Combined Use of Oligonucleotide and Tissue Microarrays Identifies Cancer/Testis Antigens as Biomarkers in Lung Carcinoma

Michio Sugita, Mark Geraci, Bifeng Gao, Roger L. Powell, Fred R. Hirsch, Gary Johnson, Razvan Lapadat, Edward Gabrielson, Roy Brennes, Paul A. Bunn, and Wilbur A. Franklin

Departments of Pathology [M. S., R. L. P., W. A. F.], Medicine [M. G., B. G., F. R. H., R. B., P. A. B.], and Pharmacology [G. J., R. L.], University of Colorado Health Sciences Center, Denver, Colorado 80262, and Department of Pathology, F. S. Key Medical Center, Baltimore, Maryland 21224 [E. G.]

ABSTRACT

High density oligonucleotide microarrays (OMAs) have been used recently to profile gene expression in lung carcinoma tissue homogenates. The length and the lists of potentially interesting genes generated by these studies is daunting, and biological and clinical relevance of these lists remains to be validated. Moreover, specific identification of individual biomarkers that might be used for early detection and surveillance has not been the objective of these early studies. We have developed a schema for combining the data derived from the OMA analysis of a few lung cancer cell lines with immunohistochemical testing of tissue microarrays to rapidly identify biomarkers of potential clinical relevance. Initially, we profiled gene expression in lung tumor cell lines using the Affymetrix HGU95Av2 OMA. RNA from 2 non-small lung cancer (NSCLC) cell lines (A549 and H647) and 2 small cell lung cancer (SCLC) cell lines (SHP-77 and UMC-19) were tested. Cells from 1 histologically and cytogenetically normal bronchial epithelial primary culture from a volunteer who had never smoked and 10 samples of histologically unremarkable lung tissue from resection specimens served as normalization controls. Array results were analyzed with Gene Spring software. Results were confirmed by reverse transcription-PCR in an expanded number of cell lines. We then validated the cell line data by immunohistochemical testing for protein expression using a tissue microarray containing 187 NSCLC clinical samples. Of the 20 most highly expressed genes in the tumor lines, 6 were members of the cancer/testis antigen (CTAG) gene group including 5 MAGE-A subfamily members and NY-ESO-1. SCLC lines strongly expressed all of the MAGE-A genes as well as NY-ESO-1, whereas NSCLC lines expressed a subset of MAGE-A genes at a lower level of intensity and failed to express NY-ESO-1. Reverse transcription-PCR of an expanded series of 25 lung cancer cell lines including 13 SCLC, 9 NSCLC, and 3 mesothelioma lines indicated that MAGE-A10 and NY-ESO-1 were expressed only by SCLC, and that MAGE-A1, 3, 6, 12, and 4b were expressed by both SCLC and NSCLC. By immunohistochemistry using the monoclonal antibody 6C1 that recognizes several MAGE-A gene subfamily members, 44% of NSCLC clearly expressed MAGE-A proteins in cytoplasm and/or nucleus. Expression of MAGE-A genes did not correlate with survival but did correlate with histological classification with squamous carcinomas more frequently MAGE-A positive than other NSCLC types (P < 0.00002). We conclude that expression of CTAG gene products, whereas apparently not of prognostic importance, may be useful for early detection and surveillance because of a high level of specificity for central airway squamous and small cell carcinomas.

INTRODUCTION

Late stage at detection significantly and adversely affects survival in lung cancer (1, 2). The most difficult obstacles to earlier detection are the inaccessibility of the sites of tumor origin and the multiplicity of sites from which tumors may arise. Several approaches to overcoming these problems are currently being tested ranging from new imaging technologies (helical CT; Refs. 3, 4) to aggressive efforts to identify high-risk cohorts (5). Biological properties of the tumor cells themselves (biomarkers) may also be exploited to identify subjects who might harbor clinically inapparent tumors. Molecules that are expressed uniquely or at high level by tumor cells in comparison to normal tissues and that may be secreted into accessible fluids such as blood, urine, or sputum may be useful as lung cancer biomarkers.

