

Autoantibody Approach for Serum-Based Detection of Head and Neck Cancer

Ho-Sheng Lin,^{1,2} Harvinder S. Talwar,² Adi L. Tarca,^{3,4} Alexei Ionan,⁵ Madhumita Chatterjee,⁵ Bin Ye,⁵ Jerzy Wojciechowski,⁵ Saroj Mohapatra,⁵ Marc D. Basson,^{1,6} George H. Yoo,² Brian Peshek,² Fulvio Lonardo,⁷ Chuan-Ju G. Pan,² Adam J. Folbe,² Sorin Draghici,^{3,4} Judith Abrams,⁸ and Michael A. Tainsky^{5,7}

¹Department of Surgery, John D. Dingell VA Medical Center; ²Department of Otolaryngology-Head and Neck Surgery, Wayne State University and Karmanos Cancer Institute; ³Bioinformatics Core, Barbara Ann Karmanos Cancer Institute; ⁴Department of Computer Science, Wayne State University; ⁵Program in Molecular Biology and Genetics, Karmanos Cancer Institute/Wayne State University; ⁶Departments of ⁷Surgery and ⁸Pathology, Wayne State University; ⁹Integrated Biostatistics Core, Barbara Ann Karmanos Cancer Institute and Wayne State University, Detroit, Michigan

Abstract

Currently, no effective tool exists for screening or early diagnosis of head and neck squamous cell carcinoma (HNSCC). Here, we describe an approach for cancer detection based on analysis of patterns of serum immunoreactivity against a panel of biomarkers selected using microarray-based serologic profiling and specialized bioinformatics. We biopanned phage display libraries derived from three different HNSCC tissues to generate 5,133 selectively cloned tumor antigens. Based on their differential immunoreactivity on protein microarrays against serum immunoglobulins from 39 cancer and 41 control patients, we reduced the number of clones to 1,021. The performance of a neural network model (Multilayer Perceptron) for cancer classification on a data set of 80 HNSCC and 78 control samples was assessed using 10-fold cross-validation repeated 100 times. A panel of 130 clones

was found to be adequate for building a classifier with sufficient sensitivity and specificity. Using these 130 markers on a completely new and independent set of 80 samples, an accuracy of 84.9% with sensitivity of 79.8% and specificity of 90.1% was achieved. Similar performance was achieved by reshuffling of the data set and by using other classification models. The performance of this classification approach represents a significant improvement over current diagnostic accuracy (sensitivity of 37% to 46% and specificity of 24%) in the primary care setting. The results shown here are promising and show the potential use of this approach toward eventual development of diagnostic assay with sufficient sensitivity and specificity suitable for detection of early-stage HNSCC in high-risk populations. (Cancer Epidemiol Biomarkers Prev 2007;16(11):2396–405)

Introduction

The American Cancer Society estimates that ~45,660 new cases of head and neck squamous cell carcinoma (HNSCC) will be diagnosed in the United States and 11,210 Americans will die from this disease in the year 2007 (1). Worldwide, HNSCC is the sixth most common malignancy with incidence of 644,000 new cases a year (2). Despite progress in diagnostic and treatment modalities in the past 30 years, long-term survival for patients affected by HNSCC has not significantly improved (3). In the most recent issue of *Cancer Facts &*

Figures - 2007, improvement in 5-year relative survival rates between 1975 and 2002 was reported for almost all types of cancer, with the only two exceptions being laryngeal and uterine cancers (4). One major impediment to improving survival in this patient population is the failure to detect this cancer at an early stage. More than two thirds of patients with HNSCC are diagnosed at an advanced stage when the 5-year survival is <40% (5). In many cases, these patients are offered radical treatments, which often result in significant physical disfigurement as well as dysfunction of speech, breathing, and swallowing. The plight of these patients with advanced-stage disease is in distinct contrast to that of patients who are diagnosed early. Early-stage HNSCC patients have an excellent 5-year survival rate of >80% and experience significantly less effect on their quality of life after treatment with single modality therapy (5). This dramatic difference in survival and quality of life underlies the importance of early detection in this disease.

Early detection can be achieved by screening asymptomatic patients at high risk for development of cancer. Although the American Cancer Society has issued guidelines for screening of breast, colon, prostate, and uterine cancers (6), no such guideline exists for HNSCC.

Received 4/10/07; revised 8/18/07; accepted 9/13/07.

Grant support: VA Merit Review Entry Program grant and Binn's fund from the Department of Otolaryngology-Head and Neck Surgery at Wayne State University (H-S. Lin); National Science Foundation grants DBI-0234806, CCF-0438970, 1R21 EB00990-01, and 1R01 NS045207-01 and NIH grant 1R01HG003491 (S. Draghici); and The Barbara and Fred Erb Endowed Chair in Cancer Genetics, The Michigan Life Science Corridor Fund (085P300470), and NIH grants R33-CA100740-03 and U01-CA117478-01 (M.A. Tainsky).

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.

Requests for reprints: Ho-Sheng Lin, Department of Otolaryngology/Head and Neck Surgery, Wayne State University, 5 East University Health Center, 4201 St. Antoine, Detroit, MI 48201. Phone: 313-577-0805; Fax: 313-577-8555. E-mail: hlin@med.wayne.edu

Copyright © 2007 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-0318

This is especially unfortunate given that patients at increased risk for development of HNSCC can be easily identified (excess alcohol and/or tobacco use and history of prior HNSCC) (7, 8) and targeted for screening. Early detection can also be improved by reducing diagnostic delays, reported to be between 3 (9, 10) and 6 months (11), in the primary care setting. It is estimated that, for every 1 week of delay, the stage of presentation will progress by 0.045 of a stage (12). Thus, a delay of several months may prove detrimental by decreasing a patient's chance of survival from 80% to 40%. Misdiagnosis at initial presentation to primary care physicians is common and may be due to the nonspecific nature of presenting symptoms, technical difficulty of examination in the head and neck region, as well as the rarity of this type of cancer (12). In one large prospective study involving patients who presented with hoarseness to 11 general practices, the performance of primary physicians on cancer diagnosis was poor with sensitivity of 46% and specificity of 24% (13). This low sensitivity was confirmed by many other studies showing a correct diagnosis of only 37% to 38% at the initial visit (9, 14, 15). The delay in diagnosis and referral to specialist has a significant negative effect on patient outcome and survival (9, 16, 17). Thus, there exists a need for a simple, noninvasive, and inexpensive test, widely accessible to physicians in the primary care setting, which can be used to screen (in asymptomatic patients) and diagnose (in symptomatic patients) HNSCC in high-risk population to improve early detection.

Autoantibodies against cancer-specific antigens have been identified in cancers of the colon (18), breast (19), lung (20), ovary (21), prostate (22), and head and neck (23). Immune response with antibody production may be elicited due to the overexpression of cellular proteins such as Her2 (24), the expression of mutated forms of cellular protein such as mutated p53 (25), or the aberrant expression of tissue-restricted gene products such as cancer-testis antigens (26) by cancer cells. Because these autoantibodies are raised against these specific antigens from the cancer cells, the detection of these antibodies in patients' sera can be exploited for cancer diagnosis in these patients. Further, the immune system is especially well adapted for the early detection of cancer because it can respond to even low levels of an antigen by mounting a very specific and sensitive antibody response. Thus, the use of immune response as a biosensor for early detection of cancer through serum-based assay holds great potential as an ideal screening and diagnostic tool.

