

*Featured Article*

## Recursive Partitioning as an Approach to Selection of Immune Markers for Tumor Diagnosis

James A. Koziol, Jian-Ying Zhang,  
Carlos A. Casiano, Xuan-Xian Peng, Fu-Dong Shi,  
Anne C. Feng, Edward K. L. Chan, and  
Eng M. Tan<sup>1</sup>

Division of Biomathematics [J. A. K., A. C. F.] and W. M. Keck Autoimmune Disease Center, Department of Molecular and Experimental Medicine [J.-Y. Z., F.-D. S., E. K. L. C., E. M. T.], The Scripps Research Institute, La Jolla, California; Department of Biological Sciences, The University of Texas at El Paso, El Paso, Texas [J.-Y. Z.]; Department of Biochemistry and Microbiology, Center for Molecular Biology and Gene Therapy, Loma Linda University School of Medicine, Loma Linda, California [C. A. C.]; Department of Biology, School of Life Sciences, Xiamen University, Xiamen, People's Republic of China [X.-X. P.]; and Department of Oral Biology, University of Florida Health Sciences Center, Gainesville, Florida [E. K. L. C.]

### Abstract

**Purpose and Experimental Design:** Cancer sera contain antibodies which react with a unique group of autologous cellular antigens called tumor-associated antigens (TAAs), but the low frequency of positive reactions against any individual antigen has precluded use of autoantibodies as useful diagnostic markers. With enzyme immunoassay, we examined antibody frequencies to a panel of seven TAAs, c-myc, cyclin B1, IMP1, Koc, p53, p62, and survivin, in 527 cancer patients (64 breast cancer patients, 45 colorectal cancers, 91 gastric cancers, 65 hepatocellular carcinomas, 56 lung cancers, and 206 prostate cancers), and 346 normals. We used recursive partitioning to assess whether we could accurately classify individuals as either cancer patients or normals on the basis of the profile of antibody reactivity to the seven TAAs for each individual.

**Results:** Recursive partitioning resulted in the selection of subsets of the seven-panel TAA, which differentiated between tumors and controls, and these subsets were unique to each cancer cohort. The classification trees had sensitivities ranging from 0.77 to 0.92 and specificities ranging from 0.85 to 0.91 in the cancer cohorts when normal means +2

SDs were used as standard cutoffs for immunoassay positivity. Antibody to cyclin B1 was the initial discriminating node for gastric and lung cancers, and for hepatocellular carcinoma, and was a subsequent discriminating node in all of the other cancer cohorts. c-myc was the initial discriminating node in breast cancer, p62 in prostate cancer, and IMP1 in colon cancer. Recursive partitioning demonstrated that no more than three of the seven TAAs were needed for any cancer cohort to arrive at these levels of sensitivity and specificity.

**Conclusions:** This initial study shows that multiple antigen miniarrays can provide accurate and valuable tools for cancer detection and diagnosis. Performance of the miniarrays might be enhanced by other combinations of TAAs appropriately selected for different cancer cohorts.

### Introduction

Autoimmunity as manifested by the development of antibodies to autologous cellular antigens is increasingly observed to be associated with cancer. For example, autoantibodies to p53, the tumor suppressor protein, were first reported in 1982 (1), and since then there have been numerous reports confirming and extending this finding (reviewed in Ref. 2). Although factors leading to the production of such autoantibodies are not completely understood, we have proposed (3) that autoantibodies might be used as reporters identifying aberrant cellular mechanisms in tumorigenesis. In addition, the use of autoantibodies as serological markers for cancer diagnosis is feasible because of the general absence of particular autoantibodies in normal individuals and in noncancer conditions. In a preliminary investigation in a large population of cancer patients (4), we found that antibody frequencies to two TAAs,<sup>2</sup> p62 and Koc, were 11.6% and 12.2%, respectively. Some sera that were negative for antibody to one TAA were positive for antibody to the other so that with a two-antigen panel as the substrate, the number of positive reactors increased to 20.5%. This raised the possibility that if other TAAs were selected and the antigen panel was increased in number, the frequency of antibodies might be additionally increased.

In the current study, we examine antibody frequencies to seven TAAs, c-myc, cyclin B1, IMP1, Koc, p53, p62, and survivin, in 527 cancer patients (64 breast cancer patients, 45 colorectal cancers, 91 gastric cancers, 65 HCCs, 56 lung cancers, and 206 prostate cancers), and 346 normals. We use a multivariate statistical methodology, recursive partitioning, to assess whether we can accurately classify individuals as either cancer patients or normals on the basis of the profile of antibody reactivity to the seven TAAs in each individual. We show that

Received 12/10/02; revised 7/8/03; accepted 7/18/03.

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.

This work was supported in part by grants from the National Cancer Institute (CA56956), the Division of Research Resources (RR00833), and by technical resources from the Stein DNA Core Facility of The Scripps Research Institute. F.-D. S. is supported by a development grant from the Muscular Dystrophy Association.

<sup>1</sup> To whom requests for reprints should be addressed, at W. M. Keck Autoimmune Disease Center, Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037. Phone: (858) 784-8686; Fax: (858) 784-2131; E-mail: emtan@scripps.edu.

<sup>2</sup> The abbreviations used are: TAA, tumor-associated antigen; HCC, hepatocellular carcinoma.

with a panel of seven selected TAAs reasonable sensitivity and specificity can be achieved in discriminating between cancer patients and normals; hence, TAA miniarrays might provide a viable approach to tumor detection and diagnosis.

