Downsizing Multigenic Predictors of the Response to Preoperative Chemotherapy in Breast Cancer

Conference Paper · September 2008

DOI: 10.1007/978-3-540-85565-1_20 · Source: DBLP

8 authors, including:

René Natowicz
University of Paris-Est
83 PUBLICATIONS **244** CITATIONS
SEE PROFILE

Roberto Incitti
King Abdullah University of Science and Technology
45 PUBLICATIONS **354** CITATIONS
SEE PROFILE

Roman Rouzier
Institut Curie
590 PUBLICATIONS **13,387** CITATIONS
SEE PROFILE

Cela Arben
Ecole Supérieure d'Ingénieur en Electronique et Electrotechnique - Paris
126 PUBLICATIONS **775** CITATIONS
SEE PROFILE

Downsizing Multigenic Predictors of the Response to Preoperative Chemotherapy in Breast Cancer

René Natowicz[1], Roberto Incitti[2], Roman Rouzier[3], Arben Çela[1], Antõnio Braga[4], Euler Horta[4], Thiago Rodrigues[5], Marcelo Costa[4], and Carmen D. M. Pataro[4]

[1] Université Paris-Est, ESIEE-Paris, France, {r.natowicz, a.cela}@esiee.fr
[2] Institut Mondor de Médecine Moléculaire, Créteil, France, roberto.incitti@inserm.fr
[3] Hôpital Tenon, department of Gynecology, Paris, France, roman.rouzier@tnn.aphp.fr
[4] Universidade Federal de Minas Gerais, Depto. Engenharia Eletrônica, Brazil, {apbraga, eulerhorta, carmen}@cpdee.ufmg.br, macosta.est@gmail.com
[5] Universidade Federal de Lavras, Depto. Ciência da Computação, Brazil, thiago@dcc.ufla.br

Abstract. We present a method for designing efficient multigenic predictors with few probes and its application to the prediction of the response to preoperative chemotherapy in breast cancer. In this study, each DNA probe was regarded as an elementary predictor of the response to the chemotherapy and the probes which were selected performed a faithful sampling of the training dataset. In a first stage of the study, the prediction delivered by a multigenic predictor was that of the majority of the elementary predictions of its probes. For the data set at hand, the best majority decision predictor (MD predictor) had 30 probes. It significantly outperformed the best predictor previously published, which was designed on probes that had been selected by p-value of a t-test. In a second stage, the majority decision was replaced by a support vector machine (SVM) with linear kernel. With the same set of probes, the performances of the SVM predictor were slightly better for both training and testing sets of data. The main improvement was that the performances of the best MD predictor were achieved by the SVM predictors with only 17 probes. This more than 40% downsizing of the predictors is an interesting property for the potential use of the predictors in clinical routine, and for the task of modeling the biological mechanisms underlying the patient's response to the chemotherapy.

1 Introduction

Nowadays, adjuvant and neoadjuvant (preoperative) administration of chemotherapy is based on prognostic factors, not on predictive ones. It is well known that the prognostic factors do not provide enough information for tailoring the treatment to the individual patients. Hence, nearly all breast cancer patients are given a standard chemotherapy treatment, despite their potentially poor response to the therapy, adverse side effects, and healthcare costs.

The ability to predict the patients' response to the chemotherapy would be of high interest in the treatment of breast cancer for avoiding useless chemotherapy treatments

and for selecting the most effective regimen for every patient. To this end, no single factor or biomarker ever has been in position to discriminate the patients who would respond to the treatment from those who would not. It appears that primary chemotherapy provides an ideal opportunity to correlate the gene expressions with the response to the treatment. Although gene expression microarrays provide novel tools and hold great promise in cancer research, the achievements in terms of improved prediction of drug sensitivity have been thus far rather moderate [1]. A strategy for translating microarray profiles into efficient clinical tests could consist in identifying small diagnostic gene-expression profiles with the help of microarrays then, in a second step, to validate the clinical usefulness of these genes, either retrospectively or prospectively, by making use of a simple and robust conventional assay, such as the quantitative reverse-transcriptase polymerase chain reaction (RT-PCR). Such a strategy requires microarrays analysis methods able to provide oncogenic signatures made out of few probe sets.

In the present study, every selected probe delivered an elementary prediction of the response to the treatment: pathologic complete response (*pcr*), residual disease (*nopcr*), or *unspecified*. In a first stage, we have defined very simple predictors whose predictions of the patient's response were those of the majority of their probes' elementary predictions: *PCR* if the majority of the probes predicted the response to be *pcr*; *NoPCR* if the majority was *nopcr*; and *UNSPECIFIED* in case of tie. In the second part of the study, the classification criterion of majority decision (MD) was replaced by support vector machines (SVM) with linear kernel. The resulting classifiers had slightly better training and testing performances with the same numbers of probes. Moreover, the performances of the best MD predictor were achieved by a SVM predictor with significantly less probes, 17 probes instead of 30, i.e. were more than 40% downsized.