In this study, we adopted an approach previously described in our laboratory on ovarian cancer (21, 27, 28) that combines phage display technology, protein microarrays, and bioinformatics tools to select and profile a panel of biomarkers for cancer diagnosis. We have coined the term "epitomics" to describe this global profiling of the immune response to antigens (29). We showed a high degree of diagnostic accuracy for cancer detection based on analysis of patterns of serum antibody immunoreactivity against a panel of cancer antigens. The essential features of our approach included the use of very specific antibody-antigen interaction, departure from reliance on any single biomarker for

cancer detection, unbiased selection of cloned cancer antigen, use of high-throughput technology, and specialized bioinformatics techniques for feature selection and classifier development.

Materials and Methods

Serum Samples. Blood samples from HNSCC patients (stages I-IV) and controls were obtained after informed consent. Both HNSCC and control patients were recruited from the otolaryngology/head and neck surgery clinic population. All enrolled HNSCC patients have cancer confirmed on pathology. Control patients underwent thorough head and neck examination and/or imaging to rule out the presence of cancer after they initially presented with nonspecific head and neck symptoms, such as sore throat, hoarseness, dysphagia, coughing, neck mass, otalgia, choking, and foreign body sensation. Ten milliliters of peripheral blood were collected into red top vacutainers without anticoagulant using standardized phlebotomy procedures. These samples were allowed to clot for 20 to 30 min at room temperature and then centrifuged at 2,500 rpm at 4°C for 15 min. The supernatants or sera were immediately aliquoted and stored in a -70°C freezer. This study was conducted under protocols approved by the Wayne State University Human Investigation Committee (HIC #21802MP4E).

Construction of T7 Phage Display cDNA Libraries. HNSCC specimens were obtained at the time of surgical extirpation and immediately placed in RNAlater solution (Ambion). Total RNA extraction was done using Trizol reagent (Invitrogen Corp.). After extraction, polyadenylated RNAs were purified twice using Straight A's mRNA Isolation System (EMD Biosciences-Novagen) per protocols from the manufacturer. The construction of T7 phage cDNA display libraries was done using Novagen OrientExpress cDNA Synthesis (Random Primer System) and Cloning System according to the protocol by EMD Biosciences-Novagen. The number of clones in each of the three libraries was titered by plaque assay per manufacturer's protocol (EMD Biosciences-Novagen).

Differential Biopanning of HNSCC Phage Display cDNA Libraries. The three phage cDNA display libraries were pooled together and the combined library was then biopanned separately against each of the 12 HNSCC sera. Differential biopannings using sera from control and HNSCC patients were done as per manufacturer's protocol (T7Select System, TB178; EMD Biosciences-Novagen). Protein G Plus-agarose beads (Santa Cruz Biotechnology, Inc.) were used for serum immunoglobulin (IgG) immobilization. Three to five rounds of biopanning were done using serum from each of the 12 HNSCC patients. Each cycle of biopanning consisted of passing the entire phage library through protein G beads coated with IgGs from pooled sera of healthy controls, passage through beads coated with IgGs from individual serum from each of the 12 HNSCC patients, followed by final elution of bound phage clones from the beads.

Protein Microarray Immunoreaction. Individual clones were picked and arrayed in replicates of five or six onto FAST slides (Schleicher & Schuell/Whatman)

using a robotic microarrayer ProSys 5510TL (Cartesian Technologies) with 32 Micro-Spotting Pins (TeleChem). Protein microarrays were blocked with 4% milk in $1 \times$ PBS for 1 h at room temperature followed by another hour of incubation with primary antibodies consisting of human serum at a dilution of 1:300 in PBS, mouse anti-T7 capsid antibodies (0.15 $\mu\text{g}/\text{mL}$; EMD Biosciences), and BL21 *Escherichia coli* cell lysates (5 $\mu\text{g}/\text{mL}$). The microarrays were then washed thrice in PBS/0.1% Tween 20 solution 4 min each at room temperature and then incubated with Alexa Fluor 647 (red fluorescent dye)-labeled goat anti-human IgG antibodies (1 $\mu\text{g}/\text{mL}$) and Alexa Fluor 532 (green fluorescent dye)-labeled goat anti-mouse IgG antibodies (0.05 $\mu\text{g}/\text{mL}$; Molecular Probes) for 1 h in the dark. Finally, the microarrays were washed thrice in PBS/0.1% Tween 20 for 4 min each and then twice in PBS for 2 min each and air dried.

Data Acquisition and Preprocessing. Following immunoreaction, the microarrays were scanned using the GenePix 4100A scanner (Axon Laboratories) using 635 and 532 nm lasers to produce a red (Alexa Fluor 647) and green (Alexa Fluor 532) composite image. Using the ImaGene 6.0 (BioDiscovery, Inc.) image analysis software, the binding of each of the cancer-specific peptides with IgGs in each serum was then analyzed and expressed as a ratio of red-to-green fluorescent intensities. The microarray data were further read into the R environment v2.3.0 (30) and processed by a sequence of transformations, including background correction, omission of poor quality spots, base 2 log transformation, loess-based global normalization, and combining spot replicates into a single value for each marker. Specialized bioconductor⁹ libraries, such as limma (31), were used to this end.

Data Analysis. Before building and assessing the performance of neural network model for cancer classification, HNSCC and control serum samples were split at random into 10 equal groups (folds), each having about the same number of HNSCC and controls. Then, iteratively, nine tenth ($n = 142$) of the data set was used to select clones (features) and train a model, whereas the independent one tenth ($n = 16$) of the data set was used to test the resulting model. A complete pass through the 10-fold ensures obtaining reliable performance indices because each sample can be either in the training set or in the test set but never in both. To obtain even more robust estimate, the 10-fold cross-validation strategy was repeated 100 times each time splitting the data differently into 10 partitions. Thus, a total of 1,000 feature selections with 1,000 training and 1,000 independent testing sessions were done.

Each of the feature selection process based on the training data (142 samples) included several steps. First, clones that immunoreacted, on average, less with sera from cancer patients than controls were discarded. The remaining clones were then ranked using the P value from a t test and the top 250 retained. To assess the potential of each of these 250 clones to discriminate cancer from noncancer sera in the training set, each clone

was used individually to derive a receiver-operating characteristic curve and then ranked in decreasing order of area under the receiver-operating characteristic curve (AUC).

Once the features were ranked, we built three-layered feed-forward neural network models (Multilayer Perceptron Classifiers) using as predictors the top-ranked clones. The number of clones used varied from 10 to 180 in increments of 10. Each resulting model was applied on the independent test set and the average performance indices were computed over the 1,000 independent trials. The nnet package (32) under the R environment v2.3.0 (30) was used for model training. Special attention was paid to avoid data overfitting by using a reduced number of hidden nodes ($n = 5$) as well as using a training method (Broyden-Fletcher-Goldfarb-Shanno), which included regularization, as implemented in the nnet package (32).

