

Gene Expression Profile Associated with Response to Doxorubicin-Based Therapy in Breast Cancer

Maria Aparecida Azevedo Koike Folgueira,¹ Dirce Maria Carraro,⁴ Helena Brentani,⁴ Diogo Ferreira da Costa Patrão,⁵ Edson Mantovani Barbosa,⁶ Mário Mourão Netto,⁵ José Roberto Fíguro Caldeira,⁸ Maria Lucia Hirata Katayama,¹ Fernando Augusto Soares,⁵ Célia Tosello Oliveira,⁶ Luiz Fernando Lima Reis,⁴ Jane Haruko Lima Kaiano,⁴ Luiz Paulo Camargo,⁴ Ricardo Zorzetto Nicolliello Vêncio,⁷ Igor Moysés Longo Snitcovsky,² Fabiana Baroni Alves Makdissi,⁵ Paulo José da Silva e Silva,³ João Carlos Guedes Sampaio Góes,⁶ and Maria Mitzi Brentani¹

Abstract Purpose: This study was designed to identify genes that could predict response to doxorubicin-based primary chemotherapy in breast cancer patients.

Experimental Design: Biopsy samples were obtained before primary treatment with doxorubicin and cyclophosphamide. RNA was extracted and amplified and gene expression was analyzed using cDNA microarrays.

Results: Response to chemotherapy was evaluated in 51 patients, and based on Response Evaluation Criteria in Solid Tumors guidelines, 42 patients, who presented at least a partial response ($\geq 30\%$ reduction in tumor dimension), were classified as responsive. Gene profile of samples, divided into training set ($n = 38$) and independent validation set ($n = 13$), were at first analyzed against a cDNA microarray platform containing 692 genes. Unsupervised clustering could not separate responders from nonresponders. A classifier was identified comprising *EMILIN1*, *FAM14B*, and *PBEF*, which however could not correctly classify samples included in the validation set. Our next step was to analyze gene profile in a more comprehensive cDNA microarray platform, containing 4,608 open reading frame expressed sequence tags. Seven samples of the initial training set (all responder patients) could not be analyzed. Unsupervised clustering could correctly group all the resistant samples as well as at least 85% of the sensitive samples. Additionally, a classifier, including *PRSS11*, *MTSS1*, and *CLPTM1*, could correctly distinguish 95.4% of the 44 samples analyzed, with only two misclassifications, one sensitive sample and one resistant tumor. The robustness of this classifier is 2.5 greater than the first one.

Conclusion: A trio of genes might potentially distinguish doxorubicin-responsive from non-responsive tumors, but further validation by a larger number of samples is still needed.

Primary chemotherapy in breast cancer is associated to the same survival benefit as adjuvant chemotherapy and offers the

advantage of an increased likelihood of breast conservation (1, 2). Many drug regimens have been used for a varied number of cycles, and response rates from 65% to 100% have been achieved in operable breast cancer; two of the most used, doxorubicin and cyclophosphamide, when given before surgery, are associated with an 80% response rate of breast tumor size (1, 3). Contrariwise, some patients may not experience a tumor reduction with a particular drug regimen, and if identified, they could be offered other active drug regimens or be submitted, at once, to surgical intervention.

Although predictive factors might help selection of the appropriate treatment for each individual patient, to date, there is no single marker with a predictive value for a patient's response to chemotherapy (4). A few studies have been looking for a gene profile that might predict response to primary chemotherapy in breast cancer (5–8). There is therefore much interest in breast cancer transcriptional profiling and its role in tailoring therapy.

This study was undertaken to identify genes that could predict response to doxorubicin-based primary chemotherapy in breast cancer patients.

Authors' Affiliations: ¹Disciplina de Oncologia do Departamento de Radiologia da Faculdade de Medicina da Universidade de São Paulo, ²Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, ³Instituto de Matemática e Estatística da Universidade de São Paulo, ⁴Instituto Ludwig de Pesquisa sobre o Câncer, ⁵Hospital do Câncer A.C. Camargo, ⁶Instituto Brasileiro de Controle do Câncer, ⁷Instituto Israelita de Ensino e Pesquisa Albert Einstein, São Paulo, and ⁸Hospital Amaral Carvalho, Jaú, São Paulo, Brazil

Received 3/18/05; revised 7/13/05; accepted 7/21/05.

Grant support: Fundação de Amparo à Pesquisa do Estado de São Paulo grant 01/001468-8 and Conselho Nacional de Desenvolvimento Científico e Tecnológico.

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

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Requests for reprints: Maria Mitzi Brentani, Disciplina de Oncologia do Departamento de Radiologia da Faculdade de Medicina da Universidade de São Paulo, Av. Dr. Arnaldo #455, São Paulo, São Paulo, Brazil. Phone: 55-11-3082-6580; Fax: 55-11-3082-6580; E-mail: mbrentani@lim24.fm.usp.br.

©2005 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-04-0548

Patients and Methods

Patients. Patients with newly histopathologically confirmed invasive breast cancer in samples obtained by core or incisional biopsy, clinical stages II or III, were invited to participate in this study of gene profile associated to response to primary chemotherapy based in doxorubicin, in a routine treatment basis protocol, if they followed the following eligibility criteria: ages 21 to 70 years, Eastern Cooperative Oncology Group performance status ≤ 1 , and adequate hematologic, renal, and hepatic functions. Cardiac disease was an excluding criterion.

Seventy-nine patients were prospectively accrued in three reference centers for cancer treatment in São Paulo, Brazil (Instituto Brasileiro de Controle do Câncer, São Paulo; Hospital do Câncer A.C. Camargo, São Paulo; and Hospital Amaral Carvalho, Jaú) from January 2002 to March 2005. This study was approved by the Brazilian National Ethics Committee (Comitê Nacional de Ética em Pesquisa) and a written informed consent was signed by all participants.

Nine samples were excluded, as clinical response to chemotherapy could not be evaluated due to the last clinical evaluation being done before the fourth chemotherapy cycle ($n = 3$); inflammatory carcinoma diagnosed on histopathologic exam of the breast sample ($n = 5$); and previous chemotherapy for contralateral breast cancer ($n = 1$). Other 19 tumor specimens could not be tested due to the following reasons: prechemotherapy sample was not collected ($n = 5$); invasive cancer was not the predominant feature on histologic analysis ($n = 6$); and poor RNA quality on extraction or impaired RNA amplification ($n = 8$).

Fifty-one samples, collected during tumor biopsy, were available for gene expression analysis, and 38 were primarily included in a training set and the other 13 were later analyzed as an independent validation set. Analysis was done on two different cDNA microarray slides, one containing 692 and the other containing 4,608 spotted sequences. Seven samples of the training set could not be analyzed on both slides, as no more material was available for the second hybridization.

Median age of the 51 patients, who had their samples analyzed, was 48 years (31-67 years), 54.2% were premenopausal, and infiltrating ductal carcinoma was diagnosed in 82% of the patients (Table 1). The great majority presented large lesions with a mean diameter of primary breast tumor and axillary lymph nodes of 88.0 mm. Axillary lymph nodes were detected by palpation in 80.4% of the patients, and supraclavicular node was identified in two patients before treatment. Patients with clinically negative axillae included in this study had primary tumors of at least 30 mm on physical examination. Six patients were clinically staged as IIa, 2 as IIb, 23 as IIIa, 18 as IIIb, and 2 as IV (palpable supraclavicular lymph node; American Joint Committee on Cancer, 1997). Two patients had their contralateral breast previously operated on, one due to a benign phyllodes tumor and the other due to a ductal carcinoma *in situ*, 23 and 5 years, respectively, before invasive breast cancer was diagnosed.