In this paper, we will present the low level treatment through which a probe delivered an elementary prediction of the patient's response; the valuation function by which the probes were ranked then selected in this ranking; and we will give the performances of the MD predictors then those of the SMV predictors for the same dataset.

2 Patients and Data

This work was conducted based on data from Hess *et al.* The clinical trial was conducted at the Nellie B. Connally Breast Center of The University of Texas M.D. Anderson Cancer Center [2]. One hundred thirty-three patients with stage I-III breast cancer were included. All patients underwent a single-pass, pretreatment fine-needle aspiration of the primary breast tumor before starting chemotherapy. Pretreatment gene expression profiling was performed with oligonucleotide microarrays (Affymetrix U133A) on fine-needle aspiration specimens. Patient cases were separated into patient *training cases* (82 cases) and patient *testing cases* (51 patient cases). At the completion of neoadjuvant chemotherapy, all patients had surgical resection of the tumor bed, with negative margins. Pathologic complete response (PCR) was defined as no histopathologic evidence of any residual invasive cancer cells in the breast, whereas residual disease was defined as any residual cancer cells after histopathologic study. The low level treatment of the microarray data was performed by software dCHIP V1.3 to generate probe level intensities. This program normalizes all arrays to one standard array that represents a

chip with median overall intensity. Finally, normalized gene expression values were transformed to the $\log_{10}$ scale for analysis because in microarrays, the log-values of the expression levels are closer to normal distributions than the non-transformed ones.

The set of training cases was composed of 82 patient data, each of which being the response to the treatment and the expression levels of the 22283 DNA probes. Among the training set, the response to the treatment was PCR for 21 patient cases and NoPCR for 61 cases. The testing set was composed of 51 patient data among which the response to the treatment was PCR for 13 patient cases and NoPCR for 38 patients. Hence, the ratios of PCR to NoPCR patient cases were the same for both the training and testing datasets.

3 Probes Valuation

A first research on these data is reported In [2]. In this study, the authors the probes were selected in the ranking of the p-value of a t-test, and the study was more focused on the question of choosing the best decision model for the prediction. To this end, the authors checked several models of classification: k-nearest neighbors, support vector machine, diagonal linear discriminant analysis, with various parameters for each of them. A total of 780 different classifiers have been evaluated and statistically assessed, all of them taking as input the expression levels of the probes selected by the p-values of a t-test. This study has shown that, for this criterion of probes selection, the best predictor was a one taking as input the expression levels of the 30 probes of least p-values, these expression levels being weighted thanks to a diagonal linear discriminant analysis (predictor DLDA-30). For the set of validation cases, the performances of the DLDA-30 predictor were[1]: accuracy=0.76, sensitivity=0.92, specificity=0.71, PPV=0.52, NPV=0.96.

Hence, the predictor DLDA-30 had very high sensitivity and negative predictive values. Taking the results of this study as a starting point, we wanted to design predictors with higher specificity values, while preserving the very high sensitivity value of the DLDA-30 predictor. Furthermore, because the study reported in [2] had put the emphasis on the classification models and was very complete in this regard, we have decided to investigate the process of probes selection, with the wish of improving the performances with models of classification as simple as possible.

For the dataset at hand, it has appeared to us that the probes selection criterion of p-value to a t-test favored probes that mainly gave some information on the membership of the patients to the NoPCR class. For instance, the probe of smallest p-value was the probe 203929_s_at, of the gene MAPT. The box-plot of its expression levels (fig. 1)

[1] Definitions: let TP and TN be the numbers of true positives and negatives returned by a given predictor, and let FP and FN be the numbers of false positives and negatives. The accuracy of the predictor is the proportion of correctly predicted cases, the sensitivity that of correctly predicted PCR cases, the specificity is the proportion of correctly predicted NoPCR cases, the positive predictive value (PPV) is the probability of a case predicted as PCR to be a PCR case, and the negative predictive value (NPV) is the probability of a case predicted NoPCR to be a NoPCR case. Therefore, these criteria are: accuracy=(TP+TN)/(TP+FN+TN+FP), sensitivity=(TP)/(TP+FN), specificity=(TN)/(TN+FP), the positive predictive value PPV is (TP)/(TP+FP) and the negative predictive value is (TN)/(TN+FN).

shows that given a high expression level, the probability of NoPCR membership was high. It also shows that a low expression level for this probe did not provide any information on the patient's class: in this case, the probability to predict a PCR was more or less that of a random choice with probabilities $\frac{P}{P+N}$ and $\frac{N}{P+N}$, where P and N are the respective numbers of PCR and NoPCR cases of the training set.