## Materials and Methods

**Patients and Sera.** The serum bank of the Tumor Cell Engineering Laboratory of Xiamen University has a collection of sera from cancer patients that has been used for epidemiological and other studies, and this laboratory made available sera from 321 patients with different types of cancer, including 64 with breast cancer, 45 with colorectal cancer, 91 with gastric cancer, 65 with HCC, and 56 with lung cancer. Sera from 206 prostate cancer patients treated at Loma Linda University Medical Center were also available for this study. As controls, sera were obtained from 82 individuals from Fujian province who were having annual health examinations and who had no evidence of malignancy. In addition, 264 control sera from the San Diego area came from a large epidemiological and genetic study of hemochromatosis in which these "normal" sera were obtained from the nonhospitalized adult San Diego population (courtesy of Dr. Ernest Beutler, The Scripps Research Institute). In total, then, we had sera from 527 cancer patients and 346 normals. This study was approved by the Institutional Review Boards of the respective academic centers.

### Expression and Purification of Recombinant TAAs.

Seven antigens, c-myc, cyclin B1, IMP1, Koc, p53, p62, and survivin, were selected for expression of recombinant proteins. Recombinant p62 was expressed from a clone derived from cDNA expression library by immunoscreening with antibody from a patient with HCC (5). p62 cDNA was subcloned into pET28a vector producing a fusion protein with NH<sub>2</sub>-terminal 6x histidine and T7 epitope tags. The recombinant protein expressed in *Escheria coli* BL21 (DE3) was purified using nickel column chromatography. Koc cDNA cloned in the pcDNA3 vector (6) was similarly subcloned into pET28a vector and recombinant protein expressed as above. IMP1 construct pCMV5-IMP1 was kindly provided by Finn C. Nielsen (Copenhagen, Denmark; Ref. 7) and p53 clone (p53SN3) by Yuxin Yin of Columbia University (New York, NY), and were also subcloned into pET28a for protein expression. cDNA from c-myc was amplified by PCR from human fetal liver tissue and survivin cDNA from human survivin expressed sequence tag clone (BG258433) before subcloning in pET28a vector. Recombinant cyclin B1 had been prepared and used previously (8), and was isolated from a pGEX construct expressing cyclin B1 with glutathione S-transferase fusion partner. Expression of adequate amounts of recombinant protein was examined in SDS-PAGE, and Coomassie Blue staining was used to determine whether expression products of expected molecular sizes were produced. In addition, Western immunoblot analysis was used to confirm that the bands seen in SDS-PAGE were reactive with reference antibodies. The antibodies used were rabbit polyclonal anti-IMP1 from Nielsen *et al.* (7) and anti-Koc/IMP3 from Müller-Pillasch *et al.* (6), which were raised against specific COOH-terminal peptides of the respective proteins and anti-p62, which was raised against the full-length protein (5). Reactivities of p53, c-myc, and cyclin B1 were determined with monoclonal

antibodies obtained from Oncogene Research Products (Boston, MA.) Rabbit polyclonal antisurvivin antibody raised against the COOH-terminal peptide was obtained from ProSci Inc. (Poway, CA).

**ELISA.** Purified recombinant proteins were diluted in PBS to a final concentration of 0.5 µg/ml, and 200 µl aliquots were pipetted into each well to coat Immulon 2 microtiter plates (Dynatech Laboratories, Alexandria, VA). All of the human sera were diluted 1:200 in PBS, incubated with antigen-coated wells at room temperature for 90 min followed by washing with PBS, and developed with horseradish peroxidase-conjugated goat anti-human IgG (Caltag Laboratories, San Francisco, CA) using the substrate 2,2'-azinobis (3-ethylbenzothiazole-6-sulfonic acid; Boehringer Mannheim GmbH, Mannheim, Germany). Each serum sample was tested in duplicate, and the average absorbance at 490 nm was used for data analysis. Because several hundred test sera were analyzed at different time periods, each run of ELISA included 10 normal human serum samples and 2 positive control samples. The 10 NHS samples were selected to represent a range of 2 SD above and below the mean of the normals, and the average value of the 10 normals was used in each run to normalize all of the absorbance values to the standard mean of all of the normals. In addition, all of the positive sera were confirmed with repeat testing, as were some negative sera. All of the immunoassays were performed in one laboratory (E. M. T.) skilled in this technique.

**Statistical Considerations.** A multivariate statistical methodology, recursive partitioning (9–11), was used to classify individuals as either normals or cancer patients on the basis of the reactivity of each individual to the panel of 7 antigens. Recursive partitioning is nonparametric in nature, imposing no *a priori* restrictions on the distributional forms of the predictor variables. The recursive partitioning algorithm is simple and intuitive. At each step, the recursive partitioning program determines for each variable (in this case for each of the seven antibodies) a cut point that optimally splits all of the individuals into cases (cancer) and controls (normal) and selects the variable that performs best. It then takes the resulting subpopulations and repeats the process on each until no additional partitioning is warranted: either a subpopulation contains one class of individuals or the subpopulation is too small to subdivide anew. The program RPART (12); freely available on the internet,<sup>3</sup> implemented in R, was used to generate the decision trees depicting the classification rules generated through recursive partitioning. When growing each tree, we assigned equal prior probabilities to the normal and cancer cohorts, and equal misclassification costs. Pruning of the trees (to correct for overtraining) was undertaken using the 1 SE rule described by Breiman *et al.* (9).