Therapy. All patients were planned to receive doxorubicin and cyclophosphamide therapy every 21 days at 60 and 600 mg/m², respectively, for four cycles. Patients received all four courses of chemotherapy, except for two patients who had their breasts operated on for stable disease after the third course. Median duration of chemotherapy was 68 days, mean administered dose of doxorubicin was 96.7%, and only the two previously reported patients received <90% of the planned dose.

Assessment of clinical response. The longest diameter of palpable breast and lymph node lesions, identified as target lesions (all measurable lesions up to a maximum of 5 lesions per organ and 10 lesions in total, representative of all involved organs) according to the Response Evaluation Criteria in Solid Tumors guidelines (9), was clinically measured before each cycle of primary chemotherapy and after the last one. Disappearance of all lesions was defined as complete response. Patients were classified as responsive if at least a 30% decrease in the sum of the diameters of target lesions, taking as reference the baseline sum of the longest diameter, was detected. Otherwise, they were grouped as nonresponsive. The median interval between the last

cycle of chemotherapy and clinical evaluation was 26.5 days. Surgical therapy followed within a median time of 37 days from the last cycle of chemotherapy; 41 patients underwent mastectomy and 6 underwent breast-conserving surgery. Axillary dissection was done in all patients, and a median of 19 lymph nodes was dissected.

cDNA microarray assembly, hybridization, and analysis. Literature and Serial Analysis of Gene Expression libraries were reviewed to select genes expressed in mammary tissue and breast cancer to assemble a breast tissue-specific cDNA microarray glass slide. Some genes related to chemotherapy resistance as well as some open reading frame expressed sequence tags (ORESTES; ref. 10), identified as expressed in other cancer types as head and neck and stomach, were also selected, some of them corresponding to unknown genes (QT-02 platform). Sequences representing 692 genes were then chosen in the Human Cancer Genome Project bank (Fundação de Amparo à Pesquisa do Estado de São Paulo/Instituto Ludwig de Pesquisa sobre o Câncer) or synthesized by PCR reactions. Inserts were amplified by PCR using M13 reverse and forward primers from the cDNA clones. Amplicons were purified by gel filtration, and clones were printed as three or six replicates onto Corning slides using a Flexys Robot (Genomic Solutions, Ann Arbor, MI). Some genes were represented by two clones corresponding to different regions of the cDNA.

ORESTES representing 4,608 genes, of which the full-length sequence is known, were chosen in the Human Cancer Genome Project bank. All the ORESTES tags were at least 300 bp long and were contained in the 3' end of genes but 5' to the first polyadenylation signal (4.8K-01 platform; ref. 11). Inserts were amplified as described and amplicons were purified by gel filtration. After sequencing to verify identity, clones were printed onto Corning slides using a Flexys Robot. One hundred ninety-two reference sequences were also spotted on the slides. Both cDNA microarray platforms, complying with Minimum Information About a Microarray Experiment format, were submitted to the Gene Expression Omnibus data repository under the accession nos. GPL 1727 and GPL 1930 (<http://www.ncbi.nlm.nih.gov/projects/geo>), respectively. Both cDNA microarray slides were assembled at Instituto Ludwig de Pesquisa sobre o Câncer. Raw data can be achieved at <http://www.lbc.ludwig.org.br/doxorubicin>.

Samples obtained from tumor biopsies were hand dissected to eliminate normal tissue, fibrosis, and adipose tissue, and after microscopic analysis, only samples composed of at least 80% malignant cells were further processed. Total RNA from frozen or RNAlater preserved specimens was isolated using Trizol reagent (Invitrogen Corp., Carlsbad, CA) according to the manufacturer's protocol or CsCl gradient centrifugation. RNA quality was verified by electrophoresis through agarose gel on visualization with ethidium bromide. Only RNA samples with a ratio of >1 for 28S/18S rRNA were further processed. A two-round RNA amplification procedure was carried out by combining antisense RNA amplification with a template-switching effect following a previously described protocol (12) with some minor modifications (13). At the start, total RNA (3 μ g) was used to yield amplified RNA (~60 μ g). Amplified RNA (3-5 μ g) was then used in a reverse transcriptase reaction in the presence of random hexamer primer (Invitrogen/Life Technologies, Carlsbad, CA), Cy3- or Cy5-labeled dCTP (Amersham Biosciences), and SuperScript II (Invitrogen/Life Technologies). HB4A normal epithelial mammary cell line (kindly donated by Drs. Mike O'Hare and Alan Mackay, Ludwig Institute for Cancer Research-University College London, London, United Kingdom; ref. 14) was used as reference for hybridization. These cells were processed in the same manner as tumor samples.

Equal amounts of breast tumor samples and HB4A cDNA labeled probes were concurrently hybridized against cDNA microarray slides. Dye swap was done for each sample analyzed to control for dye bias. Prehybridization was carried out in a humidified chamber at 42°C for 16 to 20 hours and hybridization at 65°C on a GeneTac Hybridization Station (Genomic Solutions).

Hybridized arrays were scanned on a confocal laser scanner (ArrayExpress, Packard Bioscience) using identical photomultiplier