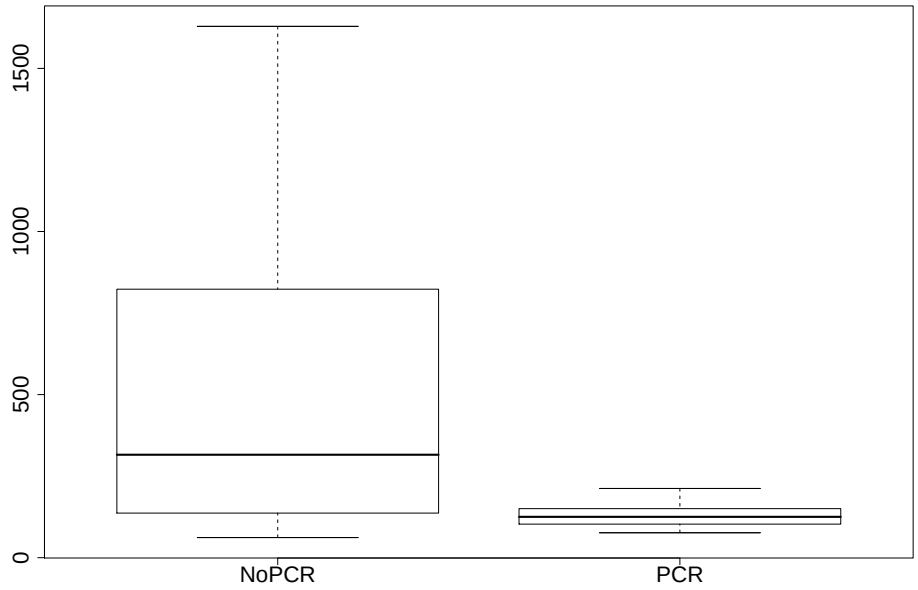

Fig. 1. Probe 203929_s_at of the gene MAPT, probe of smallest p-value to a t-test: box-plot of the expression levels.

Because we aimed at designing predictors with better performances and simple classification criteria, we were interested in putting into light DNA probes conveying information on the two classes of patients, PCR and NoPCR. Qualitatively speaking, we wanted to find probes whose interval of expression levels respectively computed on the PCR and on the NoPCR subsets of learning cases, would have a "small intersection", and which would deliver an information non biased by the overrepresentation of the NoPCR cases in the learning set.

To this end, we have chosen to assign two sets of expression levels to each probe s, the sets $E_p(s)$ and $E_n(s)$, computed from the training data as follows [3,4]. Let $m_p(s)$ and $sd_p(s)$ be the mean and standard deviation of the expression levels of the probe s for the PCR training cases, and let $m_n(s)$ and $sd_n(s)$ be those of the NoPCR training cases. The set of expression levels of the PCR training cases was defined as the set difference $E_p(s)$, $E_p(s) = [m_p(s) - sd_p(s), m_p(s) + sd_p(s)] \setminus [m_n(s) - sd_n(s), m_n(s) + sd_n(s)]$ and conversely for the NoPCR training cases, $E_n(s) = [m_n(s) - sd_n(s), m_n(s) + sd_n(s)] \setminus [m_p(s) - sd_p(s), m_p(s) + sd_p(s)]$.

Discrete probes' predictions. For any patient case, the individual prediction of a probe was a discrete value in the set $\{pcr,\ nopcr,\ unspecified\}$: *pcr* if the expression level of patient p was in the interval $E_p(s)$ and *nopcr* if it was in $E_n(s)$. Otherwise, the individual prediction value was *unspecified*.

Probes' valuation function. Let $p(s)$ be the number of PCR training cases correctly predicted pcr by the probe s, and let $n(s)$ be the number of NoPCR training cases correctly predicted $nopcr$ by the probe. The valuation function of the probes was defined so as to favor probes which correctly predicted high numbers of training cases and whose sets of correctly predicted training cases were 'good' samplings of the training set. To this end, we have considered the sensitivity and specificity values of the probe s, i.e. the ratios $p(s)/P$ and $n(s)/N$ of correctly predicted training cases. The valuation function $v(s)$, $v(s) \in [0, 1]$, was defined as the sum of its sensitivity and specificity, $v(s) = 0.5 \times \left(\frac{p(s)}{P} + \frac{n(s)}{N}\right)$ (the coefficient ensures that the valuation is in the unit interval).

The figure 2 is the box-plot of the expression levels of a DNA probe of the gene BTG3, for the patients of the training set. This probe was one of the two equally top ranked probes, (cf. table 3.) From this figure, one can see that, given that the expression