ELISA results are typically dichotomized by selection of a threshold or cutoff absorbance value: an observed absorbance value above the threshold connotes a positive reaction, below the threshold, a negative reaction. Trees were initially constructed using prechosen cutoffs of normal means +3 SDs and normal means +2 SDs on all of the antibody assays. That is, input for each individual consisted of dichotomous variables

<sup>3</sup> Internet address: <http://lib.stat.cmu.edu>.

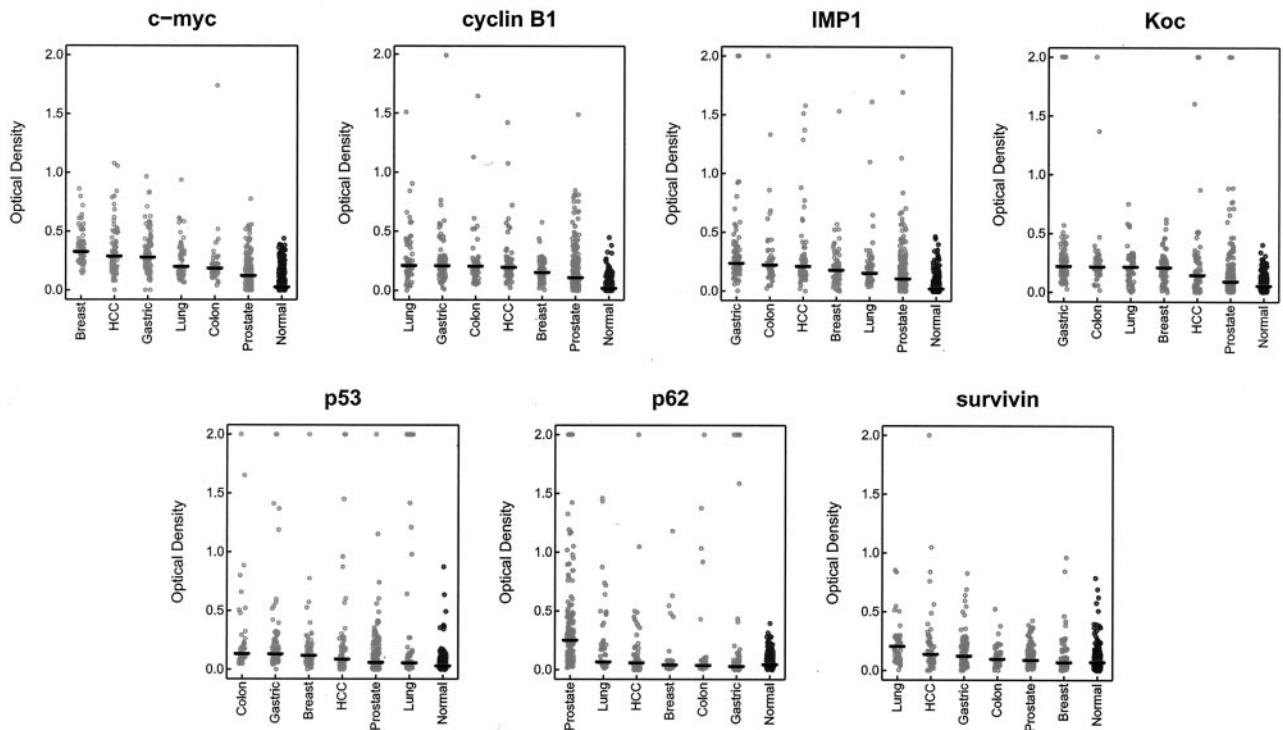


Fig. 1 Values of antibody titers to different cancer-associated antigens are expressed as absorbance units obtained from enzyme immunoassays. The sera were obtained from 64 breast cancer patients, 45 colorectal cancers, 91 gastric cancers, 65 HCCs, 56 lung cancers, 206 prostate cancers, and 346 normals. Horizontal lines mark median absorbances within each cohort.

“positive” or “negative” on each antibody assay, along with group membership (cancer cohort or normal). Normal mean + 3 SD and normal mean + 2 SD have conventionally been used to distinguish abnormal or positive from normal or negative in antibody assays, so these trees should provide benchmarks of “standard” performance. Error rates associated with these trees were determined by simple resubstitution, and are summarized by sensitivity or true positive rate (likelihood of correctly classifying cancer patients) and specificity or true negative rate (likelihood of correctly classifying normals). Because resubstitution may be overly optimistic, we also used 10-fold cross validation (13) to evaluate prospectively the error rates of these trees, as follows. Each data set was randomly divided into 10 equally sized subsets, and a classification tree was derived with 9 subsets (the learning set) and tested with the remaining subset (the training or validation subset). This cross-validation was repeated 10 times for complete coverage, and the results on the training sets combined to calculate the predictive accuracy and error rates for the tree.