Table 1. Characteristics of the patients included in the study

ID	Age (y)	Clinical stage	Target lesions (mm)	Tumor histology	Clinical response	No. involved nodes	Estrogen receptor status	Progesterone receptor status	ErbB2
J-04	51	IIIb	90.0	D	Y	0	+	+	-
I-02	67	IIIa	82.0	D	Y	0	+	-	+
I-03	58	IIIa	120.0	D	Y	0	+	-	+
I-05	56	IIIa	150.0	D	N	NA	+	-	-
I-06	63	IIIb	85.7	D	Y	0	+	-	-
I-07	39	IIIb	98.0	D	Y	6	+	+	+
I-09	44	IIIb	120.0	D	Y	19	+	+	+
I-10	32	IIIa	390.0	D	Y	NA	-	-	+
I-15	48	IIIb	92.0	D	Y	3	-	-	-
I-16	55	IIIa	116.0	D/L	Y	9	+	+	+
I-18	37	IIIa	148.0	D	Y	2	+	-	+
I-19	60	IIIa	82.0	D	N	5	+	+	+
I-20	48	IIIa	109.0	D	Y	4	+	+	+
I-21	54	IIIa	120.0	M	Y	0	+	+	+
I-22	43	IIIa	102.0	D	N	16	+	+	+
I-23	63	IIIa	112.0	C	Y	14	-	-	-
I-24	56	IIb	80.0	D	Y	3	+	+	-
I-25	46	IIIa	140.0	D	Y	0	-	-	+
I-36	51	IIIa	71.0	D	Y	4	+	+	-
I-34	34	IIIa	115.0	L	Y	16	+	+	-
Q-06 13	64	IIIb	130.0	P	N	2	+	-	+
Q-28	40	IIIb	110.0	D	N	21	+	+	+
Q-32	65	IIa	40.0	D	Y	0	+	-	-
Q-47	46	IIb	45.0	D	Y	0	+	-	-
Q-48	55	IIa	35.0	D	Y	0	+	+	+
Q-27	57	IIIa	45.0	D	Y	2	+	+	-
Q-44	49	IIIb	70.0	D	Y	17	-	+	-
Q-4546	51	IIIb	48.0	D	Y	3	+	+	+
Q-17	64	IIIb	40.0	D	Y	1	+	-	+
Q-100	53	IIa	30.0	D	Y	2	+	+	-
Q-104	67	IIIb	110.0	D	N	17	+	+	-
Q-107	31	IIIa	65.0	D	Y	9	+	+	+
Q-129	33	IV	120.0	D	Y	0	-	-	-
Q-113	56	IIIb	35.0	L	Y	2	+	-	-
Q-115	40	IIIb	45.0	D	Y	5	+	+	+
Q-127	47	IIIa	60.0	D	N	1	-	-	+
Q-130	47	IIa	30.0	D	Y	1	+	+	-
Q-136	55	IIIa	60.0	D	Y	7	+	+	-
J-11 (V)	42	IIIa	70.0	D	Y	0	-	-	-
I-33 (V)	61	IIIa	95.0	D	N	NA	-	-	-
Q-144 (V)	45	IIIa	85.0	A	Y	23	-	-	ND
Q-52 (V)	56	IIa	35.0	D	Y	0	+	+	+
J-01 (V)	39	IIIb	59.0	D	Y	3	+	+	-
I-31 (V)	45	IIIa	138.0	L	Y	5	+	+	-
Q-138 (V)	42	IIIb	50.0	D	Y	6	+	+	+
Q-170 (V)	42	IIa	50.0	D	Y	1	+	+	+
Q-164 (V)	58	IIIb	35.0	D	Y	1	+	+	-
Q-183 (V)	47	IV	100.0	D	Y	NA	+	-	+
J-15 (V)	48	IIIb	60.0	D	Y	NA	-	+	-
I-01 (V)	42	IIIa	110.0	D	N	4	-	-	+
J-03 (V)	47	IIIb	60.0	D	Y	6	+	+	-

Note: Estrogen receptor, progesterone receptor, and ErbB2 were determined by immunohistochemistry: +, positive; -, negative.

Abbreviations: V, validation set; ND, not determined; NA, not available; clinical response: Y, yes; N, no; tumor histology: D, ductal; L, lobular; D/L, mixed ductal and lobular; C, cribriform; P, papillary; M, medullary; A, apocrine.

voltage (PMT 50) for all slides and data were recovered by QuantArray software (Packard Bioscience) using histogram methods. After image acquisition and quantification, saturated spots (signal intensity >63,000) as well as low-intensity spots (QT-02: within the 95% percentile of intensity distribution of known empty spots) were removed from the analysis. Average signal intensity between technical replicates was determined for each spotted sequence. In platform QT-02 (three to six times replica spots representing the same gene), average signal intensity was determined and spots with low reproducibility between technical replicates (mean \pm 2 SDs cutoff) were excluded; then, the average signal was once again evaluated without these spot values. For platform 4.8K, we did a local background subtraction. Quantified signals were then submitted to log transformation and Lowess normalization within each array followed by a global Lowess normalization for all arrays.

Permuted Student's *t* test (10,000 permutations) was used to determine the significance level of the expression of each individual gene, and false discovery ratio (FDR) was employed as a multiple analysis correction. Hierarchical clustering analysis based on Euclidean distance and complete linkage was done using the genes differentially expressed. Reliability of the clustering was assessed by Bootstrap technique using TMEV software (15).

To predict the response to doxorubicin, linear classifiers were then designed, as reported by Kim et al. (16, 17), to have a small error with respect to a spread sample data. To minimize the computational effort to select good feature sets, the preprocessing technique support vector machine (SVM)-based feature selection algorithm was used (18). The feature selection step relies on a modified linear SVM, which uses the maximum distance instead of the usual Euclidean distance. The genes used to define the best separating plane are selected for the next step, and the other genes are fed back to the modified SVM, repeating the procedure described, until a fixed number of genes is selected. Once the preselection phase is completed, an exhaustive search for classifiers based on triplets of genes is done. To assess model performance, leave-one-out cross-validation testing was applied. Finally, to further evaluate the model, validation testing with 13 new samples was done.

To investigate a relationship between differentially expressed genes, Fatigo (<http://www.fatigo.org>; ref. 19) and Onto-Express Tool (<http://vortex.cs.wayne.edu:8080/index.jsp>) were used. Onto-Express Tool searches for biological processes, molecular functions, cellular components, and chromosomes, corresponding to differentially expressed genes. It also gives a significance level for each one of those, based on the total number of genes involved in that function, both in the differentially expressed list and in the whole array. It assesses *P*s values (based on χ^2) and FDR for each.

Quantitative real-time PCR. Reverse transcription was done using 2 μ g total RNA, random hexamer primer, and SuperScript II reverse transcriptase. Primers were designed in different exons to avoid amplification of genomic DNA following sequences deposited at <http://www.ncbi.nlm.nih.gov/nucleotide> using Primer 3 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) and synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). Primer sequences used in this study are provided in Supplementary Table S1.

PCR reactions were done in a LightCycler (Roche Diagnostics GmbH, Mannheim, Germany) or Rotor-Gene System (Corbett Research, Mortlake, Sydney, Australia). Thermocycling was done in a total volume of 20 μ L containing 5 μ L cDNA sample (diluted 1:10); 1.5 to 2.0 mol/L MgCl₂; 0.2 μ mol/L of each primers; 1 μ L LightCycler-DNA Master SYBR Green I (Roche Diagnostics) or 0.1 μ L SYBR Green I (Sigma) working dilution (1:100) and 1.25 units Platinum Taq DNA polymerase (Invitrogen), reaction buffer, and deoxynucleotide triphosphate mixture; 5% DMSO; and 0.5 μ L of 10 mg/mL bovine serum albumin (Promega). After 2 minutes at 95°C, the cycling conditions were as follows: 40 cycles of denaturation at 95°C for 15 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 30 seconds. For β -actin, denaturation occurred at 95°C for 60 seconds,

annealing at 64°C for 60 seconds, and extension at 72°C for 60 seconds. All samples were tested in duplicate, and average values were used for quantification.

Relative expression of genes of interest was normalized to that of β -actin, and gene expression in each sample was then compared with expression in HB4A cells. The comparative C_T method ($\Delta\Delta C_T$) was used for quantification of gene expression and relative expression was calculated as $2^{-\Delta\Delta C_T}$.

Results

Clinical response. We have first analyzed clinical response and gene expression in 38 patients as a training set, and 13 other patients were later analyzed as a validation set. Thirty-one (81.6%) patients who presented a \geq 30% tumor reduction, based on Response Evaluation Criteria in Solid Tumors guidelines, were classified as responders and 7 (18.4%) as nonresponders in the training set. Only two complete pathologic responses were detected.