It might be expected that, because of their stark morphological distinction from normal lung cells and their aggressive biological behavior, lung cancer cells may exhibit many molecular differences from non-neoplastic lung cells. To date there have been numerous attempts to identify such molecules with limited success. Lung cancer biomarkers measurable in the peripheral blood have included carbohydrate-rich cell matrix molecules such as carciinoembryonic antigen (6), cytokeratin-derived intermediate filament molecules such as CYFRA-21.1 (7, 8), tissue polypeptide antigen (TPA) (9), and tissue polypeptide specific antigen (TPS) (10), peptides such as proGRP (11), neural markers such as neuron specific enolase (12, 13) and chromogranin A (14, 15), and antibodies to immunogenic molecules such as Hu (16), calcium channel proteins (17), and p53 (18, 19). Thus far tests for these molecules have had limited clinical impact because of low specificity or low frequency of positive results in early stage patients. However, it is likely that the list of biomarkers already tested represents only a fraction of the molecular changes that occur in tumor cells, and that more sensitive and specific biomarkers remain to be discovered.

Recently, high-density OMAs3 have been introduced that permit rapid analysis of expression levels simultaneously for large numbers of genes (20). This approach overcomes limitations inherent in expression analysis of single genes. With completion of human genome (21) sequencing, comprehensive OMA expression profiles can be created for individual tumors as well as for large classes of tumors. Early OMA analyses of lung cancers have centered on phenotypic classification of specific tumor type (22) and have not specifically focused on biomarker discovery.

Our objective in the present study was to discover potentially useful biomarkers for lung cancer by first identifying large gene expression differences between tumor cell lines and normal lung using high density OMAs. The microarray used (Affymetrix HGU95Av2) incorporates 12,600 probes accounting for a large fraction of the expressed human genome. We searched for biomarkers that were overexpressed in relation to normal tissue, because they are more likely to be useful for detection and screening of accessible specimens such as sputum, peripheral blood, or urine than biomarkers that are underexpressed. We confirmed the expression levels of the gene group (CTAG) most frequently represented on a list of highly expressed genes by testing a broader series of cell lines using relatively inex-
pensive RT-PCR methodology. Finally, results of this preliminary testing were confirmed at the protein level by IHC using a TMA containing 187 early stage NSCLCs. This algorithm for biomarker development takes advantage of two high throughput microarray technologies to rapidly identify potentially important biomarkers linked to clinical outcomes and prognostic importance.

MATERIALS AND METHODS

Samples and RNA Extraction Procedures. Four established cell lines including 2 SCLC (SHP-77 and UMC-19) and 2 NSCLC (A549 and H647) lines were analyzed. Before harvesting, SCLC and NSCLC cells were grown in RPMI 1640 supplemented with 5–10% fetal bovine serum. Substrate adherent cultures (SHP-77, A549, and H647) were grown to subconfluence and harvested by rapid removal of medium and application of RNeasy extraction medium (Qiagen, Valencia, CA) containing guanidinium isothiocyanate. Nonadherent cells were harvested at a concentration of ~1 million cells/ml after ~1 week in culture after thawing.

Two controls were used for comparison to expression profiles of tumor cell lines. For one control, bronchial epithelial primary cell cultures were obtained from a bronchoscopic biopsy of a healthy 48-year-old female who had never smoked and who had volunteered under a Colorado Combined Institutional Review Board-approved protocol. The biopsy was explanted onto a T25 culture flask containing bronchial epithelial cell growth medium (Clonetics, Inc., Walkersville, MD) and epithelial cells were allowed to grow from the explant to a diameter of 1 cm (10 days). Cells were then passaged into a second T25 flask and grown to ~90% confluence (4 days). The culture cells were again split onto glass coverslips to perform spectral karyotyping on metaphase cells according to the manufacturer’s protocol (Applied Spectral Imaging, Inc., Carlsbad, CA). A second aliquot was split into two T75 flasks and again grown to 90% confluence (4 days). One flask was additionally split into three T75 flasks and expanded for an additional 3 days. Finally, 90% confluent cells were harvested by removal of culture medium followed by immediate addition of RNeasy extraction medium as described above. The total time from biopsy date to RNA harvest was 21 days. Primary cultures processed in this way grew as substrate-adherent monolayers, which are 100% cytokeratin positive on immunohistochemical staining (23). Spectral imaging karyotype was diploid with no detectable subchromosomal abnormalities.

A second set of controls consisted of archival data obtained from experiments in which RNA was extracted from benign lung tissue obtained at the time of surgical resection for carcinoma elsewhere in the lung. For these experiments, duplicate tissue samples from 10 lung specimens were snap frozen and stored in liquid nitrogen until use. For RNA extraction, frozen tissue fragments were placed in RNeasy extraction medium and homogenized with a Tissue Tearer homogenizer (Biospec Products, Bartlesville, OK) followed by filtration through a QIAshredder column. The filtrate was used for RNA extraction using the Qiagen RNeasy Mini protocol.