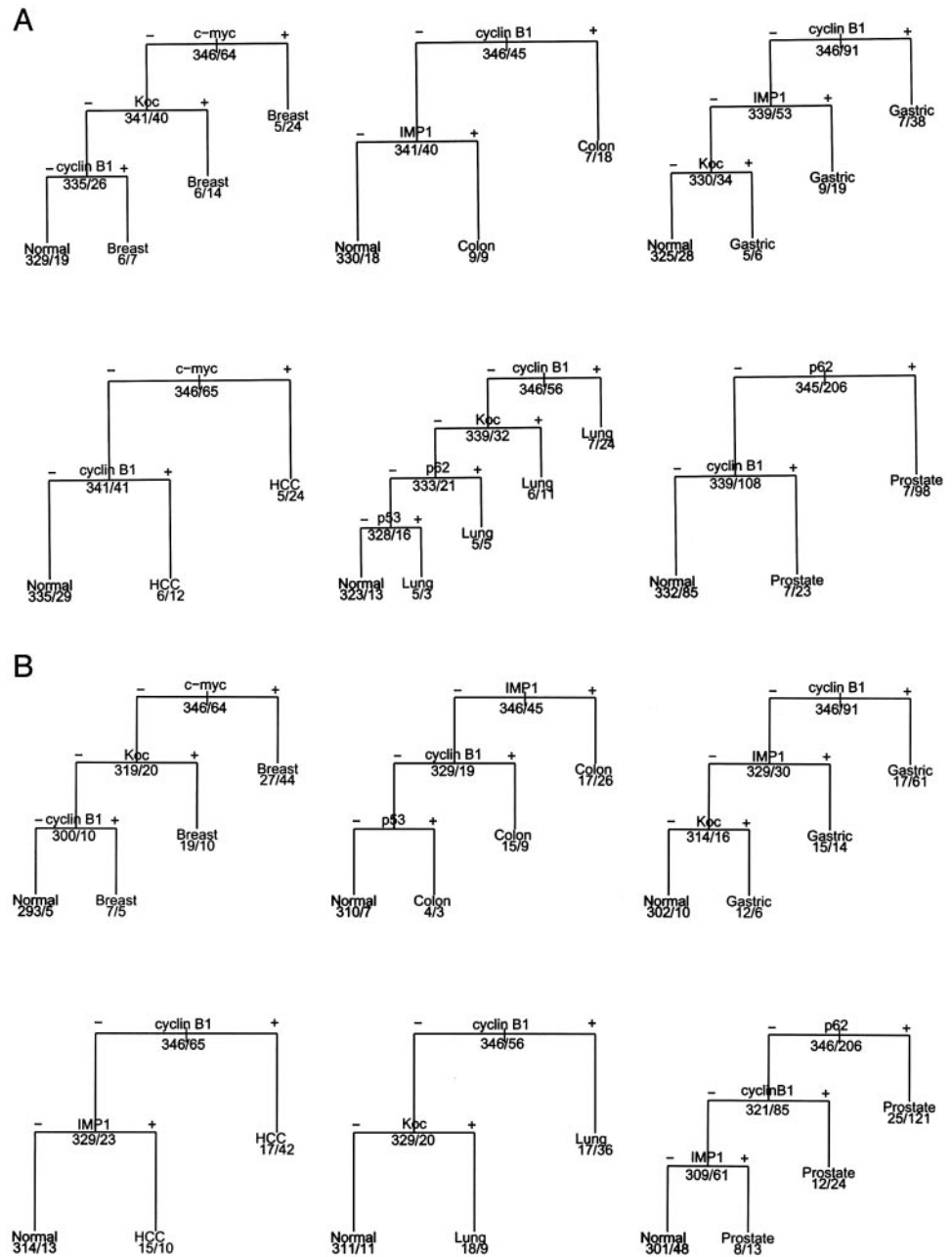
Choosing identical cutoffs in terms of means plus multiples of SDs may not be optimal in terms of operating characteristics of the recursive partitioning decision rules, so we also investigated trees in which input consisted of the actual absorbance values on each of the antibody assays for each individual rather than the dichotomous positive or negative declaration. The RPART program was allowed to select suitable cutoffs on each of the absorbance scales for classification purposes without respect to means and SDs. Again, 10-fold cross-validation was used to evaluate prospectively the error rates of these trees.

## Results

**Antibody Titers to a Panel of TAAs.** In Fig. 1 we display the absorbances relating antibody titers to TAAs with the various cancer cohorts and normal controls. It is apparent by examining the median titers that antibodies to certain TAAs should prove useful for distinguishing particular cancer cohorts from normal controls: for example, breast, hepatocellular, and gastric cancer patients appear to have substantially higher levels of antibodies to c-myc than normals; and, cyclin B1, IMP1, and Koc appear elevated in virtually all of the cancer cohorts relative to the normal controls. On the other hand, there is substantial overlap in antibody levels between normals and cancer cohorts for other antigen-antibody systems; for example, it appears unlikely that antibodies to survivin by themselves will be useful in distinguishing most cancer patients from normals. Gastric cancer patients seem well differentiated from normals on the basis of five of the seven antibody levels, the exceptions being p62 and survivin. We next addressed the issue of whether there exist optimal subsets of the seven antigen-antibody panel for discriminating the cancer cohorts from normals, and for this, we used the recursive partitioning approach as described above.