We could not find a difference in clinical response, menopausal status, clinical stage, estrogen receptor, progesterone receptor, ErbB2 and P53 (detected in 50.7% of all the samples analyzed) immunohistochemistry expression, and tumor histologic type and grade (only 10.3% grade 1) between the 51 patients who were included in the study and the remaining 28 patients who had not their samples analyzed. No differences were detected between these two groups of patients for the following variables: age, sum of the diameter of the lesions (breast and lymph node) before chemotherapy, percentage of tumor reduction, percentage of dose of doxorubicin administered, duration of chemotherapy, and number of dissected and involved lymph nodes, indicating that there was no selection bias. In addition, among all included patients, we could not find any association between response to chemotherapy and menopausal status, histologic type and grade, estrogen receptor, progesterone receptor, and P53 and ErbB2 expression. There was neither a difference in patient age, chemotherapy dose, or duration between groups of responders and nonresponders.

Differentially expressed genes. At first, gene expression was analyzed on a cDNA microarray slide with 692 gene sequences (QT-02), most of them identified previously as expressed in normal or tumoral mammary tissue. Using a nominal *P* of 0.05 (permuted Student's *t* test), 25 (3.8%) transcripts were differentially expressed between responders and nonresponders; however, not a single one resisted a 5% FDR. The maximum differential gene expression ratio observed was 2.3-fold overexpression in nonresponders and 1.87-fold overexpression in responders. Unsupervised hierarchical clustering analysis was unable to recognize two patterns of doxorubicin response.

As the primary aim of our work was to find a predictor for clinical response or nonresponse, and the number of samples was small, we considered classifiers that used only a subset of the genes, and trios were our best choice, based on the size of our training set. Using the SVM-based feature selection algorithm, we were able to define sets of three transcripts (Supplementary Table S2). Spread error of the 10 first classifiers varied from 0.0982 to 0.1049. We then chose the first ranked classifier trio as our best candidate, as it could correctly classify up to 94.7% of our training data set (100% of sensitive tumors and 71.4% of the resistant ones). Genes comprised in this classifier were *family sequence similarity 14, member B*

(*FAM14B*), *pre-B-cell colony-enhancing factor (PBEF)*, and *elastin microfibril interface-located protein 1 (EMILIN1)*; Fig. 1). This trio was tested by the cross-validation leave-one-out procedure, resulting in 5.41% error. Using the SVM classifier, 100% sensibility and 94.7% specificity were attained to identify responsive patients among the training set. On the other hand, this trio misclassified 4 of the 13 new patients.

We have additionally analyzed the gene expression of 31 samples (24 responsive and 7 nonresponsive) from the former training group, and 13 samples from the validation group, using another cDNA microarray platform, with 4,608 sequences. Only a few genes were commonly ($n = 183$) spotted on both cDNA microarray platforms, and the three classifier genes, *EMILIN1*, *PBEF*, and *FAM14B*, defined in the first one, were not among them. After Lowess normalization and log transformation, discriminatory genes were selected as those with a FDR < 0.01.

Among the training samples, 187 genes were differentially expressed between resistant and sensitive tumors (representing 4.1% of the sequences analyzed), 98 overexpressed and 89 underexpressed in resistant samples. Hierarchical clustering using the differentially expressed genes identified two groups of tumors, with high reliability, as shown by the bootstrap technique (Fig. 2). All resistant tumors clustered together as well as 91.7% of the sensitive ones. Differentially expressed transcripts were mainly located on some specific chromosomes: 1, 2, 9, 5, 6, 10, 17, 18, and 8 (pFDR < 0.05). In addition, our data were searched for main differences between resistant and sensitive tumors, considering the biological processes in which differentially expressed genes are involved against all sequences spotted on the slides. Differentially expressed genes were involved in cell cytoskeleton and migration (microspike biogenesis, microtubule-based process, sequestering of actin monomers, and cell substrate junction assembly), cell homeostasis and metabolism (cell homeostasis; ornithine metabolism; folic acid and derivative biosynthesis; rRNA catabolism; and protein-nucleus import, docking), apoptosis (induction of

apoptosis via death domain receptors, induction of proapoptotic gene products, and apoptotic nuclear changes), mitosis (mitosis and centrosome cycle), and cell differentiation (odontogenesis and spermatid development; pFDR < 0.05; <http://vortex.cs.wayne.edu>).

Using the same 4.8K-01 platform and all the available 44 samples from the training and validation sets (9 resistant and 35 sensitive tumors), 228 genes, or 4.9% of the sequences spotted on the cDNA microarray slides, were found to be differentially expressed, 95 underexpressed and 133 overexpressed in resistant compared with sensitive tumors. Unsupervised hierarchical clustering analysis using the 228 differentially expressed genes could, once more, distinguish all the resistant as well as 85.7% of the sensitive tumors (Fig. 3).

Most of the differentially expressed genes, on both previous analysis, were involved in cellular physiologic process (84%), metabolism (57%), regulation of physiologic process (24%), cell communication (25%), and regulation of cellular process (22%) according to their gene ontology annotated as biological process at level 3 (19). In addition, differential expression varied from 1.2 to 4.5 times and 1.3 to 4.2 times, under-expression and overexpression, respectively, in resistant tumors. Among the 187 and 228 genes differentially expressed in samples of the training set alone and the training plus validation set, respectively, 124 were common, representing 66.3% and 54.4% of all the differentially expressed genes.

Therefore, using the SVM-based feature selection algorithm, we have again defined sets of three transcripts as done previously, and *serine protease 11 or insulin-like growth factor-binding protein 5 protease (PRSS11)* and *metastasis suppressor 1 (MTSS1)* were present in 6 of the 10 first ranked; however, these two genes together could not separate the samples. Eight of trios presented a spread error of 0.04 and we have chosen one of them, *PRSS11*, cleft lip and palate-associated transmembrane protein 1 (*CLPTM1*), and *MTSS1*. This trio could properly group all samples of the training set, and cross-validation leave-one-out procedure resulted in a 100% correct classification. In addition,

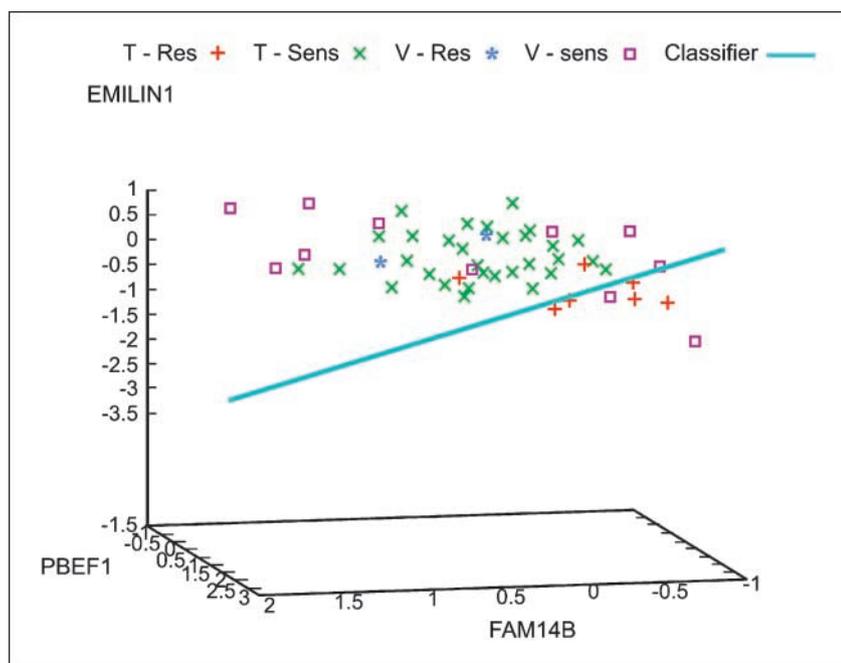


Fig. 1. Three-dimensional distribution of tumor samples according to the expression of three expressed genes as evaluated in QT-02 platform: *PBEF*, *EMILIN1*, and *FAM14B*. A linear classifier was designed using three genes that were found by a SVM-based feature selection algorithm as the best classifying triple among all. Expression values for each gene are represented on each axis. Each tumor is represented by a signal: training set (T): red cross, drug resistant (Res); green cross, drug sensitive (Sens); validation set (V): purple square, drug sensitive; blue star, drug resistant. The classifier is a plane in light blue. The trio misclassified two samples from the training set and four from the validation set (Q127, I22, I33, I01, Q52, and Q164), which gives recall/precision of 93.94%/100.00% for the training set and 66.67%/81.81% for validation set.