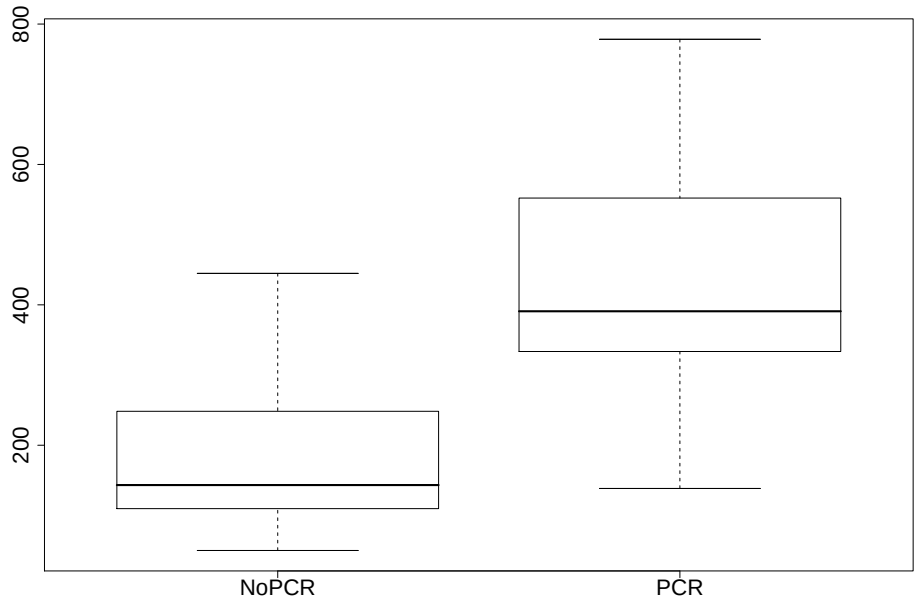

Fig. 2. Probe 205548_s_at of the gene BTG3, top ranked for the valuation function $v(s)$: box-plot of the expression levels.

level of this probe was high, there was a high probability for the patient to have a pathologic complete response and symmetrically, given that the expression level was low, a high probability of residual disease. Hence, this probe delivered an information about both PCR and NoPCR classes membership. Up to a high rank, the probes selected by

decreasing values conveyed information on both classes. In the ranking of the valuation function, the first *mono-informative* probe (giving information about only one class of patients) was at rank 63 (probe 207067_s_at of the gene HDC, cf. fig. 3.)

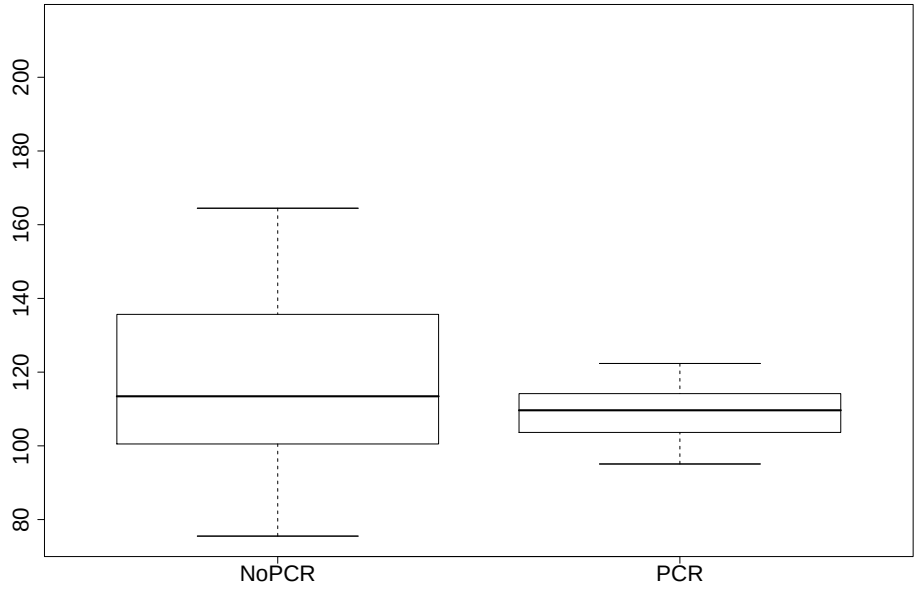

Fig. 3. Probe 207067_s_at of the gene HDC, top ranked mono-informative probe for the valuation function $v(s)$: box-plot of the expression levels.

4 Multigenic Predictors with Majority Decision

We have defined the k-probes majority decision predictor (MD predictor) as the k top ranked probes for the valuation function $v(s)$ together with the classification criterion of majority decision: for each patient case, the prediction was PCR if the number of elementary *pcr* predictions was strictly greater than the number of elementary *nopcr* ones, the prediction was NoPCR for the converse situation, and UNSPECIFIED in case of tie. In figure 4 are the training and testing accuracies of the first 41 MD k-predictors ($0 \leq k \leq 40$.) In table 3 are the numbers of false positives and false negatives of the same set of MD k-predictors.

The predictor of highest training accuracy was the 38-probes predictor: acc.=0.85, sens.=0.86, specif.=0.85, npv=0.945 (negative predictive value), ppv=0.67 (positive predictive value), corresponding to 3 FN (false negatives) and 9 FP (false positives) out of 21 PCR and 61 NoPCR cases. Its performances on the test set were: acc.=0.84, sens.=0.77, specif.=0.87, npv=0.917, ppv=0.67, corresponding to 3 FN and 5 FP out of 13 PCR and 38 NoPCR cases.