**Trees for Discriminating between Cancer Cohorts and Normal Controls.** We first examine two *a priori* decision rules: an individual is considered positive for antibody to a particular antigen if the absorbance of that individual exceeds the mean plus three SDs or the mean plus two SD of the normal absorbances in the corresponding immunoassay. The resulting

**Fig. 2** Classification trees for breast, colorectal, gastric, HCC, lung, and prostate cancer, based on recursive partitioning analysis for positive (+) or negative (-) antibodies to each of seven TAAs. Trees are based on the entire samples of cancer cohorts and normals as depicted in Fig. 1. Cutoff for positivity in each assay is taken as normal mean +3 SDs (A) and normal mean +2 SDs (B). In each tree, decision points or nodes are indicated in *bold* as for example for breast cancer, the initial node uses c-myc for partitioning. The numbers below each node represent the number of normals (initially 346) and the number of cancer cases (with breast cancer, initially 64), which were analyzed there. In breast cancer, the positive branch of the initial node detected 5 normals and 24 breast cancer cases with antibodies to c-myc (A). The remaining subpopulation to the left of the node (341 normals and 40 breast cancer) was additionally partitioned for antibody to Koc, and the negative normals and breast cancer patients were again partitioned for antibodies to cyclin B1. The terminal leaves of the trees (classification sites) are also indicated in *bold*. Note that in the case of breast cancer, 14 patients were positive for anti-Koc and further down the tree, 7 of 26 cases were positive for anti-cyclin B1. In this particular tree, 329 of 346 controls were correctly classified as normal, and similarly 45 of 64 breast cancer patients were correctly classified as cancer. The trees for other cancer cohorts in A and B follow the same principle as that described for breast cancer in A.



pruned trees are displayed in Figs. 2, A and B, and operating characteristics of these trees are summarized in Table 1.

In general, the trees associated with the two cutoff standards within each cancer cohort are consistent in that they entail examination of responses to the same antigens. Interestingly, cyclin B1 appears in every tree and is the first determinant or the initial node for both lung cancer trees as well as both gastric trees, one of the colon trees and one of the HCC trees. Positivity to c-myc is the initial split (determinant) in both breast cancer trees and p62 in both prostate cancer trees. c-myc appears solely in breast and HCC trees; Koc appears in breast, gastric, and lung cancers, and survivin does not appear in any tree. P53 appears

only in the lung (3 SD cutoff) and colon (2 SD cutoff) cancer trees.

Note from Table 1, the clear trade-off between sensitivity and specificity as we vary the cutoffs: sensitivity tends to decrease and specificity tends to increase as the assay cutoffs increase. Over this range of cutoffs, sensitivities are more variable than specificities: that is, there can be dramatic differences in sensitivities (for example colon from 0.60 to 0.84 or prostate from 0.59 to 0.77) depending on the cutoff chosen for assay positivity.

Fig. 2, A and B, and Table 1 also reveal some interesting relationships obtained with the recursive partitioning approach.

Table 1 Operating characteristics of the classification trees depicted in Fig. 2, using normal means + 3 SD or normal means + 2 SDs as cutoffs for positivity

Cancer	3 SD cutoff		Cross-validated values		2 SD cutoff		Cross-validated values	
	True pos	True neg	True pos	True neg	True pos	True neg	True pos	True neg
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Breast	0.70	0.95	0.67	0.94	0.92	0.85	0.88	0.87
Colon	0.60	0.95	0.50	0.94	0.84	0.90	0.63	0.92
Gastric	0.69	0.94	0.70	0.91	0.89	0.87	0.83	0.89
HCC	0.55	0.97	0.55	0.91	0.80	0.91	0.78	0.85
Lung	0.77	0.93	0.67	0.95	0.80	0.90	0.77	0.89
Prostate	0.59	0.96	0.52	0.96	0.77	0.87	0.64	0.89

For discrimination between breast cancer and normal controls, the use of *c-myc*, *Koc*, and *cyclin B1* of the seven TAAs is sufficient to reach fairly good sensitivity and specificity levels, from 0.70 to 0.92 for sensitivity and from 0.85 to 0.95 for specificity. Different cancer cohorts approach these levels of sensitivity and specificity but with the use of different TAAs. For colon cancer, this is *IMP1*, *cyclin B1*, and *p53*; for gastric cancer, HCC, and lung cancer, the relevant antigens are *cyclin B1*, *IMP1*, and *Koc*, and for prostate cancer, they are *p62*, *cyclin B1*, and *IMP1*. It needs to be mentioned at this point that the sensitivities and specificities obtained have to be considered in the context of the seven TAAs used for analysis. Although these seven TAAs have all been reported to be targets of autoantibodies in cancer (4–8, 14–16), they might not be the optimal candidate autoantigens for particular types of tumors. It is likely that with some cancer cohorts, other TAAs that we have not tested may result in yet higher degrees of sensitivity and specificity.

**Optimal Discrimination between Cancer Cohorts and Normals.** Although performance with fixed cutoffs is remarkably good, we investigated whether we could develop even better classification trees by allowing the cutoffs to vary among the antibodies. The resulting trees are displayed in Fig. 3, and operating characteristics of these trees are given in Table 2. We note that we could outperform the fixed cutoffs, with no increased complexity in the trees (with the exception of prostate cancer). Note also that the absorbance trees generally involve the same antibodies as the fixed cutoff trees (although *cyclin B1* is no longer so prominent a player). A more valid assessment of the operating characteristics of the optimal trees is obtained from 10-fold cross-validation; we find that the cross-validated estimates of sensitivity drop by ~0.10 to 0.18, and specificity 0.05 to 0.13, relative to the retrospective trees.