this trio could correctly classify 84.6% of the 13 samples from the validation set, and only samples I-01 (resistant) and Q-52 (sensitive) were misclassified (Fig. 4).

To validate our cDNA microarray data, 14 genes were chosen, 8 of them present on the 692-gene cDNA microarray slides (QT-02 platform) and 6 of them present on the 4,608-gene slides (4.8K-01 platform), to verify whether expression values derived from cDNA microarray were correlated with values obtained by quantitative reverse transcription-PCR. Spearman rank correlation between cDNA microarray and real-time PCR

measurements were significantly positive for 5 of 8 (62.5%) genes represented on the QT-02 platform and for 50% of the 6 genes represented on the 4.8K-01 platform (Table 2).

Discussion

Treatment with chemotherapy is often empirical despite the observation that patients are not equally susceptible to the same regimen. In cancer, drug resistance mechanisms are not clearly understood and may arise intrinsically from the plethora

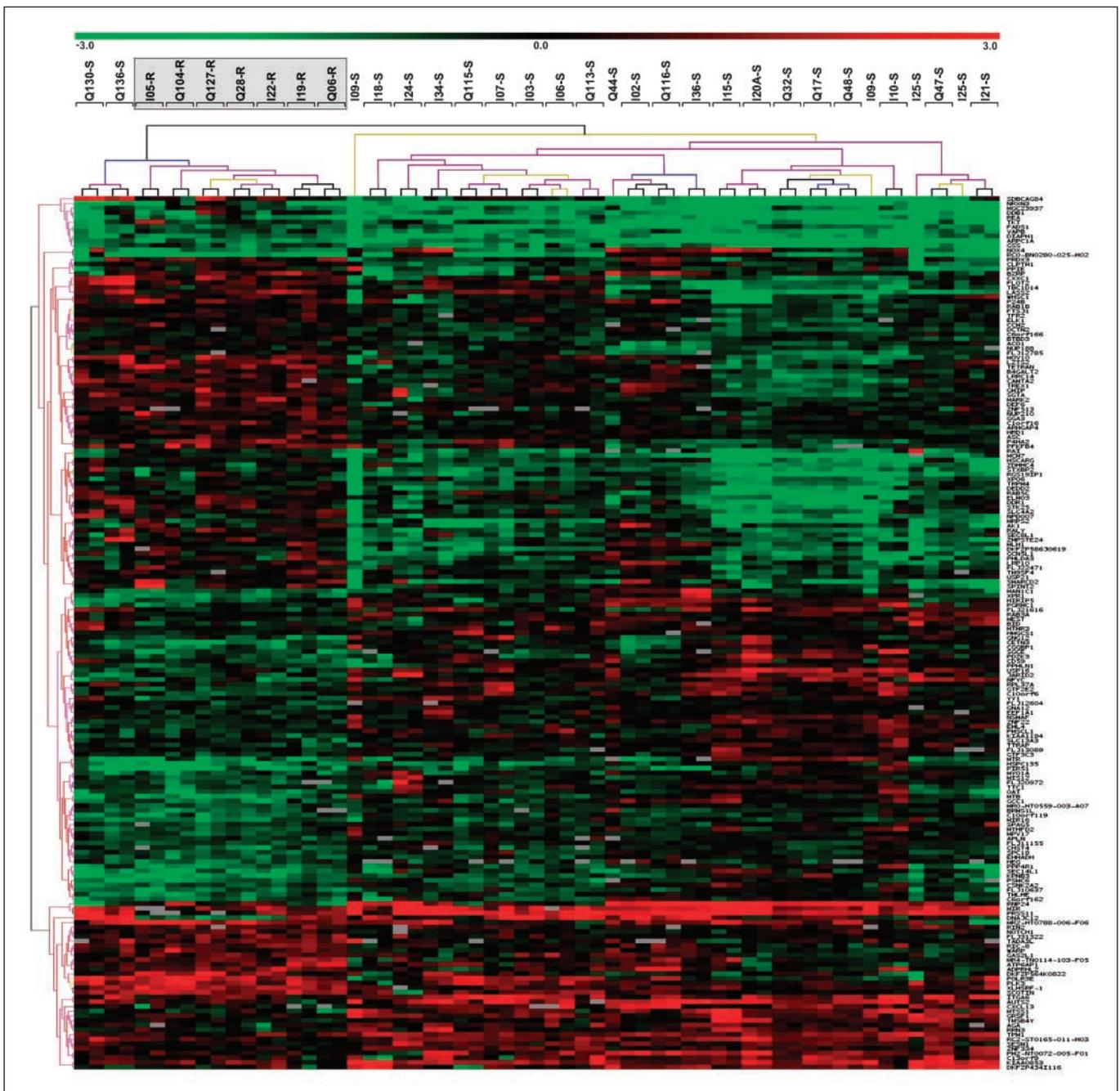


Fig. 2. Hierarchical clustering of 31 samples of the training set [24 chemotherapy sensitive (S) and 7 chemotherapy resistant (R)] as evaluated by 4.8K-01 platform containing 4,608 spotted genes. One hundred eighty-seven genes were found to be differentially expressed according to a FDR < 0.01 criterion. The colored lines of the dendrogram stand for the support for each clustering: *black* and *gray lines*, more reliable; *yellow* and *red lines*, less reliable. The metric used was Euclidean distance, with complete linkage for distance between clusters. Chemotherapy resistant samples are outlined into a *box*.

of genetic alterations during tumor progression or may be acquired through selection during chemotherapy. Studies examining the transcriptional expression profiles of cancer cell lines and human cancer xenografts have begun to identify genes that may be associated with response or resistance to doxorubicin (20, 21). However, currently, there are no molecular markers or clinical features that may, by themselves, predict response to doxorubicin chemotherapy.

We have at first analyzed the gene profile of resistant and sensitive samples against a cDNA microarray platform containing 692 genes, most of them previously described as expressed in breast tissue or breast cancer or involved in chemotherapy resistance. Although 3.8% of the genes were found to be differentially expressed, unsupervised clustering could not separate responders from nonresponders, suggesting that gene expression was homogeneous and highly superimposed between groups and that genes included in the cDNA microarray platform were unsuitable to distinguish response to chemotherapy.