After the final panel of biomarkers was determined, a completely new data set of 80 samples was used to validate its performance. Multilayer Perceptron Classifier, based on this final panel of biomarkers, was then trained and tested on these new data sets. Before building and testing of the model, the new data set was randomly split into a training set of 52 samples (26 HNSCC and 26 controls) and a test set of 28 samples (14 HNSCC and 14 controls).

Sequencing of Phage cDNA Clones. Individual phage clones were PCR amplified using forward primer 5'-GTTCTATCCGCAACGTTATGG-3' and reverse primer 5'-GGAGGAAAGTCGTTTTTTGGGG-3' and sequenced using forward primer by Wayne State University Sequencing Core Facility.

Results

Differential Biopanning Results in Enrichment of T7 Phage HNSCC cDNA Display Libraries. We constructed T7 phage cDNA display libraries from three HNSCC specimens (floor of mouth, base of tongue, and larynx). The insertion of foreign HNSCC cDNAs into the T7 phage capsid genes results in the production of fusion capsid proteins. Foreign peptides displayed in this fashion have been shown to fold in their native conformations (33), thus exposing both linear and conformational antigens on the surface of the bacteriophage where they are accessible for selection and analysis (34). A potential limitation of the T7 phage display system, however, is the absence of posttranslational modifications, such as glycosylation, sulfation, methylation, and phosphorylation, which may influence the folding and binding of these peptides (35). Each of these three cDNA phage libraries was titered by plaque assay and found to contain between 10^6 and 10^7 primary recombinants. Because the majority of the clones in the HNSCC cDNA libraries carried normal self-proteins, differential biopanning was done to enrich the cDNA libraries with clones expressing the HNSCC-specific antigenic peptides (Fig. 1). This technique relied on specific antigen-antibody reactions to remove clones that bind to IgGs from control sera while retaining clones with peptides of interest (HNSCC-specific antigens) using antibodies in HNSCC sera as bait. To increase the diversity of HNSCC-specific peptides, the three cDNA

⁹ <http://www.bioconductor.org>

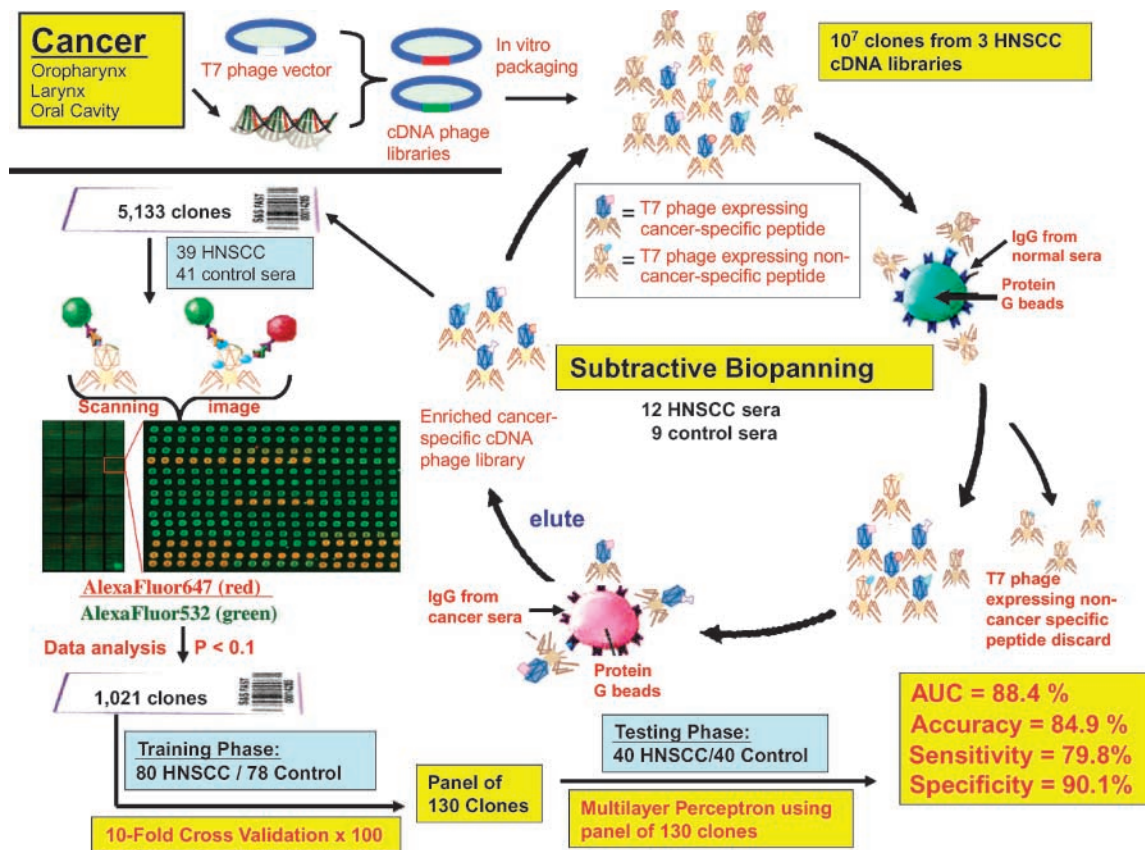


Figure 1. Schema showing the process of combining phage display technology, protein microarrays, and bioinformatics tools to profile and select a panel of 130 clones from 10^7 initial clones in the three HNSCC cDNA phage display libraries. First, three cDNA libraries were constructed from HNSCC specimens. Because the majority of the clones in the HNSCC cDNA libraries carried normal self-proteins, subtractive biopanning was done to enrich the cDNA libraries with clones expressing the HNSCC-specific peptides. This technique relied on specific antigen-antibody reactions to remove clones that bind to IgGs from control sera while retaining clones with peptides of interest (HNSCC-specific antigens) using antibodies in HNSCC sera as bait. Protein G Plus-agarose beads were used for serum IgG immobilization. Three to five rounds of biopanning were done using serum from each of the 12 HNSCC patients. Each cycle of biopanning consisted of passing the entire phage library through protein G beads coated with IgGs from pooled sera of healthy controls, passage through beads coated with IgGs from individual serum from HNSCC patients, followed by final elution of bound phage clones from the column. Following biopanning, a total of 5,133 clones were randomly chosen from the 12 highly enriched pools of T7 phage cDNA libraries. These clones were arrayed and immunoreacted against serum samples from 39 HNSCC and 41 control patients. Phage clones were spotted in replication of five to six in an ordered array onto FAST nitrocellulose-coated glass slides. The binding of arrayed HNSCC-specific peptides with antibodies in sera was quantified with the Alexa Fluor 647 (red fluorescent dye)-labeled anti-human antibody. The use of mouse anti-T7 capsid antibodies, detected with the use of Alexa Fluor 532 (green fluorescent dye)-labeled goat anti-mouse IgG antibody, was necessary to normalize for any small variation in the amount of phage particles spotted throughout the microarray chip. Following immunoreaction, the microarray data were analyzed and processed by a sequence of transformations. To reduce the number of clones for further analysis, one-tailed *t* test was used to select 1,021 clones (from the original 5,133 clones) with increased reactivity to cancer sera compared with control sera ($P < 0.1$). Sera from 80 cancer and 78 noncancer controls, not previously used for biopanning or selection of clones, were immunoreacted against the previously selected 1,021 HNSCC-specific peptides. The reactivity of each of the 1,021 cancer-specific peptides with each of these 158 sera was then analyzed. A 10-fold cross-validation procedure was used to assess the performance of a neural network model (Multilayer Perceptron) to classify the patients based on patterns of serum immunoreactivity against a panel of biomarkers. Both the clone (feature) selection and model training were based solely on nine tenths of the data set and the model was tested on the remaining (completely independent) one tenth of the data set. The entire 10-fold cross-validation was repeated 100 times to minimize any potential bias due to random partition of training and test sets. For each of the 1,000 splits, a slightly different panel of 130 biomarkers was selected from each of the 1,000 randomly generated training sets. To establish a final set of 130 markers that can be used for further studies, we ranked the top clones based on the number of times out of the 1,000 that they were selected as one of the top 130 biomarkers used in the panel to build the classifier for distinguishing cancer from noncancer sera. This panel of 130 markers was then immunoreacted against an independent validation set of sera from 40 HNSCC and 40 controls. These 80 sera samples were not used previously for biopanning, selection of clones, or building and training of the classifiers. Multilayer Perceptron Classifier, based on this final panel of 130 biomarkers, was then trained and tested on these new data set, achieving an AUC of 88.4% and accuracy of 84.9% with sensitivity of 79.8% and specificity of 90.1%.