## Discussion

We have found that classification trees based on the panel of seven TAAs can discriminate between cancer cohorts and normals with reasonably high sensitivity and specificity, both typically exceeding 0.80. The multivariate approach under which the joint distribution of antigen profiles is considered yields far higher values of sensitivity and specificity than would be obtained had we solely looked at individual TAAs in isolation (17).

Choice of cutoffs for positivity in the underlying immuno-

assays can dramatically affect performance of the classification trees. For example, selection of the normal mean + 3 SDs for positivity in each immunoassay seems overly stringent, and leads to unacceptably high false-negative rates (ranging from 0.23 to 0.41). When selection of the cutoff for positivity used the normal mean + 2 SDs, the false-negative rates were reduced to a range of 0.08–0.23 (Table 1). We have also analyzed these sets of data using a cutoff of 1 SD of the normal mean. Although the false-negative rates were lower, the false-positive rates were higher, ranging from 0.13 to 0.21. The false-positive rates at this cutoff were considered too high to be of value when compared with the false positive rates using mean plus either 2 SD or 3 SD cutoffs (Table 1).

Recursive partitioning is a well-established statistical methodology, and its application to classification problems in oncology first appeared some 20 years ago (18). There have been several recent studies reporting the use of recursive partitioning to address different issues related to cancer, including the use of DNA gene microarrays to analyze expression data from 2000 genes in normal and colon cancer (19), generalized recently (20). In addition, recursive partitioning has been used by the Radiation Therapy Oncology Groups to determine subgroups in head and neck tumors having unique outcomes (21) and identification of prognostic subgroups in patients with non-small cell lung cancer (22). Classification and regression tree analysis has been used in the study of unknown primary carcinoma to estimate the survival probability of an individual patient (23) and also in the analysis of recurrence in breast cancer after radiation and chemotherapy (24, 25). A particular advantage of recursive partitioning for classification relative to other methodologies we might have used (*e.g.*, discriminant analysis and logistic regression) is that the trees are easy to interpret and often capture much of the relevant covariate structure of the data, including complex interactions and nonlinearities. The very low error rates of the classification trees are reassuring, if not surprising. Of greater import is that the error rates under prospective evaluation (10-fold cross-validation) remain rather small, suggesting that this methodology may well be useful for cancer diagnosis. In this regard, we must emphasize that positive and negative predictive values will critically influence accuracy of the trees in any clinical setting; that is, the effect of prevalence on the predictive values of the trees will be pronounced (26). On the other hand, the classification trees can be