Gene	Probe	$v(s)$	$p(s)$	$n(s)$	Gene	Probe	$v(s)$	$p(s)$	$n(s)$
BTG3	213134_x_at	0.61	12	40	GATA3	209602_s_at	0.41	13	13
BTG3	205548_s_at	0.61	12	40	BBS4	212745_s_at	0.41	3	42
GATA3	209604_s_at	0.59	15	29	DAPK1	203139_at	0.41	9	24
GATA3	209603_at	0.49	12	26	SAS	203226_s_at	0.40	7	29
THRAP2	212207_at	0.46	8	34	FLJ10916	219044_at	0.40	8	26
SCCPDH	201826_s_at	0.46	12	22	E2F3	203693_s_at	0.40	8	26
SIL	205339_at	0.45	10	27	AHNAK	220016_at	0.40	9	23
KRT7	209016_s_at	0.45	6	38	KLHDC3	214383_x_at	0.40	9	23
MCM5	201755_at	0.45	7	35	SFRS12	212721_at	0.40	9	23
NME3	204862_s_at	0.44	10	25	SRPK1	202200_s_at	0.39	6	31
METRN	219051_x_at	0.44	11	22	CXCR4	217028_at	0.39	8	25
PDE4B	211302_s_at	0.43	9	27	KIF3A	213623_at	0.39	8	25
PHF15	212660_at	0.42	7	32	MGC4771	210723_x_at	0.39	8	25
SSR1	200891_s_at	0.42	7	32	C11orf15	218065_s_at	0.39	9	22
PISD	202392_s_at	0.42	11	20	CELSR1	41660_at	0.39	12	13
MELK	204825_at	0.41	8	28	LAD1	203287_at	0.39	4	36
CA12	215867_x_at	0.41	10	22	LU	203009_at	0.38	6	30
CA12	214164_x_at	0.41	10	22	LIPE	213855_s_at	0.38	7	27
MAPK3	212046_x_at	0.41	10	22	GAMT	205354_at	0.38	7	27

Table 1. The 38 probes of highest valuations. Gene: gene name in Hugo Gene nomenclature [5]; probe: reference of the Affymetrix DNA probe set; $v(s)$: probe valuation; $p(s)$, $n(s)$: numbers of correct pcr and nopcr predictions for the 21 PCR and 61 NoPCR cases of the training set. Total numbers of pcr and nopcr predictions: 301 and 900, ratio=0.33.

Gene	Probe	$v(s)$	rank	$p(s)$	$n(s)$	Gene	Probe	$v(s)$	rank	$p(s)$	$n(s)$
MAPT	203929_s_at	0.22	780	0	28	AMFR	202204_s_at	0.23	662	0	29
MAPT	203930_s_at	0.291	218	2	30	CTNND2	209617_s_at	0.27	337	0	33
BBS4	212745_s_at	0.41	21	3	42	GAMT	205354_at	0.38	38	7	27
MAPT	203928_x_at	0.22	781	0	28	CA12	204509_at	0.24	566	1	27
THRAP2	212207_at	0.46	5	8	34	FGFR1OP	214124_x_at	0.37	52	6	28
MBTPS1	217542_at	0.26	391	0	32	KIAA1467	213234_at	0.25	475	3	22
MAPT	206401_s_at	0.22	900	0	27	METRN	219051_x_at	0.44	11	11	22
PDGFRA	215304_at	0.32	118	4	28	FLJ10916	219044_at	0.40	24	8	26
ZNF552	219741_x_at	0.24	564	1	27	E2F3	203693_s_at	0.40	25	8	26
RAMP1	204916_at	0.22	774	0	28	ERBB4	214053_at	0.21	1040	0	26
BECN1	208945_s_at	0.30	165	4	26	JMJD2B	215616_s_at	0.37	45	7	26
BTG3	213134_x_at	0.61	1	12	40	RRM2	209773_s_at	0.37	51	3	37
SCUBE2	219197_s_at	0.15	3078	0	19	FLJ12650	219438_at	0.27	293	0	34
MELK	204825_at	0.41	16	8	28	GFRA1	205696_s_at	0.18	1994	0	22
BTG3	205548_s_at	0.61	2	12	40	IGFBP4	201508_at	0.38	39	7	27

Table 2. The 30 probes of smallest p-values to the t-test [1] (*ranks* are the probes' rankings for the valuation function $v(s)$.) Total numbers of pcr and no pcr predictions: 123 and 894, ratio=0.13

The MD predictors of second highest training accuracy had $k = 30$, $k = 33$ and $k = 34$ probes: acc.=0.84, sens.=0.81, specif.=0.85, npv=0.929, ppv=0.65 (4 FN and 9 FP). On the set of testing cases the performances of the 30-probes predictor were: acc.=086, sens.=0.92, specif.=0.84, npv=0.970, ppv=0.67, corresponding to 1 FN and 6 FP. The performances of the 33 and 34-probes predictors were approximatively the same on the set of testing cases (cf. table 3.)