libraries were pooled and biopanned against individual serum from 12 HNSCC patients with tumors representing different subsites of head and neck (Supplementary Table S1), producing 12 antigen-enriched T7 phage cDNA libraries.

High-Throughput Protein Microarray Immunoscreeing for Selection of Informative HNSCC-Specific Biomarkers. A total of 5,133 clones were randomly selected from the 12 highly enriched pools of T7 phage cDNA libraries (Fig. 1). These clones were arrayed and immunoreacted against serum samples from 39 HNSCC and 41 controls (Supplementary Table S2). The binding of each of the arrayed HNSCC-specific peptides with antibodies in sera was quantified with the Alexa Fluor 647 (red-fluorescent dye)-labeled anti-human antibody. The amount of phage particles at each spot throughout the microarray was detected using mouse monoclonal antibody to the T7 capsid protein and quantified using Alexa Fluor 532 (green fluorescent dye)-labeled goat anti-mouse antibody. To correct for any small variation in the amount of antibody binding in each spot that may be due to different amounts of phage particles spotted on the microarray, the ratio of the intensity of Alexa Fluor 647 over Alexa Fluor 532 was calculated for each spot.

Following immunoreaction, the microarray data were processed by a sequence of transformations and then analyzed. The interassay reproducibility of the immunoreaction assay was assessed by comparing the results of immunoreactivity among chips printed and immunoreacted with the same serum samples at different times (Fig. 2). Further, the intraassay reproducibility was assessed by comparing the results among the six replicates printed within the same chip for each clone. The calculated interassay coefficient of variance was 16% and intraassay coefficient of variance was 7% (Supplementary Fig. S1). To reduce the number of clones for further analysis, one-tailed *t* tests under the R environment v2.3.0 (30) were used to select clones with increased binding to IgGs present in cancer sera compared with control sera using the criterion of $P < 0.10$; 1,021 clones met the criterion.

Selection of a Panel of Biomarkers and Estimation of Neural Network Classifier Performance (Training Phase). Sera from 80 HNSCC patients and 78 controls, not previously used for biopanning or selection of clones, were immunoreacted against the previously selected 1,021 HNSCC-specific peptides. Of the 80 HNSCC patients, 18 had early-stage disease (I and II) and 62

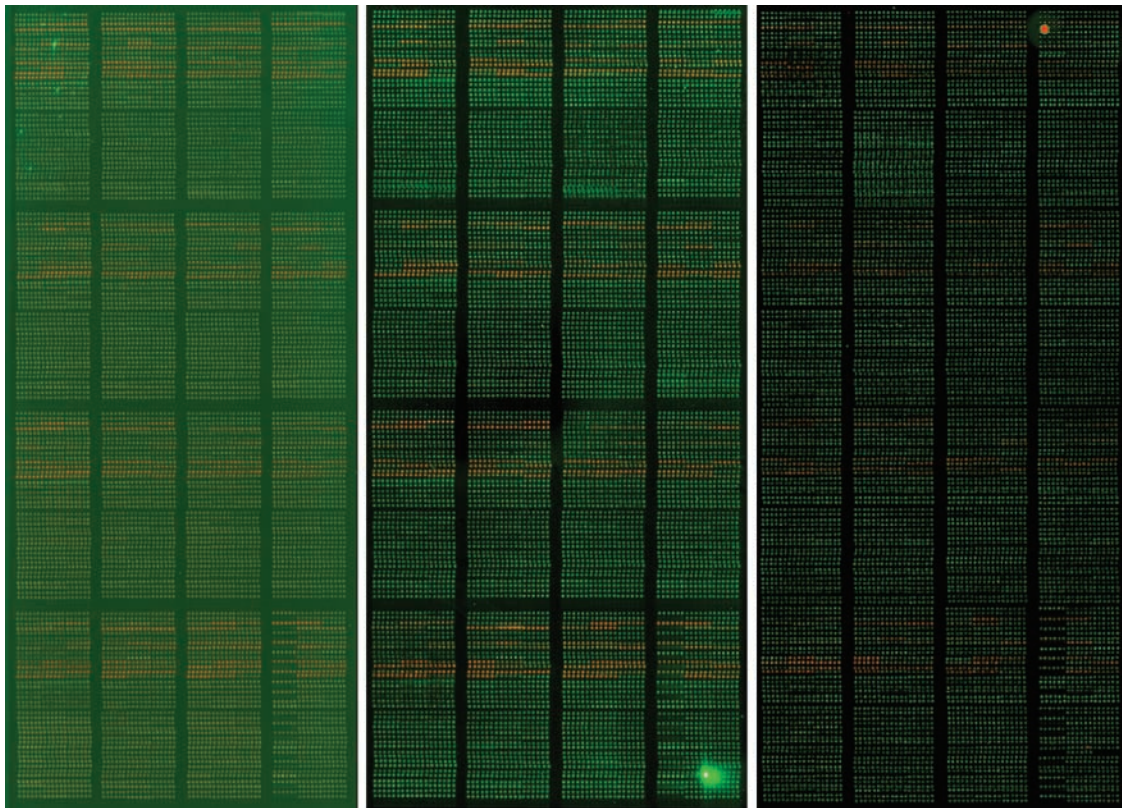


Figure 2. Representative images of three microarray chips that were printed and immunoreacted against serum from cancer patient SCC006 at different times. *Orange spots*, immunoreactivity with serum samples; *green spots*, no reactivity. Because clones were spotted in replicate of six onto the slides, each positive clone was represented by a row of six *orange spots*. Despite differences in background intensity and image quality, the visually positive clones were reproducible across all three immunoreacted chips (interassay reproducibility). Further, the six spots that represent each clone seemed to have uniform intensity visually (intraassay reproducibility).