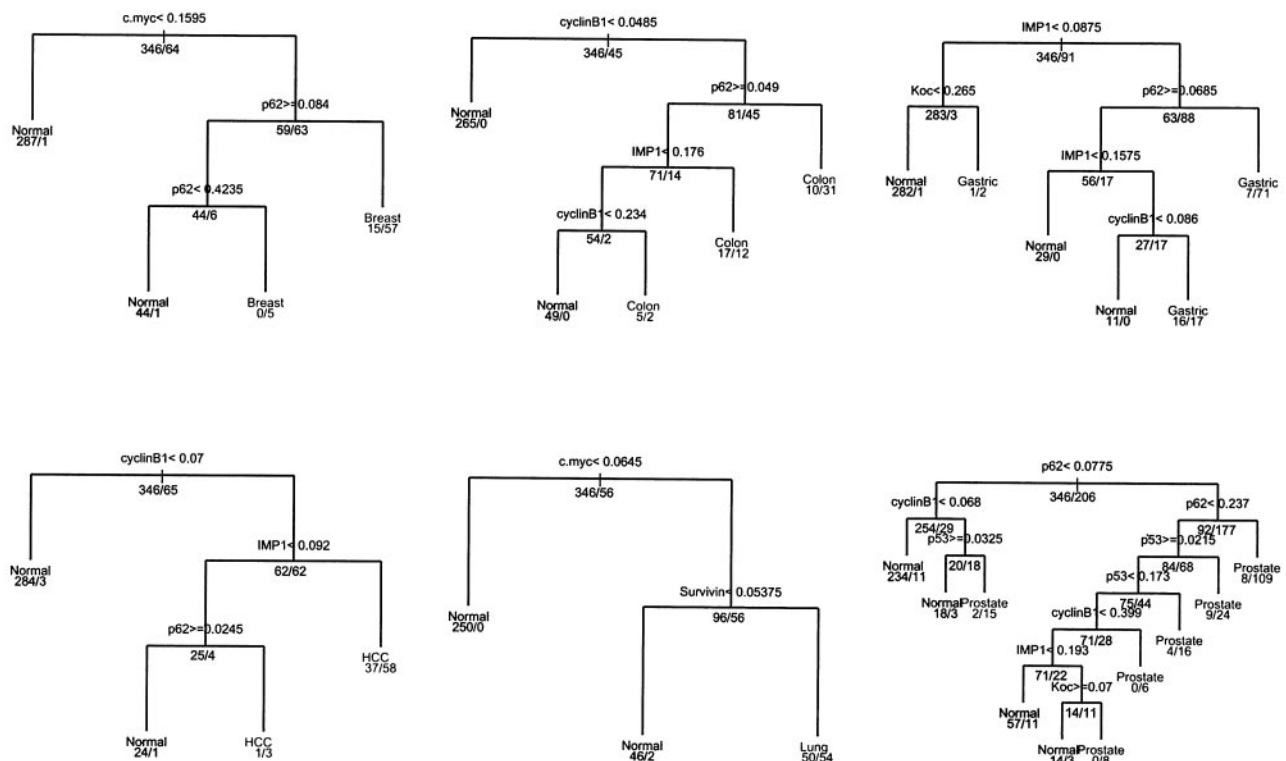


Fig. 3 Classification trees for breast, colorectal, gastric, HCC, lung, and prostate cancer, based on the continuous coding of absorbances to each of the 7 TAAs for each individual. Here, decision points or nodes are again indicated in *bold*, but are based on actual immunoassay absorbance values of antibody response. If the decision criterion at a node is satisfied, one proceeds to the left, and if it is not satisfied, one proceeds to the right. The numbers of subjects at each node, and the classification at terminal leaves of each tree are as in Fig. 2.

Table 2 Operating characteristics of the classification trees depicted in Fig. 3, using optimal cutoffs in absorbances for positivity in the immunoassays

Cancer	Absorbance cutoff		Cross-validated values	
	True pos	True neg	True pos	True neg
	Sensitivity	Specificity	Sensitivity	Specificity
Breast	0.97	0.96	0.90	0.93
Colon	1.00	0.91	0.88	0.88
Gastric	0.99	0.93	0.81	0.93
HCC	0.94	0.89	0.93	0.84
Lung	0.96	0.86	1.00	0.84
Prostate	0.86	0.93	0.79	0.86

tuned to reflect different prior probabilities, or costs of misclassification, of the normal and cancer categories.

The TAAs studied here include oncogene product *c-myc* and tumor suppressor gene product *p53*. Cyclin B1 has been reported to be antigenic in humoral and cell-mediated immune responses in HCC and epithelial cell tumors, respectively (8, 15). *p62* (5), *Koc* (6), and *IMP1* (7) are three proteins containing similar RNA-binding motifs, and their mRNAs are either overexpressed in cancer (6, 14) or antibodies have been identified to the protein products (4, 5). *Survivin* is a cellular protein that is antiapoptotic (27), and antibodies have been found in cancer sera (16). Full-length recombinant proteins rather than peptide

fragments of these proteins were used as antigens in immunoassay so that theoretically all of the antigenic determinants of each protein were available and sera from other autoimmune disorders were examined to establish that antibody reactivities observed in the present study were cancer-specific and not because of nonspecific immunoreactivity (17).

This initial study was based on seven selected TAAs. In this regard, we could expect that incorporation of additional antigens or of other antigens replacing those in this seven TAA panel might additionally improve the operating characteristics of the screens for the various types of cancer. There is support at the genetic level for this notion. Gene abnormalities such as mutations or overexpression tend to occur in combinations that vary from tissue to tissue (28), and if autoimmune responses represent immune system reactions to abnormal gene expression (3), one might then expect differences in autoantibody profiles from one type of tissue cancer to another. Panels or arrays of TAAs could be expanded to include, among others, several other known TAAs such as *HER-2/neu* and *ras*, cell-cycle proteins besides cyclin B1 and antigens involved in cell-mediated immune reactions, many of which might have concomitant humoral immune responses (29, 30).

## Acknowledgments

We thank Drs. Israel Gutierrez and Herbert Ruckle of Loma Linda University Medical Centers for the prostate cancer sera. We also thank

Xiao Wang for help in many laboratory experiments. This is publication 14818-MEM from The Scripps Research Institute.