	Training		Test	
k	FP	FN	FP	FN
1	21	9	18	7
2	21	9	18	5
3	19	6	16	4
4	22	7	15	4
5	22	7	15	4
6	21	6	15	3
7	18	6	13	3
8	17	6	12	4
9	15	6	9	4
10	16	6	7	4
11	15	6	7	3
12	14	6	7	3
13	15	5	7	3
14	11	4	8	3
15	12	3	7	3
16	12	3	8	3
17	12	3	8	3
18	12	3	8	3
19	11	3	9	2
20	12	3	8	2
30	**9**	**4**	**6**	**1**
21	11	3	8	2
22	11	3	6	2
23	11	4	7	2
24	10	4	7	3
25	10	4	7	2
26	10	4	6	2
27	10	4	6	1
28	10	4	5	3
29	10	4	6	1
33	**9**	**4**	**4**	**2**
34	**9**	**4**	**5**	**2**
38	**9**	**3**	**5**	**3**
31	10	4	5	2
32	10	4	5	2
35	10	4	5	2
36	10	4	5	2
37	10	4	5	2
39	10	4	5	2
40	10	4	5	2

Table 3. Confusion table for majority decision predictors. FP, FN: number of false positives and false negatives for training and testing cases. In bold: 1st and 2nd maximum training accuracies.

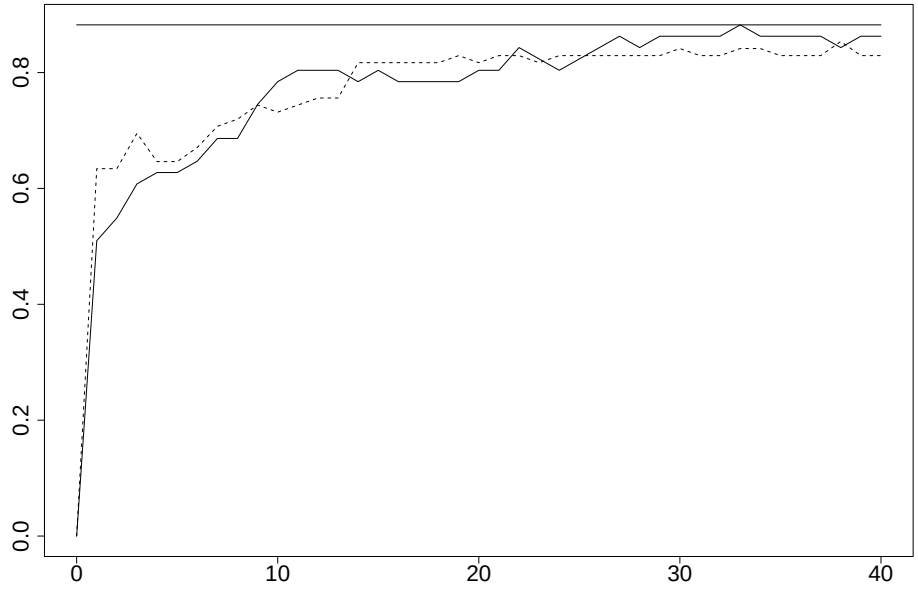

Fig. 4. Training and testing accuracies of the majority decision predictors. Solid line: testing accuracy, dashed line: training accuracy, horizontal line: maximum testing accuracy (0.88) for $k = 33$ probes. X axis: number of probes.

5 Training Set Sampling

Because the criterion the most widely used for selecting DNA probes in microarray studies for cancer research is the p-value of a t-test [1], and because the predictors designed with probes selected by our valuation function outperformed those designed with probes selected by the p-value [3], we were interested in finding a parameter which could account for the difference of performances. It has appeared that the quality of the sampling performed by a given set of probes could be this explicative parameter.

The ratio of the numbers of PCR to NoPCR training cases of the data set at hand was $\frac{P}{N} = \frac{21}{61} = 0.34$. For the valuation $v(s)$, this ratio was in excellent agreement with that of the total numbers of *pcr* to *nopcr* correct predictions of the k top ranked probes. For $20 \leq k \leq 50$ probes, the values of the ratios were between 0.33 and 0.34, and below 20 probes, the ratios were between 0.30 and 0.38. For the set of 30 top ranked probes, the mean number of correct predictions per probe was 35.16 and the ratio of the *pcr* to *nopcr* numbers of predictions was 0.34 (equal to the ratio of PCR to NoPCR numbers of cases in the training set).

For the p-value of a t-test, the set of 30 top ranked probes comprised 11 *mono-informative* probes (in the present case, their numbers of pcr predictions was null, cf. table 3). The mean number of correct predictions per probe was 37.67 (approximatively equal to that of the probes selected according to our probes valuation function), but the ratio of the *pcr* to *nopcr* numbers of predictions was far lesser, this ratio value being precisely equal to that of the whole set of probes (0.14). From this, one could see that, with the same mean number of predictions per probe, the quality of the sampling performed by a set of probes, measured by the ratio of the pcr to nopcr numbers of predictions, could explain the difference of performances between the two methods of probes selection. The 30 probes of least p-values to a t-test performed a sampling in which the *nopcr* predictions were over-represented while the 30 probes of highest values $v(s)$ performed a more faithful sampling of the training set, the numbers of *pcr* and *nopcr* predictions being in the ratio of the numbers of PCR to NoPCR testing cases.