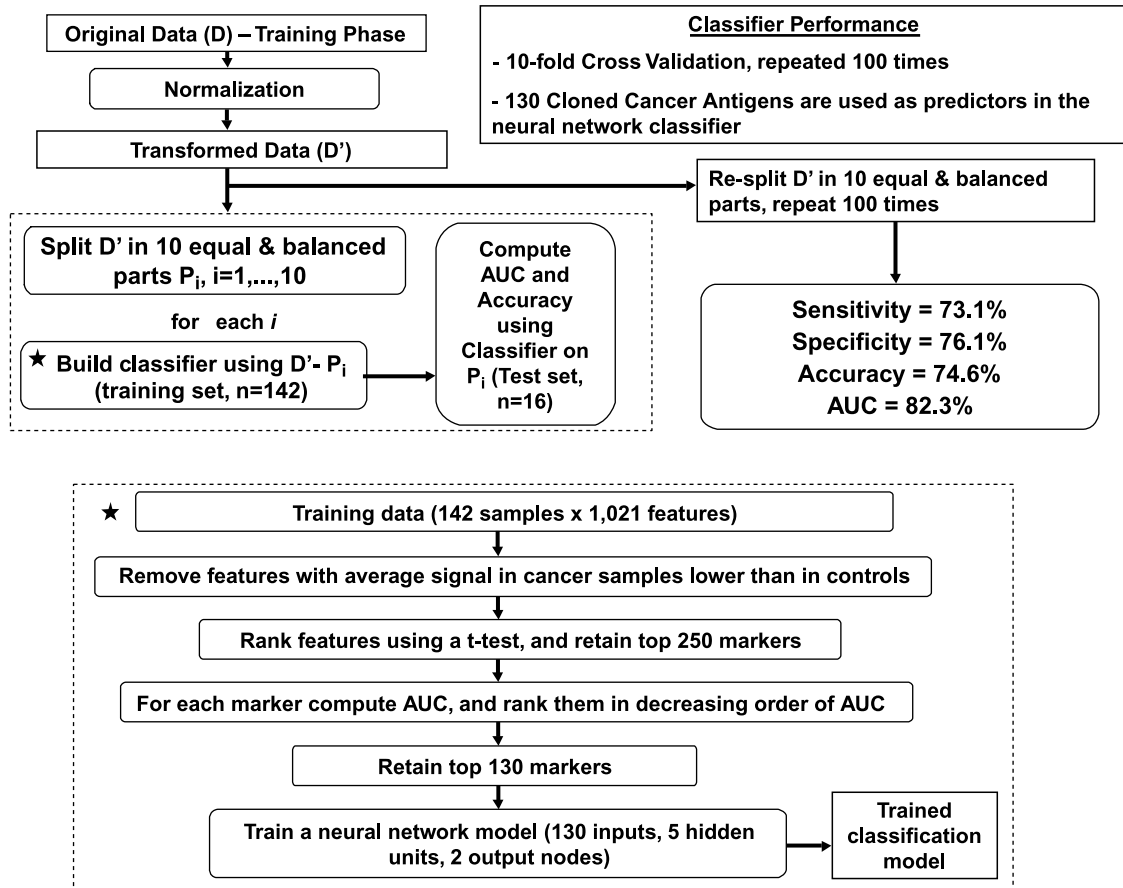


Figure 3. Schema showing the calculation of classifier performance based on 10-fold cross-validation. The transformed data (D') was split into 10 equal and balanced parts to obtain a 10-fold cross-validation data partition. At each fold ($i = 1 \dots 10$), the i^{th} fraction of the data P_i (~ 16 samples) was kept aside for testing purposes. The remaining set $D' - P_i$ (~ 142 samples) was used to select markers and train the model. Marker selection started with ranking the 1,021 clones using a one-tailed t test. The top 250 markers were then reranked using the AUC. The top 130 clones were retained to build a classification model based on a three-layer feed-forward neural network. Special attention was paid to avoid data overfitting by using a reduced number of hidden nodes ($n = 5$) as well as using a training algorithm that includes regularization of the variables in the model. The resulting classifier was then tested against the independent test set P_i . The entire 10-fold cross-validation was repeated 100 times to minimize any potential bias due to random partitioning of training and test sets. This classifier, used to differentiate cancer from noncancer serum samples, has an accuracy of 74.6% (95% confidence interval, 52.5-96.7%) with an AUC of 82.3%, sensitivity of 73.1%, and specificity of 76.1%.

had advanced-stage disease (III and IV), reflecting the distribution of HNSCC in our clinical practice. HNSCCs from almost all subsites of head and neck were represented (Supplementary Table S3A). To reflect the target screening population, control sera used were taken from patients who presented with signs and symptoms similar to that of HNSCC patients. Many of these control patients also have history of moderate to excessive tobacco and/or alcohol use. Cases and controls were matched in terms of age, race, and gender (Supplementary Table S3B).

With this data set, we did both the clone (feature) selection and model training based on nine tenth ($n = 142$) of the data set and the model testing on the remaining one tenth ($n = 16$) of the data set, which were independent and not used previously in the selection or training of the model. We then calculated the performance of the classifiers built using panels with varying

number of the top-ranked biomarkers, ranging from 10 to 180, in increments of 10 (Supplementary Table S4). Although similar accuracy could be obtained using a smaller panel of biomarkers, the use of a panel of 130 biomarkers was found to be the best compromise to maximize accuracy, preserve diversity of clones for future studies, and keep as low as possible the complexity of the models. Thus, the top 130 clones were retained to build a classification model, which was then tested against the independent test set of 16 samples to assess the performance of the classifier. Considering that any result obtained from a random partitioning of data into training and testing sets only once may not be truly reflective of the accuracy of this model in the real world, we repeated this process 1,000 times (10-fold cross-validation $\times 100$; Fig. 3). Thus, for each of the 1,000 splits, a different training set was randomly selected to build a classification model using the top 130 clones selected

from this particular training set. This classifier was then applied to a completely separate test set, not involved in selection of the top 130 clones or building the classifier. By averaging the performance of these 1,000 classifiers, we obtained an average accuracy of 74.6% (95% confidence interval, 52.5-96.7%), AUC of 82.3%, sensitivity of 73.1%, and specificity of 76.1%. Notably, this classifier was able to detect early-stage HNSCC (72.8%) at least as well as late-stage cancers (73.2%). The sensitivity of this classifier in detecting cancer from different subsites of head and neck region was 73.9% (glottis), 72.6% (supraglottis), 83.7% (hypopharynx), 74.9% (oropharynx), 87.5% (nasopharynx), 67.3% (oral cavity), and 60% (unknown primary).

To further verify that there is a true link between the immunoreaction level of the different clones and the class membership of the samples (HNSCC versus control), we randomly permuted the class identifiers among the patients and recalculated in the same way the performance indices with the permuted data (36). As expected, the estimate of accuracy and AUC obtained in these permuted cases was ~50% and is statistically significantly different ($P < 1e-15$) from the accuracy (74.6%) and AUC (82.3%) obtained using the actual class identifiers (Supplementary Fig. S2).