Total RNA extracted from each sample described above was tested for degradation and applied to a separate HG-U95Av2 microarray. Each control RNA from cultured normal bronchial cells or whole lung homogenate was used as a separate normalization control in the GeneSpring filtering algorithms described below.

Preparation of Labeled cRNA and Hybridization to OMAs. Before application to test chips the quality of RNA was tested using the one step duplex RT-PCR assay (24). In this assay, the ratio of short to long segment β-actin PCR product is used to quantify the extent of RNA degradation. All of the samples in this study had ratios of <2.6 indicating a low level of degradation.

Double-stranded cDNA was synthesized from 16 to 20 μg total RNA using an oligodeoxynucleotidyl acid 24 primer with a T7 RNA polymerase promoter site added to the 3′ end (Superscript cDNA Synthesis System; Life Technologies, Inc., Rockville, MD). After second-strand synthesis, in vitro transcription was performed using a T7 Megascript kit (Ambion, Austin, TX) in the presence of biotin-11-CTP and biotin-16-UTP (Enzo Diagnostics, Farmingdale, NY) to produce biotin labeled cRNA. Twenty μg of the cRNA product was fragmented at 94°C for 35 min into 35–200 bases in length. The sample was then added to a hybridization solution containing 100 mmol/liter 4-morpholinepropanesulfonic acid, 1 mol/liter Na+ , and 20 mmol/liter of EDTA in the presence of 0.01% Tween 20 to a final cRNA concentration of 0.05 mg/ml. Hybridization was performed for 18–20 h by incubating 200 μl of the sample to HG-U95Av2 microarrays, and each microarray was stained with streptavidin-phycerythrin and scanned at 6-μm resolution by Gene Array scanner G2500A (Hewlett Packard, Boise, ID) according to procedures developed by Affymetrix.

Statistical Analysis. Detailed protocols for data analysis of Affymetrix microarrays, and extensive documentation of the sensitivity and quantitative aspects of the method have been described (20, 25).

Briefly, mismatch probes act as specificity controls that allow the direct subtraction of both background and cross-hybridization signals. To determine the quantitative RNA abundance, the average of the difference representing perfect match – mismatch for each genespecific probe family is calculated. This data were transferred to GeneSpring software (Silicon Genetics, Redwood City, CA) for additional analysis.

Using the GeneSpring software package, a two step filtering algorithm was implemented to select genes highly expressed by tumor cells in comparison with non-neoplastic lung cells and tissue. In the first step, cultured normal epithelial cells were compared with tumor cell lines using the following settings: the 80th percentile of all measurements was used as a positive control for each sample, and each measurement was divided by this control. The 0.1% measurement was used as a control for background correction. The measurement for each gene was then divided by the corresponding value for the sample of normal bronchial epithelium. A list was compiled of all of the genes expressed by at least two tumor cell lines at >100× over the normal control. This filtering step resulted in the identification of 42 genes. In the second filtering step, the 50th percentile of all measurements was used as a positive control for each sample, and each measurement was divided by this control. The measurement for each gene was then divided by the corresponding value for 19 samples from 10 non-neoplastic lung specimens. A list was then compiled of 107 genes that were expressed in the tumor cell lines at >20× over the non-neoplastic tissue. The contents of the two lists was then compared using the GeneSpring Venn diagram feature and a list of 20 highly overexpressed genes common to the two lists was compiled.

The selected genes were annotated using the GeneOntology database4 within the NetAffx® analysis system offered by Affymetrix. GeneOntology stores a dynamic controlled vocabulary organized on molecular function, cellular component, and biological process that can be applied to all organisms. The cellular component attributes were used to search for genes that were either extracellular (secreted) or transmembrane molecules as potential biomarkers.

The secretory attributes of the selected genes were further investigated by looking at the leader sequence signal. Briefly the master protein model sequences were obtained from the LocusLink database

---

5 Internet address: http://www.netaffx.com.
and analyzed using the program SignalP\(^6\) (SignalP version2.0) that detects secretory signal peptides in amino acid sequences. The program splices the first 70 amino acids and runs two different types of detection algorithms: one based on neural network prediction and the other based on Hidden Markov Models. Both are trained against a library of known signal peptides and calculate a final score, which will assign the protein to one of three classes: (a) nonsecretory; (b) signal anchor (NH\(_2\) terminus of type II membrane proteins, uncleaved signal peptides); and (c) signal peptide (secretory signal).