Therefore, we have searched our data for predictors of clinical response or nonresponse, and a trio, including *EMILIN1*, *FAM14B*, and *PBEF* (the first two genes being also differentially expressed with the least *Ps*), was identified, with a leave-one-out estimation error of 5.41% for the training set. *EMILIN1*, an extracellular matrix with adhesive properties (22), and *FAM14B*, at first described as an IFN-induced gene but however contribute to combating cellular stress independent of the IFN system (23), were up-regulated in responsive tumors. *PBEF*, which promotes growth of B-cell precursors (24) and functions as an inhibitor of apoptosis (25), was more expressed in resistant tumors, and both mechanisms, reduced apoptosis or impaired cell cycle regulation, might be implied in drug resistance. Unfortunately, this trio could not separate samples from the validation set according to response to chemotherapy.

Our next step was to analyze gene profile in a more comprehensive cDNA microarray platform, containing 4,608 ORESTES tags (11), and ~4% genes were found differentially expressed between responsive and resistant tumors. Unsupervised clustering could correctly group all the resistant as well as at least 85% of the sensitive samples.

Consistent with an apoptosis induction mode of action for doxorubicin, sensitive tumors had higher expression of apoptosis-related genes as *BID*, *PAWR*, *CARD8*, and *TTRAP*. Genes involved or correlated with cell proliferation, such as *AREG* (*amphiregulin*), *CDKL1* (*cyclin-dependent kinase like-1*), *ESP8* (*epidermal growth factor receptor pathway substrate 8*), *PRSS11*, *GMFB* (*glia maturation factor*), and several other genes coding for proteins involved in cell cycle checkpoint control, including *PIR51* (*RAD51*) and *BRAP* (*BRCA1-associated protein*), or M phase of cell cycle, such as *CETN3* (*centrosome duplication*), *PCM1* (*pericentriolar material*), *MML4* (*microtubule-associated protein*), *MIS12*, *MTB* (*chromosome segregation*), *SPAG5* (*dynamic regulation of mitotic spindles*), and metastasis suppressors (*BRMS1L1*, *MTSS1*, and *DLC1*), were more expressed in responsive tumors. In addition, a group of genes involved in the ubiquitin-proteasome pathway, including *PSMC6* (*proteasome 26*), *USP* (*ubiquitin-specific protease*), *HSPC135* (*proteasome subunit*), and *RANBP2* (*GTP-binding protein*) were up-regulated in sensitive samples.

A differential expression of transcription factors was characterized, and *RPN3* (*RNA polymerase I transcription factor*), *GTF2E2*, *GTF3C3* (*general transcription factors*), *TEAD1*, *JARID2*,

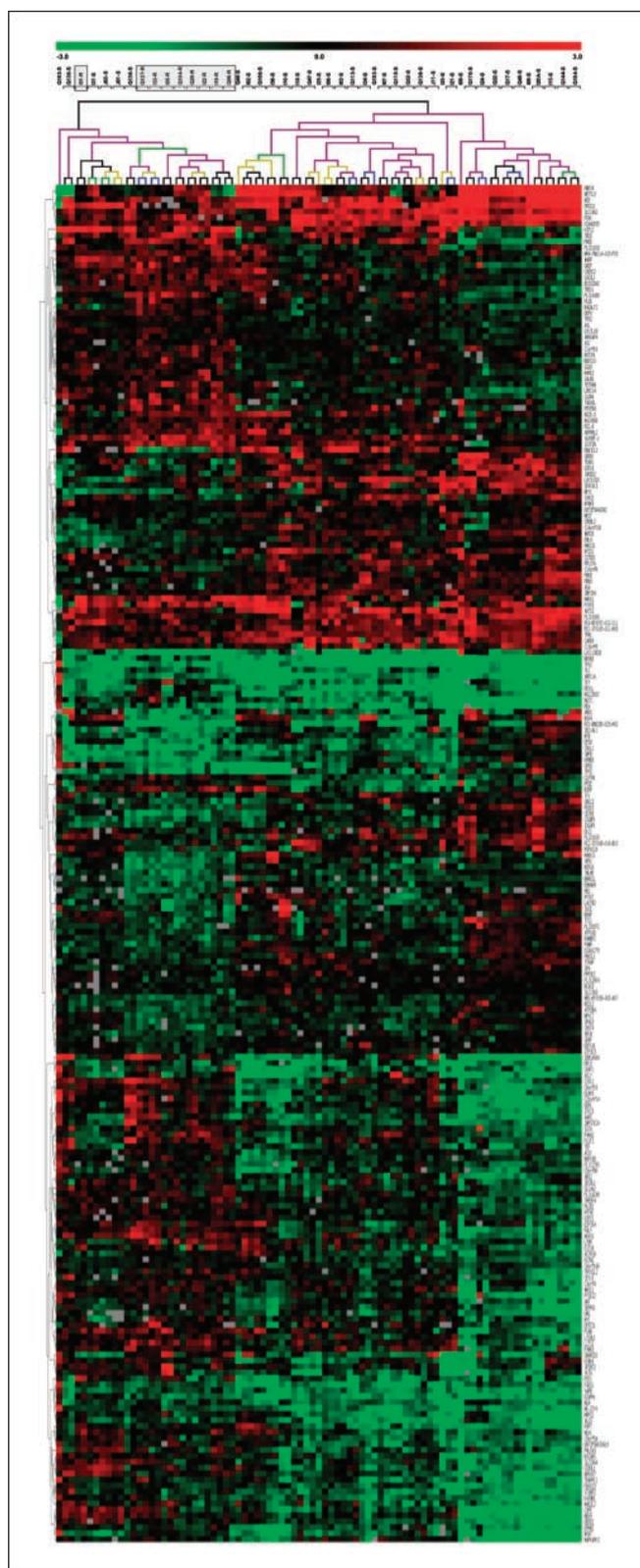
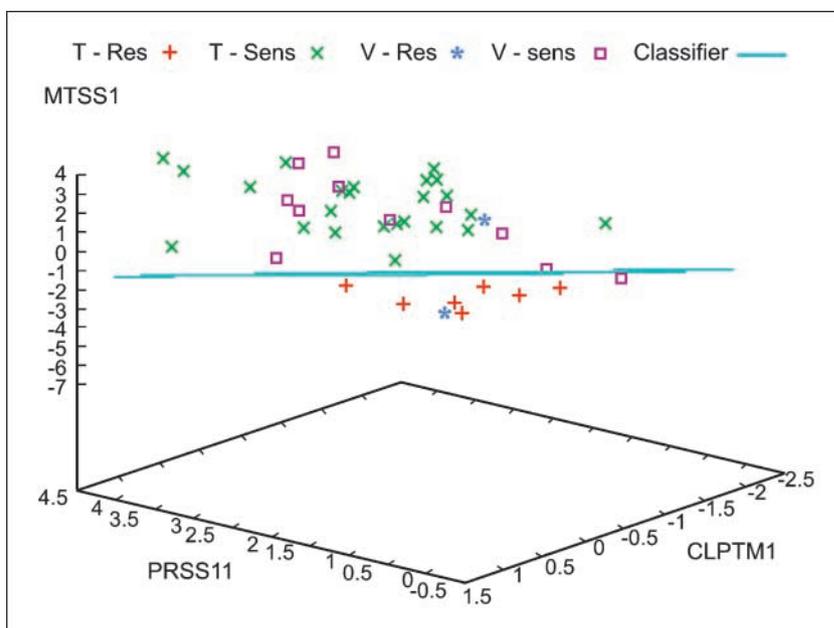


Fig. 3. Hierarchical clustering of 44 samples of the training and validation sets [35 chemotherapy sensitive (S) and 9 chemotherapy resistant (R)] as evaluated by 4.8K-01 platform containing 4,608 spotted genes. Two hundred twenty-eight genes were found to be differentially expressed according to a FDR < 0.01 criterion. The colored lines of the dendrogram stand for the support for each clustering: *black and gray lines*, more reliable; *yellow and red lines*, less reliable. The metric used was Euclidean distance, with complete linkage for distance between clusters. Chemotherapy resistant samples are outlined into a box.