In addition to providing a robust estimate of the classification performance, the 10-fold cross-validation repeated 100 times allowed us to identify the most reliable markers. For each of the 1,000 splits, a slightly different panel of 130 biomarkers was selected from each of the 1,000 randomly generated training set. To establish a final set of 130 markers that can be used for further studies, we ranked each of the clones based on the number of times out of the 1,000 feature selection processes that it was selected as one of the top 130 biomarkers used in the panel to build the classifier for distinguishing cancer from noncancer sera (Supplementary Table S5).

Characterization of the Panel of 130 Biomarkers. The panel of 130 markers was sequenced and analyzed for

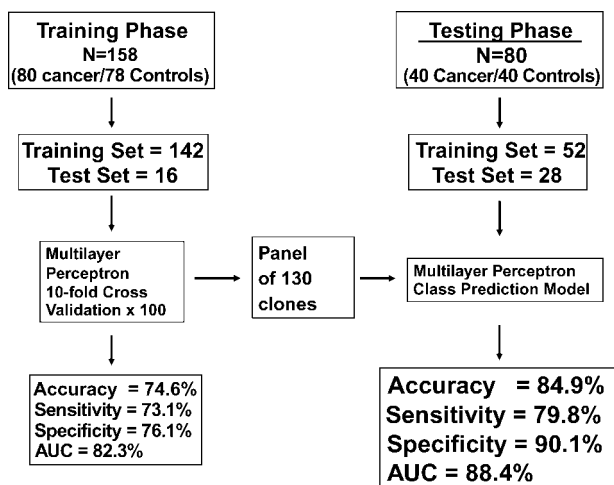


Figure 4. Overview of the strategy used for the development and validation of the panel of 130 biomarkers for serum-based cancer detection.

homology to mRNA and genomic entries in the Genbank databases using BLASTn. We also determined the predicted amino acids in-frame with the phage T7 gene 10 capsid protein. Of the top 130 clones, there were 8 clones that contained known gene products in the reading frame of the T7 gene 10 capsid proteins. These included multiple myeloma overexpression gene 2, ubiquinone binding protein, NADH dehydrogenase subunit 1, C10 protein, and a hypothetical protein LOC400242 (Supplementary Table S6). The remaining 122 clones contained peptides that were different from the original proteins coded by the inserted gene fragments. This occurred because the inserted gene fragments were out of frame with the open reading frame of the T7 10B gene ($n = 61$; Supplementary Table S7A), represented untranslated region of known genes ($n = 18$; Supplementary Table S7B), or contained sequences from unknown genes ($n = 43$; Supplementary Table S7C). It is likely that the recombinant gene products of these clones mimic some other natural antigens and hence can be termed mimotopes (21, 37). It is also possible that some of these products may represent cancer antigens produced as a result of altered reading frame or alternative splicing (38-40). BLASTp search of the SWISSPROT database for homology to each in-frame mimotope revealed that many of these gene products mimic known cancer proteins and as such represent putative tumor antigens.

Validation of Panel of 130 Biomarkers. Sera from completely new set of 40 HNSCC patients and 40 controls, not previously used for biopanning or selection of clones, were immunoreacted against the panel of 130 biomarkers selected from the previously described training phase. Of the 40 HNSCC patients, 11 had early-stage disease (I and II) and 29 had advanced-stage disease (III and IV). HNSCCs from all subsites of head and neck were represented (Supplementary Table S8A). Cases and controls were matched in terms of age, race, and gender (Supplementary Table S8B). The new samples were randomly split into a training set of 52 samples (26 HNSCC and 26 controls) and a test set of 28 samples (14 HNSCC and 14 controls). Multilayer Perceptron Classifier, based on this final panel of 130 biomarkers, was then trained on the training set of 52 samples and tested against the independent test set of 28 samples. This resulted in an AUC of 88.4% and accuracy of 84.9% with sensitivity of 79.8% and specificity of 90.1%. Similar accuracy was achieved by reshuffling of the training and test sets two more times and by using other types of class prediction models, such as logistic regression (accuracy = 87.5%, sensitivity = 90%, specificity = 85%) and sequential minimal optimization algorithm in Weka (accuracy = 85%, sensitivity = 92.5%, specificity = 77.5%; Fig. 5; ref. 41).

Discussion

In this study, we used protein microarray technology for high-throughput quantitative analysis of the antibody-antigen reaction between 238 serum samples and 5,133 cloned cancer antigens preselected via biopanning. Using specialized bioinformatics, we mined through this massive data set to identify a panel of top 130 cancer biomarkers useful for building classifier models (Fig. 1). Many of these markers represent or mimic known cancer

antigens. To further validate the ability of this panel of markers for cancer detection, we used Multilayer Perceptron Classifier, based on these 130 cancer biomarkers, to train and classify an independent set of 80 serum samples. The performance of this panel of 130 markers in discriminating cancer from noncancer serum samples was excellent with an accuracy of 84.9% (sensitivity of 79.8% and specificity of 90.1%; Fig. 4), which represented a significant improvement over current diagnostic accuracy in the primary care setting with reported sensitivity of 37% (9, 14, 15) to 46% (13) and specificity of 24% (13).

As previously stated, currently, there are no guidelines or tests for the early detection of HNSCC. Because the prevalence of HNSCC in the general population is low (~1 in 1,510 American), a screening assay will need to have such a high degree of sensitivity and specificity that may be unattainable. Thus, it is essential that high-risk patients be identified and targeted for screening. Data from one of the largest multicenter population-based case-control study for oral and pharyngeal cancers in the United States showed that, for those patients who consume large amount of alcohol and tobacco, their risk of developing HNSCC can be up to 37.7 times over the general population (7). The prevalence of HNSCC in this population is therefore 2.5%. Patients with history of prior HNSCC, now free of disease, represent another group of high-risk patients. These patients have approximately 30% to 45% chance of locoregional recurrences (if they are within the first 5 years after completion of treatment; refs. 42, 43) as well as 5% risk per year of developing second primary cancer of the upper aerodigestive tract (3, 44). Thus, this group of patients has an estimated prevalence of ~10%. Despite close surveillance by head and neck specialists, early cancer detection in these patients can still be challenging given the difficulty of examining and imaging the heavily irradiated and/or operated head and neck region. Given the approximate prevalence of 2.5% to 10% in these two high-risk populations, a screening test with a sensitivity of 79.8% and specificity of 90.1% will yield positive predictive values of 17.1% to 47.3% and negative predictive values of 97.6% to 99.4%, which should be sufficient for clinical testing in these patients (Fig. 5).

Although the result of the diagnostic test shown here seems promising and can achieve very high positive predictive value and negative predictive value when applied to high-risk populations (Fig. 5), this high degree of accuracy in the laboratory may not hold true when we increase the sample size and extend our study to the population outside the laboratory. Indeed, our control patients were not selected from the two high-risk populations described above. Efforts are ongoing to recruit and study sera from patients in these high-risk populations. However, to reflect the target screening population in the primary care practice, we did intentionally select our control population from patients who presented with symptoms or exams similar to that of head and neck cancer patients. We also included many control patients with history of moderate to excessive tobacco and/or alcohol use to minimize bias due to differential use of tobacco and alcohol between the HNSCC and control groups.