Spearman correlation was used for clustering of all of the hybridization experiments. To evaluate the expression profile for melanoma-associated antigens, a list of 64 melanoma-associated genes was compiled using the GeneSpring search feature for melanoma, and Pearson correlation was used for clustering of this list (see Fig. 1).

To evaluate the relationship between MAGE-A expression as determined by IHC (see below) and survival, log rank test was performed using the SPSS statistical package, version 11.0 (SPSS, Inc., Chicago, IL). \(\chi^2\) analysis was performed using Microsoft Excel.

**Confirmatory RT-PCR Assay.** Gene expression was confirmed by RT-PCR in 25 cell lines with SCLC, NSCLC, and mesothelioma histologies (representative gel shown in Fig. 3). The RT-PCR assay was performed using One-Step RT-PCR system (Life Technologies, Chicago, IL). The RT-PCR products were separated on 1.5% agarose gels and visualized by UV transillumination of the gels stained with ethidium bromide.

**Preparation of Cell Lines for IHC.** Protein expression was evaluated in the 19 cell lines by immunoperoxidase staining of cell pellets created by centrifugation of cultured cells that were then fixed in formalin for 30 min and embedded in paraffin. Sections of the resulting paraffin blocks were stained with the same anti-MAGE monoclonal antibody (6C1; Novacastra, Newcastle, United Kingdom) and by the same methods used for the TMAs described below.

**TMA IHC.** Paraffin blocks of tumor tissue from 187 patients with NSCLC (stages I-III) were obtained from the University of Colorado Cancer Center and Johns Hopkins Medical Institutions according to IRB-approved protocols. Follow-up of patients represented on the TMA ranged from 18 to 100 months. The distribution of tumor histologies and clinical stage in this group of patients is shown in Table 2.

The TMAs were assembled using a tissue-arraying instrument (Beecher Instruments, Silver Spring, MD), consisting of thin-walled stainless steel biopsy needles and styllets used to empty and transfer the needle content. The assembly is held in an X-Y position guide that is manually adjusted by micrometers. A large diameter stylet (1.5 mm) was used for sampling, and non-necrotic areas of the blocks were routinely over-sampled with three replicate core samples of tumor (different areas) and normal (one, if present) regions from each donor block. Normal lung and 15 other control tissues were included in each tissue array block. Four-\(\mu\)m sections of the resulting microarray blocks were cut with a Leitz microtome. Sections were transferred to adhesive coated slides using the adhesive-coated tape sectioning system (Instrumedics Inc., Hackensack, NJ; Ref. 22). Subsequently, UV light treatment of the slides for 60 s polymerized the adhesive coating into a plastic layer and sealed the sections to the slides. Thereafter, the tape could be removed in a solvent (Instrumedics Inc.).

The sections were then deparaffinized with standard xylene and hydrated through graded alcohols into water. Antigen retrieval was performed using the DAKO Target Retrieval system in a Biocare Medical decloaking chamber. Peroxide blocking was performed with 3% hydrogen peroxide in water. After incubation of the mouse monoclonal anti-MAGE-A antibody, 6C1 (Novacastra) for 1 h at room temperature, the DAKO Envision Plus detection was applied for 30 min also at room temperature. This was followed by application of diaminobenzidine chromogen. The slides were then counterstained in hematoxylin and coverslipped.

Outcome data on cases used for microarray construction was obtained from the University of Colorado tumor registry. Patients were followed for a median of 51 months (range, 18–100).

**Scoring of IHC Results.** Each core on the TMA was examined by conventional white light microscopy and the observed staining pattern graded for each core. Percentage of tumor cells positive and intensity of staining was recorded for both tumor cytoplasm and nucleus. A grading score was obtained by multiplying the intensity of staining on an arbitrary 0–4+ scale by the percentage of cells stained separately for nuclear and cytoplasmic staining. The same grading system was used for both TMA samples and lung cancer cell lines.