## References

- Crawford, L. V., Pim, D. C., and Bulbrook, R. D. Detection of antibodies against the cellular protein p53 in sera from patients with breast cancer. *Int. J. Cancer*, *30*: 403–408, 1982.
- Soussi, T. p53 antibodies in the sera of patients with various types of cancer. A review. *Cancer Res.*, *60*: 1777–1788, 2000.
- Tan, E. M. Autoantibodies as reporters identifying aberrant cellular mechanisms in tumorigenesis. *J. Clin. Investig.*, *108*: 1411–1415, 2001.
- Zhang, J. Y., Chan, E. K., Peng, X. X., Lu, M., Wang, X., Mueller, F., and Tan, E. M. Autoimmune responses to mRNA binding proteins p62 and Koc in diverse malignancies. *Clin. Immunol.*, *100*: 149–156, 2001.
- Zhang, J. Y., Chan, E. K. L., Peng, X. X., and Tan, E. M. A novel cytoplasmic protein with RNA-binding motifs is an autoantigen in human hepatocellular carcinoma. *J. Exp. Med.*, *189*: 1101–1110, 1999.
- Mueller-Pillasch, F., Lacher, U., Wallrapp, C., Micha, A., Zimmerhackl, F., Hameister, H., Varga, G., Friess, H., Buchler, M., Beger, H. G., Vila, M. R., Adler, G., and Gress, T. M. Cloning of a gene highly overexpressed in cancer coding for a novel KH-domain containing protein. *Oncogene*, *14*: 2729–2733, 1997.
- Nielsen, J., Christiansen, J., Lykke-Andersen, J., Johnsen, A. H., Wewer, U. M., and Nielsen, F. C. A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol. Cell. Biol.*, *19*: 1262–1270, 1999.
- Covini, G., Chan, E. K., Nishioka, M., Morshed, S. A., Reed, S. I., and Tan, E. M. Immune response to cyclin B1 in hepatocellular carcinoma. *Hepatology*, *25*: 75–80, 1997.
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. *Classification and Regression Trees*. Wadsworth, Monterey, CA, 1984.
- Province, M. A., Shannon, W. D., and Rao, D. C. Classification methods for confronting heterogeneity. *Adv. Genetics*, *42*: 273–286, 2001.
- Zhang, H., and Singer, B. *Recursive Partitioning in the Health Sciences*. Springer, New York, 1999.
- Atkinson, E. J., and Therneau, T. M. An introduction to recursive partitioning using the RPART routines. Technical report, Mayo Clinic Section of Biostatistics, 2000.
- Venables, W. N., and Ripley, B. D. *Modern Applied Statistics with S-PLUS*, 3rd Ed., pp. 303–327. Springer, New York, 1999.
- Doyle, G. A., Bourdeaux-Heller, J. M., Coulthard, S., Meisner, L. F., and Ross, J. Amplification in human breast cancer of a gene encoding a c-myc mRNA binding protein. *Cancer Res.*, *60*: 2756–2759, 2000.
- Kao, H., Marto, J. A., Hoffmann, T. K., Shabanowitz, J., Finkelshtein, S. D., Whiteside, T. L., Hunt, D. F., and Finn, O. J. Identification of cyclin B1 as a shared human epithelial tumor-associated antigen recognized by T cells. *J. Exp. Med.*, *194*: 1313–1323, 2001.
- Rohayem, J., Diestelkoetter, P., Weigle, B., Oehmichen, A., Schmitz, M., Mehlhorn, J., Conrad, K., and Rieber, E. P. Antibody to the tumor-associated inhibitor of apoptosis protein survivin in cancer patients. *Cancer Res.*, *60*: 1815–1817, 2000.
- Zhang, J. Y., Casiano, C. A., Peng, X. X., Koziol, J. A., Chan, E. K. L., and Tan, E. M. Enhancement of antibody detection in cancer using panel of recombinant tumor-associated antigens. *Cancer Epidemiol. Biomark. Prev.*, *12*: 136–143, 2003.
- Dillman, R. O., and Koziol, J. A. Statistical approach to immunosuppression classification using lymphocyte surface markers and functional assays. *Cancer Res.*, *43*: 417–421, 1983.
- Zhang, H., Yu, C. Y., Singer, B., and Xiong, M. Recursive partitioning for tumor classification with gene expression microarray data. *Proc. Natl. Acad. Sci. USA*, *98*: 6730–6735, 2001.
- Zhang, H., Yu, C. Y., and Singer, B. Cell and tumor classification using gene expression data: Construction of forests. *Proc. Natl. Acad. Sci. USA*, *100*: 4168–4172, 2003.
- Cooper, J. S., Berkey, B., Marcial, V., Fu, K. F., and Lee, D. J. Validation of the RTOG recursive partitioning for head and neck tumors. *Head Neck*, *23*: 669–677, 2001.
- Werner-Wasik, M., Scott, C., Cox, J. D., Sause, W. T., Byhardt, R. W., Asbell, S., Russel, A., Komaki, R., and Lee, J. S. Recursive partitioning analysis of 1999 Radiation Therapy Oncology Group (RTOG) patients with locally advanced non-small-cell lung cancer (LA-NSCLC): identification of five groups with different survival. *Int. J. Radiation Oncology Biol. Phys.*, *48*: 1475–1482, 2000.
- Hess, K. R., Abbruzzese, M. C., Lenzi, R., Raber, M. N., and Abbruzzese, J. L. Classification and regression tree analysis of 1000 consecutive patients with unknown primary carcinoma. *Clin. Cancer Res.*, *5*: 3403–3410, 1999.
- Katz, A., Buchholz, T. A., Thomas, H., Smith, C. D., McNeese, M. D., Theriault, R., Singletary, S. E., and Strom, E. A. Recursive partitioning analysis of locoregional recurrence patterns following mastectomy: implications for adjuvant irradiation. *Int. J. Radiat. Oncol. Biol. Phys.*, *50*: 397–403, 2001.
- Freedman, G. M., Hanlan, A. L., Fawble, B. L., Anderson, P. R., and Nicolau, N. Recursive partitioning identifies patients at high risk and low risk for ipsilateral tumor recurrence after breast conserving surgery and radiation. *J. Clin. Oncology*, *20*: 4014–4021, 2002.
- Sackett, D. L., Haynes, R. B., Guyatt, G. H., and Tugwell, P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Second Edition. Little, Brown and Company, Boston, 1991.
- Ambrosini, G., Adida, C., and Altieri, D. C. A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma. *Nat. Med.*, *3*: 917–922, 1997.
- Bartek, J., and Lukas, J. Are all cancer genes equal? *Nature (Lond.)*, *411*: 1001–1002, 2001.
- Boon, T., and van der Bruggen, P. Human tumor antigens recognized by T lymphocytes. *J. Exp. Med.*, *183*: 725–729, 1996.
- Stockert, E., Jager, E., Chen, Y. T., Scanlan, M. J., Gout, I., Karbach, J., Arand, M., Knuth, A., and Old, L. J. A survey of humoral immune response of cancer patients to a panel of human tumor antigens. *J. Exp. Med.*, *187*: 1349–1354, 1998.