6 Multigenic Predictors with Support Vector Machine

We have defined the k-probes support vector machine predictor (SVM predictor) as the k top ranked probes for the valuation function $v(s)$ together with a linear kernel SVM [6], [7]. The maximum training accuracy of the SVM k-predictors (table 4 and figure 5) was achieved for $k = 15$ probes: acc.=0.88, sens.=0.90, specif.=0.87, npv=0.96, ppv=0.70, corresponding to 2 false negatives and 8 false positives (out of 21 PCR and 61 NoPCR cases). The testing performances of the 15-probes predictor were: acc.=0.82, sens.=0.92, specif.=0.79, npv=0.968, ppv=0.6, corresponding to 1 false negative and 8 false positives (out of 13 PCR and 38 NoPCR cases). Hence, the SVM predictor of highest training accuracy had slightly better training performances than the best MD predictor, and approximatively the same testing performances, with twice less probes (15 vs. 30 probes).

The second maximum training accuracy was achieved with $k = 17, 18, 19, 20$ probes and $k = 27, 28$ probes. On the plateau from 17 to 20 probes, the predictors

had approximatively the performances of the $k = 15$ probes predictor for both training and testing accuracies. Because of this plateau, and because the 15-probes predictor appeared isolated, we have considered that 17-probes was a safer lower limit for downsizing the best MD predictor. On the small plateau of two predictors showing the second maximum training accuracy, $k = 27, 28$ probes, the testing performances were better than those of the best MD predictor but, obviously, without significant downsizing. The testing performances were: acc.=0.88, sens.=0.92, spec.=0.87, npv=0.971, ppv=0.71, corresponding to 1 false negative and 5 false positives. One noticed that this plateau of two predictors was the beginning of a longer plateau of predictors, from $k = 27$ to $k = 32$ probes, showing approximatively the same training performances and the same, or even better, testing performances.

Hence, from these results, one might say that the performances of the best MD predictor, the MD 30-probes predictor, were achieved by the SVM 17-probes predictor, i.e. with 43% less probes. In particular, their sensitivity values on the test set were equal (0.92) and their negative predictive values were almost equal (31/32=0.967 for the SVM 17-probes vs. 32/33=0.970 for the MD 30-probes predictor). In addition, the SVM 17-probes predictor showed slightly better training performances than the MD 30-probes predictor.

	Training		Testing		
k	FP	FN	FP	FN	SV
1	21	4	18	3	82
2	11	9	5	8	82
3	11	9	7	5	82
4	7	10	2	7	82
5	10	7	5	5	82
6	5	9	2	8	82
7	9	7	2	5	82
8	7	7	3	5	82
9	7	8	2	5	82
10	11	7	4	4	82

	Training		Testing		
k	FP	FN	FP	FN	SV
11	13	4	7	3	74
12	10	2	8	2	48
13	12	2	9	1	48
14	10	2	7	1	51
15	**8**	**2**	**8**	**1**	**54**
16	10	2	7	2	51
17	**9**	**2**	**7**	**1**	**51**
18	**9**	**2**	**7**	**1**	**52**
19	**9**	**2**	**8**	**1**	**51**
20	**9**	**2**	**8**	**1**	**50**

	Training		Testing		
k	FP	FN	FP	FN	SV
21	9	3	7	1	71
22	9	3	6	1	67
23	9	3	6	1	82
24	9	4	5	1	72
25	9	4	6	1	79
26	10	4	6	1	82
27	**7**	**4**	**5**	**1**	**79**
28	**8**	**3**	**5**	**1**	**70**
29	8	4	4	1	77
30	8	4	5	1	78

	Training		Testing		
k	FP	FN	FP	FN	SV
31	8	4	4	1	79
32	8	4	4	1	79
33	9	4	4	1	80
34	9	3	5	1	80
35	10	4	5	1	80
36	11	2	8	1	82
37	10	3	5	1	80
38	11	2	8	1	82
39	11	3	7	1	82
40	9	4	5	2	82

Table 4. Confusion table for support vector machine predictors. FP, FN: number of false positives and false negatives for training and testing cases; SV: number of support vectors. In bold: maximum training accuracies.