Fig. 4. Three-dimensional distribution of tumor samples according to the expression of three expressed genes as evaluated in 4.8K-01 platform: *MTSS1*, *PRSS11*, and *CLPTM1*. Expression values for each gene are represented on each axis. Each tumor is represented by a signal: training set (T): red cross, drug resistant (*Res*); green cross, drug sensitive (*Sens*); validation set (V): purple square, drug sensitive; blue star, drug resistant. The classifier is a plane in light blue. Therefore, it misclassified zero samples in the training set and two in the validation set (I33 and Q52), which gives recall/precision of 100.00%/100.00% for training set and 84.62%/90.90% for validation set.



and several ZN finger transcription factors were more expressed in sensitive tumors, in contrast to resistant tumors, which presented an up-regulation of *NOTCH1*, *SMARCD2* (matrix associated, actin dependent), *myc*, *MCF7*, *TEF*, *TADA3L*, and *CXXC1*.

Several genes involved in DNA repair were associated to resistant tumors (*REV1*, *MLH1*, *UNG*, and *TREX1*). Another finding was the up-regulated expression of several genes associated to protein transport, membrane traffic, or vesicle docking in the resistance group, including *STXBP2*, *SEC8L1*, *COPE*, *GGA3*, *RAB5C*, *RAB1B*, *RIN2* (*Ras* and *RAB* interactor 2), *BZRP* (*benzodiazepine receptor*), *XPO6* (*exportin 6*), *NUP188*, *NUP120*, *VAPB* (*vesicle-associated membrane protein*), *TETRAN* (*tetracycline transport-like protein*), *TM9SF4* (*transmembrane 9 superfamily protein*), and *TRAPPC1* (*trafficking protein particle*), and genes associated to cytoskeleton organization, including *TBCD* (*tubulin-specific chaperone*), *TPX2* (*microtubule-associated protein*), *KIFC2* (*microtubule-associated complex*), *MARK2* (*microtubule affinity-regulating kinase*), *KATNB1* (*disassemble microtubule*), *DCTN2* (*dynactin*), *ARPC1A* and *ACTR1B* (*actin-related proteins*), *ARHGAP4* (*Rho GTPase*), *GAS2L1* (*actin-associated protein in growth-arrested cells*), *flotilin*, and *fibromodulin*. Some of the genes differentially expressed in our list or from the same families have been shown previously to be involved with chemotherapy response (26–28).

We have again looked for a trio that could distinguish responders from nonresponders, and *PRSS11*, *MTSS1*, and *CLPTM1* could correctly classify 95.4% of the 44 samples analyzed, with only two misclassifications, one sensitive tumor and one resistant tumor. It is important to note that the mean spread error of the 10 first ranked trios determined in this analysis was 2.5-fold lower than the mean spread error of the 10 first ranked classifiers using the QT-02 platform, meaning that these trios better separate samples based on doxorubicin response.

PRSS11 may influence the activity of the insulin-like growth factor pathway, which stimulate the proliferation and differentiation of a vast number of cell types. There is also some evidence that down-regulation of *PRSS11* expression may represent an

indicator of melanoma progression. On the other hand, *PRSS11* overexpression in the metastatic melanoma cell line strongly inhibited proliferation and chemoinvasion *in vitro* as well as cell growth *in vivo* (29). *PRSS11* was the least expressed gene in resistant tumors compared with sensitive ones, which may indicate that *PRSS11* overexpression in sensitive tumors may be linked to growth inhibition on chemotherapy.

MTSS1 was also significantly less expressed in resistant versus sensitive tumors. *MTSS1* was identified in some cancer cell

Table 2. Correlation of cDNA microarray gene expression data with quantitative reverse transcription-PCR–derived values

	<i>n</i>	Pearson correlation		Spearman rank correlation	
		<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>
<i>FOS</i> *	9	0.900	0.001	0.800	0.01
<i>EMILIN1</i> *	11	0.306	0.360	0.309	0.355
<i>FAM14A</i> *	16	0.524	0.037	0.698	0.003
<i>PBEF</i> *	15	0.764	0.001	0.721	0.002
<i>MAL2</i> *	13	0.509	0.076	0.484	0.094
<i>CPNE3</i>	19	0.760	<0.001	0.731	<0.001
<i>262664.0R</i> *	17	0.320	0.211	0.328	0.198
<i>262638.0R</i> *	16	0.351	0.183	0.509	0.044
<i>CTGF</i> #	13	0.574	0.040	0.637	0.019
<i>SMOC2</i> #	12	0.581	0.048	0.860	<0.000
<i>DUSP1</i> #	12	0.727	0.011	0.655	0.029
<i>PKL3</i> #	11	0.856	0.002	0.067	0.855
<i>C20orf45</i> #	12	0.03	0.994	0.343	0.276
<i>SRPRB</i> #	13	-0.206	0.499	-0.011	0.972

NOTE: *, genes present on the 692-gene cDNA microarray slides (QT-02 platform); #, genes present on the 4,608-gene cDNA microarray slides (4.8K-01 platform); *n*, number of samples evaluated.

lines, but its expression was not detected in metastatic cells of bladder, breast, and prostate cancers (30). *MTSS1* is an actin-binding protein involved in cytoskeleton and cell projection organization, which may also associate with transcription factors to affect nuclear signaling (31). Overexpression of *MTSS1* causes formation of abnormal actin structures in NIH 3T3 cells and reduces proliferation of PC-3 prostate cancer cells (32). These data suggest that a reduction of *MTSS1* gene expression in chemotherapy-resistant tumors may contribute to tumor growth.

CLPTM1 was identified in a family with cleft lip and palate as a novel gene disrupted by a translocation t(2;19); however, its role in clefting was not well established. It encodes a putative protein with seven transmembrane domains ubiquitously expressed in both adult and embryonic tissues (33). Interestingly, *CLPTM1* is homologous with *cisplatin resistance-related gene 9 (CRR9)* up-regulated in a cisplatin-resistant ovarian tumor cell line (34). Our data showed that *CLPTM1* was more expressed in resistant compared with sensitive samples, and its similarity to *CRR9* may suggest that these genes may be involved with resistance to various chemotherapy agents.

There is no clear explanation, at the moment, why two samples were misclassified by the trio. Patient I-01 was a premenopausal woman who presented a clinical stage III invasive ductal carcinoma grade 3, hormonal receptors negative, ErbB2 positive, Ki-67 positive (75%) tumor, with no response to primary chemotherapy. She developed a bone recurrence 16.1 months after mastectomy. Patient Q-52, a postmenopausal woman with invasive ductal carcinoma clinical stage II, grade 1, hormone receptors and ErbB2 positive tumor, presented a partial response to chemotherapy and no involved lymph nodes on resection. She remains disease free after a 23.7-month observation.