In this study, we included a large number of serum samples from patients with advanced disease because

Classifier	Accuracy	Sensitivity	Specificity	Prevalence	PPV	NPV
Multilayer Perceptron	84.9%	79.8%	90.1%	0.066%	0.53%	99.99%
				2.5%	17.1%	99.4%
				10%	47.3%	97.6%
SMO	85%	92.5%	77.5%	0.066%	0.27%	99.99%
				2.5%	9.54%	99.8%
				10%	31.4%	98.94%
Logistic Regression	87.5%	90%	85%	0.066%	0.4%	99.99%
				2.5%	13.3%	99.7%
				10%	40.0%	98.7%

Figure 5. Chart showing the positive predictive value (PPV) and negative predictive value (NPV) in three populations of different disease prevalence. The prevalence of HNSCC in the general population is low, ~1 in 1,510 American or 0.066%. In patients who consume large amount of alcohol and tobacco, their risk of developing HNSCC can be up to 37.7 times over the general population, giving them a prevalence of 2.5%. Finally, patients with history of prior HNSCC, now free of disease, have approximately 30% to 45% chance of locoregional recurrences (if they are within the first 5 years after completion of treatment) as well as 5% risk per year of developing second primary cancer of the upper aerodigestive tract. Thus, this group of patients has an estimated prevalence of ~10%. Given the approximate prevalence of 2.5% to 10% in these two high-risk populations, a screening test, such as the one described here, with a sensitivity of 79.8% and specificity of 90.1% will yield positive predictive values of 17.1% to 47.3% and negative predictive values of 97.6% to 99.4%, which should be sufficient for clinical testing in these patients. SMO, sequential minimal optimization.

these sera were more likely to contain higher levels and greater variety of antibodies against the greater load of cancer antigens present in these patients. Further, the disproportionately larger representative of advanced-stage patients in this study was representative of the distribution within our serum bank, which was reflective of the late-stage presentation of the majority of our cancer patients. Although we showed here a panel of biomarkers capable of detecting all stages of HNSCC with high degree of accuracy, we cannot make any conclusion about how this panel will do in detecting early-stage cancer due to the low proportion of early-stage sera used in this study. Thus, there is a need for further validation studies to include a large number of early-stage sera.

The results presented in this study indicate the potential of a new platform for head and neck cancer detection based on analysis of pattern of serum immunoreactivity against a panel of cancer antigens. We and others have previously used this technology and showed a high degree of diagnostic accuracy for cancer detection in ovarian (21), lung (45, 46), and prostate (22) cancers. We found this pattern of immunoreactivity to be highly reproducible. In addition, we have found that serum

IgGs are extremely stable, which should minimize interlaboratory variations in clinical diagnostics setting. Further, the potential to translate our approach into an assay system already widely available in clinical practice, ELISA, represents a major advantage of this technology over that of other technologies, such as matrix-assisted laser-desorption and ionization and surface-enhanced laser desorption and ionization time-of-flight mass spectrometry. These unique ionization techniques represent other powerful and novel techniques for identification of potential diagnostic biomarkers from a variety of biological and clinical samples (47, 48) and have been applied successfully to discriminate cancer from non-cancer based on differences in the serum protein profiles. Sensitivity and specificity have been reported to be 83% and 97% for prostate cancer (49), 100% and 95% for ovarian cancer (48), and 83.3% and 90% for HNSCC (50). Although this technology seems promising with the reported high sensitivity and specificity, it is potentially limited by its applicability in routine clinical settings because the implementation of this technology requires expensive equipment and software that may be available only at research facility or tertiary care centers.

Finally, the pattern of reactivity of biomarkers with serum samples from HNSCC may be analyzed to develop other classifiers capable of predicting clinical outcome and thereby guiding the most optimal therapeutic treatments. These biomarkers may also have utility in posttreatment monitoring of HNSCC patients and may even provide new targets for therapeutic interventions or diagnostic imaging in future clinical trials. Because the host immune system can reveal molecular events (overexpression or mutation) critical to the genesis of HNSCC, this novel proteomics technology can also identify genes with mechanistic involvement in the etiology of the disease.

In conclusion, using epitomics technology based on a combination of high-throughput antigen selection using microarray-based serologic profiling and specialized bioinformatics, we identified a panel of 130 biomarkers that can provide sufficient accuracy for a clinically relevant, serum-based cancer detection test based on the pattern of serum IgG binding. The results shown here are encouraging and show the potential to use this promising technology toward the eventual development of diagnostic assay for detection of early-stage cancer. Further work with larger panels of antigens as well as refinement of this technology should provide a comprehensive set of biomarkers to facilitate the building of a classifier with sufficient sensitivity and specificity suitable for clinical testing in populations at high-risk for HNSCC.

Acknowledgments

We wish to thank Dr. Robert H. Mathog, chairman at Department of Otolaryngology-Head and Neck Surgery at Wayne State University, for his unwavering support and encouragement in this project. We also wish to thank Dr. Wei-zen Wei from Department of Immunology & Microbiology for her help in the area of immunology and Nancy K. Levin for her assistance in preparation of IRB proposal. Finally, we like to thank the Biostatistics, Bioinformatics and the Applied Genomics Core, Karmanos Cancer Institute, Wayne State University (P30CA022453) for expert assistance.