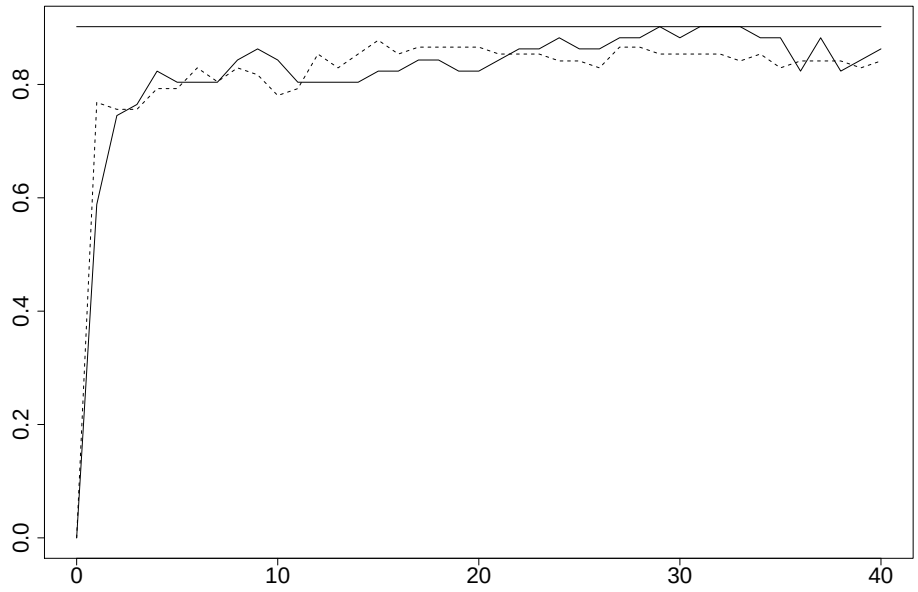

Fig. 5. Training and testing accuracies of the support vector machine with linear-kernel. Solid line: testing accuracy, dashed line: training accuracy, horizontal line : maximum testing accuracy (0.90) for $k = 29, 31, 32, 33$ probes. X axis: number of probes.

Conclusion

With our approach of features selection, the DNA probes were regarded as elementary predictors of the response to the chemotherapy. We have presented a valuation function through which the probes behaving as good samplers of the training set were assigned high values. Two classifier models were evaluated for these probes: the non weighted majority decision among the elementary predictions of the probes (MD-predictors), and a support vector machine with linear kernel (SVM-predictors). The probes making up the predictors were selected in the ranking of their values. The MD-predictor of highest accuracy for the training data was designed with the 30 probes of highest values. On the independant testing set, its accuracy, sensitivity, specificity, negative and positive predictive values were respectively 0.86, 0.92, 0.84, 0.97, and 0.67, which significantly outperformed the best predictors designed on probes selected by the p-values of a t-test (whose expression levels had been weighted by a diagonal linear discriminant analysis). We tend to think that our method of probes selection has revealed relevant genes because the significant improvement of the predictors' performances have been obtained by using a very simple decision criterion: the majority decision. In a second stage, using a support vector machine with linear kernel instead of the majority decision, the performances of the best MD predictor have been achieved with 17 probes instead of 30, hence the best MD predictor was downsized by more than 40%.

Because of the very high negative predictive value (0.97), such predictors could be of interest for supporting the decision of not allocating a patient to the treatment.

Beside, the predictor would unadvisedly allocate 16% of the non-responder patients to the treatment (specificity=0.84). This result has to be compared to the nowadays almost systematic allocation to the treatment.

The small number of probes involved in the SVM predictor (17 probes) could allow the design of predictors at very low cost, which is an important issue for their potential use in clinical routine. Furthermore, because of the focusing on a small number of potentially relevant genes, this downsizing could also be an interesting property for modeling the biological mechanisms underlying the response to the chemotherapy.

Acknowledgments. The authors would like to thank the CAPES-COFECUB french-brazilian cooperation program and the CNPq for their support.

References

1. Knudsen, S.: Cancer Diagnostics with DNA Microarrays. John Wiley & Sons, Inc. (2006)
2. Hess, K.R., Anderson, K., Symmans, W.F., Valero, V., Ibrahim N., Mejia, J.A., Booser, D., Theriault, R.L., Buzdar, A.U., Dempsey, P.J., Rouzier, R., Sneige, N., Ross, J.S., Vidaurre, T., Gómez, H.L., Hortobagyi, G.N., Pusztai, L.: Pharmacogenomic Predictor of Sensitivity to Preoperative Chemotherapy with Paclitaxel and Fluorouracil, Doxorubicin, and Cyclophosphamide in Breast Cancer. J. Clin. Onc. 24, pp. 4236-4244 (2006)
3. Natowicz, R., Incitti, R., Horta E.G., Charles, B., Guinot, P., Yan, K., Coutant, C., Andre, F., Pusztai, L., Rouzier, R.: Prediction of the Outcome of Preoperative Chemotherapy in Breast Cancer by DNA Probes that Convey Information on Both Complete and Non Complete Responses. BMC Bioinformatics 9, 149 (2008)
4. Natowicz, R., Braga, A.P., Incitti, R., Horta, E.G., Rouzier, R., Rodrigues, T.S, Costa, M.A.: A New Method of DNA Probes Selection and its Use with Multi-Objective Neural Networks for Predicting the Outcome of Breast Cancer Preoperative Chemotherapy. In: 16th European Symposium on Artificial Neural Networks, pp. 71-76. d-side ed., Evere, Belgium (2008)
5. Wain, M.H., Bruford, E.A., Lovering, R.C., Lush, M.J., Wright, M.W., Povey S.: Guidelines for Human Gene Nomenclature. Genomics, 79, pp. 464-470 (2002)
6. Cortes, C., Vapnik, V.: Support-vector networks. Machine Leaming, 20, 273–297 (1995)
7. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2, pp. 121-167 (1998)