A few studies have been looking for a gene profile that might predict response to primary chemotherapy in breast cancer.

Chang et al. (5) and Ayers et al. (6) have searched for genes associated to response to docetaxel and paclitaxel-doxorubicin, respectively. In addition, Bertucci et al. (7) and Hannemann et al. (8) have attempted to identify an expression signature associated to doxorubicin response, the first one exclusively in inflammatory breast cancer. All these works have analyzed a similar number of samples (24-42) as we did using arrays containing 8,016 to 30,721 spotted sequences.

Although distinct patterns of chemotherapy response were shown previously by Chang et al. (5), unsupervised clustering could not discriminate the groups according to other authors (6-8). Otherwise, except for one of these reports (8), a small percentage of genes, representing <1% of the initial number of genes, was able to identify a gene pattern associated with chemotherapy sensitivity, with prediction accuracy ranging from 62% to 78%, indicating that only a small subset of transcripts is connected to response to chemotherapy (6, 7). In addition, Ayers et al. (6) and Bertucci et al. (7) determined a gene profile able to recognize not all responsive patients but exclusively those that will benefit the most (as represented by attaining a complete pathologic response).

Doxorubicin is one of the mainstay drugs in the treatment of breast cancer and the most widely used worldwide. We have now analyzed patients with noninflammatory, mostly advanced breast cancer to determine gene expression pattern associated with doxorubicin sensitivity/resistance. Experiments done on a cDNA microarray platform with 4,608 sequences were able to identify differential transcripts that could distinguish groups according to response to chemotherapy. Additionally, a set of three genes, *PRSS11*, *MTSS1*, and *CLPTM1*, could correctly classify 95% of the samples. We believe this classifier needs to be optimized with a larger number of clinical samples.

References

- Fisher B, Bryant J, Wolmark N, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. *J Clin Oncol* 1998;16:2672-85.
- van der Hage JA, van de Velde CJ, Julien JP, Tubiana-Hulin M, Vandervelden C, Duchateau L. Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. *J Clin Oncol* 2001;19:4224-37.
- Wolff AC, Davidson NE. Primary systemic therapy in operable breast cancer. *J Clin Oncol* 2000;18:1558-69.
- Hortobagyi GN, Hayes D, Pusztai L. Integrating newer science into breast cancer prognosis and treatment: a review of current molecular predictors and profiles. *ASCO Annu Meet Summ* 2002;192-201.
- Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003;362:362-9.
- Ayers M, Symmans WF, Stec J, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* 2004;22:2284-93.
- Bertucci F, Finetti P, Rougemont J, et al. Gene expression profiling for molecular characterization of inflammatory breast cancer and prediction of response to chemotherapy. *Cancer Res* 2004;64:8558-65.
- Hannemann J, Oosterkamp HM, Bosch CA, et al. Changes in gene expression associated with response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* 2005;23:3331-42.
- Therasse P, Arbuuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000;92:205-16.
- Dias Neto E, Correa RG, Verjovski AS, et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc Natl Sci Acad Sci USA* 2000;97:3491-6.
- Brentani RR, Carraro DM, Verjovski-Almeida S, et al. Gene expression arrays in cancer research: methods and applications. *Crit Rev Oncol Hematol* 2005;54:95-105.
- Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM. High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol* 2000;18:457-9.
- Gomes LI, Silva RL, Stolf BS, et al. Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray. *Anal Biochem* 2003;321:244-51.
- O'Hare MJ, Ormerod MG, Monaghan P, Lane EB, Gusterson BA. Characterization *in vitro* of luminal and myoepithelial cells isolated from the human mammary gland by cell sorting. *Differentiation* 1991;46:209-21.
- Saeed AI, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34:374-8.
- Kim S, Dougherty ER, Barrera J, Chen Y, Bittner ML, Trent JM. Strong feature sets from small samples. *J Comput Biol* 2002;9:127-46.
- Kim S, Dougherty ER, Shmulevich I, et al. Identification of combination gene sets for glioma classification. *Mol Cancer Ther* 2002;1:1229-36.
- Silva PJS, Hashimoto R, Kim S, et al. Feature selection algorithms to find strong genes. *Pattern Recog Lett* 2005;26:1444-53.
- Al-Shahrour F, Díaz-Uriarte R, Dopazo J. Fatigo: a web tool for finding significant associations of Gene Ontology terms to groups of genes *Bioinformatics* 2004;20:578-80.
- Turton N, Judah DJ, Riley J, et al. Gene expression and amplification in breast carcinoma cells with intrinsic and acquired doxorubicin resistance. *Oncogene* 2001;20:1300-6.
- Zembutsu H, Ohnishi Y, Tsunoda T, et al. Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Res* 2002;62:518-27.
- Zanetti M, Braghetta P, Sabatelli P, et al. EMILIN-1 deficiency induces elastogenesis and vascular cell defects. *Mol Cell Biol* 2004;24:638-50.
- Parker N, Porter AC. Identification of a novel gene family that includes the interferon-inducible human genes 6-16 and ISG12. *BMC Genomics* 2004;5:8.
- Kitani T, Okuno S, Fujisawa H. Growth phase-dependent changes in the subcellular localization of pre-B-cell colony-enhancing factor. *FEBS Lett* 2003;544:74-8.
- Jia SH, Li Y, Parodo J, et al. Pre-B cell colony-enhancing factor inhibits neutrophil apoptosis in experimental inflammation and clinical sepsis. *J Clin Invest* 2004;113:1318-27.

26. Kudoh K, Ramanna M, Ravatn R, et al. Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. *Cancer Res* 2000;60:4161–6.
27. Levenson VV, Davidovich IA, Roninson IB. Pleiotropic resistance to DNA-interactive drugs is associated with increased expression of genes involved in DNA replication, repair, and stress response. *Cancer Res* 2000;60:5027–30.
28. Iwao-Koizumi K, Matoba R, Ueno N, et al. Prediction of docetaxel response in human breast cancer by gene expression profiling. *J Clin Oncol* 2005;23:422–31.
29. Baldi A, De Luca A, Morini M, et al. The HttA1 serine protease is down-regulated during human melanoma progression and represses growth of metastatic melanoma cells. *Oncogene* 2002;21:6684–8.
30. Lee Y-G, Macoska JA, Korenchuk S, Pienta KJ. MIM, a potential metastasis suppressor gene in bladder cancer. *Neoplasia* 2002;4:291–4.
31. Callahan CA, Ofstad T, Horng L, et al. MIM/BEG4, a Sonic hedgehog-responsive gene that potentiates Gli-dependent transcription. *Genes Dev* 2004;18:2724–9.
32. Loberg RD, Neeley CK, Adam-Day LL, et al. Differential expression analysis of MIM (MTSS1) splice variants and a functional role of MIM in prostate cancer cell biology. *Int J Oncol* 2005;26:1699–705.
33. Yoshiura K, Machida J, Daack-Hirsch S, et al. Characterization of a novel gene disrupted by a balanced chromosomal translocation t(2;19)(q112;q133) in a family with cleft lip and palate. *Genomics* 1998;54:231–40.
34. Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. *Biochem Biophys Res Commun* 2001;280:1148–54.