu H, Yanada M, Hishida A, *et al*: Prediction of risk of disease recurrence by genome-wide cDNA microarray analysis in patients with Philadelphia chromosome-positive acute lymphoblastic leukemia treated with imatinib-combined chemotherapy. Int J Oncol 31: 313-322, 2007.
9. Takata R, Katagiri T, Kanehira M, *et al*: Validation study of the prediction system for clinical response of M-VAC neoadjuvant chemotherapy. Cancer Sci 98: 113-117, 2007.
10. Kitahara O, Furukawa Y, Tanaka T, *et al*: Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Res 61: 3544-3549, 2001.
11. Okabe H, Satoh S, Kato T, *et al*: Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. Cancer Res 61: 2129-2137, 2001.
12. Ono K, Tanaka T, Tsunoda T, *et al*: Identification by cDNA microarray of genes involved in ovarian carcinogenesis. Cancer Res 60: 5007-5011, 2000.
13. Golub TR, Slonim DK, Tamayo P, *et al*: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537, 1999.
14. Yamanaka Y, Tamari M, Nakahata T and Nakamura Y: Gene expression profiles of human small airway epithelial cells treated with low doses of 14- and 16-membered macrolides. Biochem Biophys Res Commun 287: 198-203, 2001.
15. Iwao-Koizumi K, Matoba R, Ueno N, *et al*: Prediction of docetaxel response in human breast cancer by gene expression profiling. J Clin Oncol 23: 422-431, 2005.
16. ten Bokkel Huinink WW, Prove AM, Piccart M, *et al*: A phase II trial with docetaxel (Taxotere) in second line treatment with chemotherapy for advanced breast cancer. A study of the EORTC Early Clinical Trials Group. Ann Oncol 5: 527-532, 1994.
17. Alexandre J, Bleuzen P, Bonneterre J, *et al*: Factors predicting for efficacy and safety of docetaxel in a compassionate-use cohort of 825 heavily pretreated advanced breast cancer patients. J Clin Oncol 18: 562-573, 2000.
18. Egawa C, Miyoshi Y, Takamura Y, Taguchi T, Tamaki Y and Noguchi S: Decreased expression of BRCA2 mRNA predicts favorable response to docetaxel in breast cancer. Int J Cancer 95: 255-259, 2001.
19. Hasegawa S, Miyoshi Y, Egawa C, *et al*: Mutational analysis of the class I beta-tubulin gene in human breast cancer. Int J Cancer 101: 46-51, 2002.
20. Arinc E, Arslan S and Adali O: Differential effects of diabetes on CYP2E1 and CYP2B4 proteins and associated drug metabolizing enzyme activities in rabbit liver. Arch Toxicol 79: 427-433, 2005.
21. Gut I, Danielova V, Holubova J, Soucek P and Kluckova H: Cytotoxicity of cyclophosphamide, paclitaxel, and docetaxel for tumor cell lines in vitro: effects of concentration, time and cytochrome P450-catalyzed metabolism. Arch Toxicol 74: 437-446, 2000.
22. Soucek P, Filipcova B and Gut I: Cytochrome P450 destruction and radical scavenging by benzene and its metabolites. Evidence for the key role of quinones. Biochem Pharmacol 47: 2233-2242, 1994.
23. Boehme SA, Lio FM, Maciejewski-Lenoir D, Bacon KB and Conlon PJ: The chemokine fractalkine inhibits Fas-mediated cell death of brain microglia. J Immunol 165: 397-403, 2000.
24. Sawai H, Park YW, Roberson J, Imai T, Goronzy JJ and Weyand CM: T cell costimulation by fractalkine-expressing synoviocytes in rheumatoid arthritis. Arthritis Rheum 52: 1392-1401, 2005.
25 Sonnet C, Lafuste P, Arnold L, *et al*: Human macrophages rescue myoblasts and myotubes from apoptosis through a set of adhesion molecular systems. J Cell Sci 119: 2497-2507, 2006.
26. Tsurutani J, West KA, Sayyah J, Gills JJ and Dennis PA: Inhibition of the phosphatidylinositol 3-kinase/Akt/mammalian target of rapamycin pathway but not the MEK/ERK pathway attenuates laminin-mediated small cell lung cancer cellular survival and resistance to imatinib mesylate or chemotherapy. Cancer Res 65: 8423-8432, 2005.
27. McCubrey JA, Steelman LS, Chappell WH, *et al*: Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. Biochim Biophys Acta 1773: 1263-1284, 2007.
28. McCubrey JA, Steelman LS, Abrams SL, *et al*: Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance. Adv Enzyme Regul 46: 249-279, 2006.
29. Cavarretta IT, Neuwirt H, Untergasser G, *et al*: The antiapoptotic effect of IL-6 autocrine loop in a cellular model of advanced prostate cancer is mediated by Mcl-1. Oncogene 26: 2822-2832, 2007.
30. Rincon M, Broadwater G, Harris L, *et al*: Interleukin-6, multidrug resistance protein-1 expression and response to paclitaxel in women with metastatic breast cancer: results of cancer and leukemia group B trial 159806. Breast Cancer Res Treat 100: 301-308, 2006.
31. Endo K, Takeshita T, Kasai H, *et al*: STAM2, a new member of the STAM family, binding to the Janus kinases. FEBS Lett 477: 55-61, 2000.
32. Yamada M, Ishii N, Asao H, *et al*: Signal-transducing adaptor molecules STAM1 and STAM2 are required for T-cell development and survival. Mol Cell Biol 22: 8648-8658, 2002.
33. Pandey A, Fernandez MM, Steen H, *et al*: Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signaling pathways. J Biol Chem 275: 38633-38639, 2000.
34. Ma XJ, Wang Z, Ryan PD, *et al*: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 5: 607-616, 2004.
35. Lossos IS, Czerwinski DK, Alizadeh AA, *et al*: Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 350: 1828-1837, 2004.