References

1. American Cancer Society. Estimated new cancer cases and deaths by sex for all sites, U.S., 2007. Cancer facts & figures 2007. Atlanta (GA): American Cancer Society; 2007. p. 4.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
3. Clayman GL, Laramore GE, Hong WK. Head and Neck Cancer. In: Bast RC, Pollock RE, Weichselbaum RR, Holland JF, Frei E, editors. *Cancer medicine*. Hamilton: B.C. Decker, Inc.; 2000. p. 1175–220.
4. American Cancer Society. Trends in 5-year relative survival rates (%) by race and year of diagnosis, U.S., 1975–2002. *Cancer facts & figures 2007*. Atlanta (GA): American Cancer Society; 2007. p. 18.
5. Vokes EE, Weichselbaum RR, Lippman SM, Hong WK. Head and neck cancer. *N Engl J Med* 1993;328:184–94.
6. American Cancer Society. Screening guidelines. *Cancer facts & figures 2007*. Atlanta (GA): American Cancer Society; 2007. p. 52.
7. Blot WJ, McLaughlin JK, Winn DM, et al. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res* 1988;48:3282–7.
8. Castellsague X, Quintana MJ, Martinez MC, et al. The role of type of tobacco and type of alcoholic beverage in oral carcinogenesis. *Int J Cancer* 2004;108:741–9.
9. Teppo H, Koivunen P, Hyrynkanas K, Alho OP. Diagnostic delays in laryngeal carcinoma: professional diagnostic delay is a strong independent predictor of survival. *Head Neck* 2003;25:389–94.
10. Carvalho AL, Pintos J, Schlecht NF, et al. Predictive factors for diagnosis of advanced-stage squamous cell carcinoma of the head and neck. *Arch Otolaryngol Head Neck Surg* 2002;128:313–8.
11. Bruun JP. Time lapse by diagnosis of oral cancer. *Oral Surg Oral Med Oral Pathol* 1976;42:139–49.
12. Pitchers M, Martin C. Delay in referral of oropharyngeal squamous cell carcinoma to secondary care correlates with a more advanced stage at presentation, and is associated with poorer survival. *Br J Cancer* 2006;94:955–8.
13. Hoare TJ, Thomson HG, Proops DW. Detection of laryngeal cancer—the case for early specialist assessment. *J R Soc Med* 1993;86:390–2.
14. Koivunen P, Rantala N, Hyrynkanas K, Jokinen K, Alho OP. The impact of patient and professional diagnostic delays on survival in pharyngeal cancer. *Cancer* 2001;92:2885–91.
15. Tromp DM, Brouha XD, Hordijk GJ, Winnubst JA, de Leeuw RJ. Patient and tumour factors associated with advanced carcinomas of the head and neck. *Oral Oncol* 2005;41:313–9.
16. Alho OP, Teppo H, Mantyselka P, Kantola S. Head and neck cancer in primary care: presenting symptoms and the effect of delayed diagnosis of cancer cases. *CMAJ* 2006;174:779–84.
17. Allison P, Franco E, Black M, Feine J. The role of professional diagnostic delays in the prognosis of upper aerodigestive tract carcinoma. *Oral Oncol* 1998;34:147–53.
18. Scanlan MJ, Chen YT, Williamson B, et al. Characterization of human colon cancer antigens recognized by autologous antibodies. *Int J Cancer* 1998;76:652–8.
19. Disis ML, Calenoff E, McLaughlin G, et al. Existent T-cell and antibody immunity to HER-2/neu protein in patients with breast cancer. *Cancer Res* 1994;54:16–20.
20. Diesinger I, Bauer C, Brass N, et al. Toward a more complete recognition of immunoreactive antigens in squamous cell lung carcinoma. *Int J Cancer* 2002;102:372–8.
21. Chatterjee M, Mohapatra S, Ionan A, et al. Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. *Cancer Res* 2006;66:1181–90.
22. Wang X, Yu J, Sreekumar A, et al. Autoantibody signatures in prostate cancer. *N Engl J Med* 2005;353:1224–35.
23. Carey TE, Kimmel KA, Schwartz DR, Richter DE, Baker SR, Krause CJ. Antibodies to human squamous cell carcinoma. *Otolaryngol Head Neck Surg* 1983;91:482–91.
24. Disis ML, Bernhard H, Gralow JR, et al. Immunity to the HER-2/neu oncogenic protein. *Ciba Found Symp* 1994;187:198–207; discussion 207–11.
25. Soussi T. p53 antibodies in the sera of patients with various types of cancer: a review. *Cancer Res* 2000;60:1777–88.
26. Chen YT, Scanlan MJ, Sahin U, et al. A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proc Natl Acad Sci U S A* 1997;94:1914–8.
27. Chatterjee M, Draghici S, Tainsky MA. Immunotherapeutics: breaking tolerance in immunotherapy using tumor autoantigens identified on protein microarrays. *Curr Opin Drug Discov Devel* 2006;9:380–5.
28. Draghici S, Chatterjee M, Tainsky MA. Epitomics: serum screening

- for the early detection of cancer on microarrays using complex panels of tumor antigens. *Expert Rev Mol Diagn* 2005;5:735–43.
29. Chatterjee M, Ionan A, Draghici S, Tainsky MA. Epitomics: global profiling of immune response to disease using protein microarrays. *OMICS* 2006;10:499–506.
 30. R Development Core Team. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2004.
 31. Smyth GK. Limma: linear models for microarray data. In: R. Gentleman VC, Dudoit S, Irizarry R, Huber W, editors. *Bioinformatics and computational biology solutions using R and Bioconductor*. New York: Springer; 2005. p. 397–420.
 32. Venables WN, Ripley BD, Venables WN. *Modern applied statistics with S*. 4th ed. New York: Springer; 2002.
 33. Ren Z, Black LW. Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid. *Gene* 1998;215:439–44.
 34. Rosenberg A, Griffin G, Studier FW, McCormick M, Berg J, Mierendorf R. T7Select phage display system: a powerful new protein display system based on bacteriophage T7. *Innovations* 1996;6:1–6.
 35. Walter G, Konthur Z, Lehrach H. High-throughput screening of surface displayed gene products. *Comb Chem High Throughput Screen* 2001;4:193–205.
 36. Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005;23:7332–41.
 37. Geysen HM, Rodda SJ, Mason TJ. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol Immunol* 1986;23:709–15.
 38. Wang RF, Parkhurst MR, Kawakami Y, Robbins PF, Rosenberg SA. Utilization of an alternative open reading frame of a normal gene in generating a novel human cancer antigen. *J Exp Med* 1996;183:1131–40.
 39. David A, Majeesh N, Azar I, et al. Unusual alternative splicing within the human kallikrein genes KLK2 and KLK3 gives rise to novel prostate-specific proteins. *J Biol Chem* 2002;277:18084–90.
 40. Rimoldi D, Rubio-Godoy V, Dutoit V, et al. Efficient simultaneous presentation of NY-ESO-1/LAGE-1 primary and nonprimary open reading frame-derived CTL epitopes in melanoma. *J Immunol* 2000;165:7253–61.
 41. Witten IH FE, editor. *Data mining: practical machine learning tools and techniques*. 2nd ed. San Francisco: Morgan Kaufmann; 2005.
 42. Jesse RH, Sugarbaker EV. Squamous cell carcinoma of the oropharynx: why we fail. *Am J Surg* 1976;132:435–8.
 43. Shah JP, Cendon RA, Farr HW, Strong EW. Carcinoma of the oral cavity: factors affecting treatment failure at the primary site and neck. *Am J Surg* 1976;132:504–7.
 44. Benner SE, Pajak TF, Lippman SM, Earley C, Hong WK. Prevention of second primary tumors with isotretinoin in patients with squamous cell carcinoma of the head and neck: long-term follow-up. *J Natl Cancer Inst* 1994;86:140–1.
 45. Zhong L, Coe SP, Stromberg AJ, Khattar NH, Jett JR, Hirschowitz EA. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J Thorac Oncol* 2006;1:513–9.
 46. Zhong L, Hidalgo GE, Stromberg AJ, Khattar NH, Jett JR, Hirschowitz EA. Using protein microarray as a diagnostic assay for non-small cell lung cancer. *Am J Respir Crit Care Med* 2005;172:1308–14.
 47. Issaq HJ, Veenstra TD, Conrads TP, Felschow D. The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. *Biochem Biophys Res Commun* 2002;292:587–92.
 48. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002;359:572–7.
 49. Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002;62:3609–14.
 50. Wadsworth JT, Somers KD, Stack BC Jr, et al. Identification of patients with head and neck cancer using serum protein profiles. *Arch Otolaryngol Head Neck Surg* 2004;130:98–104.