**RESULTS**

**OMA Detection of Overexpressed Genes.** Cluster analysis of microarray cell line experiments in which cell lines were compared with primary cultures of normal bronchial epithelium and homogenates of non-neoplastic lung tissue indicated that tumors clustered according to histological type, with the two SCLC lines clustering together, the two NSCLC lines clustering together, and each of the tumor types clustering separately from the cultured benign epithelium and lung tissue homogenates (Fig. 1). Each sample of non-neoplastic lung tissue invariably clustered with its corresponding duplicate sample. Complete hybridization data for cell lines, the primary culture, and the tissue homogenates are available.\(^7\)

The dual normalization and filtration process yielded 20 separate genes (Table 3) represented by 21 probe sets that were highly overexpressed by at least 2 of the cell lines. Expression levels were related to cell type with 14 genes overexpressed only in SCLC, 4 in NSCLC, and 2 in both SCLC and NSCLC. With the exception of the CTAG gene group, the chromosomal distribution of the overexpressed genes appeared to be random. A wide diversity of gene functions and subcellular localizations were represented among the gene products, ranging from a membrane-associated ion transport protein to nuclear transcription factors (Table 4). Four of the genes, ASH1, claudin 10, and claudin 10.

<table>
<thead>
<tr>
<th>Table 1 Cell line properties</th>
<th>Degradation index</th>
<th>Spectral karyotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell line</td>
<td>Histology</td>
<td></td>
</tr>
<tr>
<td>UMC19</td>
<td>SCLC</td>
<td>2.58</td>
</tr>
<tr>
<td>SHP77</td>
<td>SCLC</td>
<td>1.89</td>
</tr>
<tr>
<td>H647</td>
<td>NSCLC</td>
<td>2.01</td>
</tr>
<tr>
<td>AS49</td>
<td>NSCLC</td>
<td>1.84</td>
</tr>
<tr>
<td>Bronchial epithelium (× day culture, nonsmoker)</td>
<td>Benign Epithelium</td>
<td>1.42 2N (normal)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2 Stage and histology for TMA</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3a</th>
<th>Stage 3b</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squamous</td>
<td>52</td>
<td>14</td>
<td>16</td>
<td>9</td>
<td>91</td>
</tr>
<tr>
<td>Adeno</td>
<td>30</td>
<td>18</td>
<td>16</td>
<td>7</td>
<td>71</td>
</tr>
<tr>
<td>Large cell</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Bronchioloalveolar</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>35</td>
<td>35</td>
<td>18</td>
<td>187</td>
</tr>
</tbody>
</table>

\(^6\) Internet address: http://www.cbs.dtu.dk/services/SignalP-2.0.

\(^7\) Internet address: http://uch.uchsc.edu/uccc/research/GeneExpression/index.html.
and the secretogranins I and II, contained signal peptide sequences suggesting the possibility that the gene products are secreted. Only 1 gene, ABCC2, contained a signal anchor peptide sequence.

A notable feature of the list of overexpressed genes is the frequency of the CT group with 5 MAGE-A and the NY-ESO-1 CTAG genes constituting 30% of the total list (Fig. 2). Here again levels of expression were related to tumor type with SCLC lines more highly constituting 30% of the total list (Fig. 2). Here again levels of expression were related to tumor type with SCLC lines more highly expressed on the CT group with 5

Table 3. Genes highly overexpressed in lung cancer cell lines

<table>
<thead>
<tr>
<th>Gene (abbreviation)</th>
<th>Genbank no.</th>
<th>Affymetrix probe no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Confirmatory RT-PCR. By RT-PCR, the differences in expression patterns for several of the highly expressed genes were confirmed in 25 cell lines with SCLC, NSCLC, and mesothelioma histologies (representative gel shown in Fig. 3). As graphically depicted in Fig. 4, most of the high expression markers were detected in similar percentages of cell lines, but NY-ESO-1, MAGE-A10, and ASH-1 were
expressed frequently by SCLC but infrequently or not at all by NSCLC. By χ² analysis, differences in expression frequencies were significant at Ps of <0.02, <0.03, and <0.03 for MAGE-A, NY-ESO-1, and ASH, respectively.

Correlation of RT-PCR with IHC in Cell Lines. Of the 25 cell lines tested by RT-PCR, 19 were also tested by IHC. IHC staining patterns of cell lines were similar to those described below for TMA. Cytoplasmic and nuclear staining were frequently present in the same cells but were scored separately. There was nearly complete concordance between RT-PCR and IHC results with all of the specimens with variable intensity (Fig. 5) with labeling scores (described on page 13) separately and combined for prognostic significance. Staining was of highest levels were found in SCLC lines but strong concordance (labeling score was 1 or higher. Of the 187 arrayed tumor samples, 44% were more frequently positive than adenocarcinomas, large cell carcinomas, or bronchioloalveolar carcinomas (Table 5; χ², P < 0.00002).

As expected, there was a strong association between stage and survival among the patients with tumors represented on the TMA, but log rank test indicated no significant correlation between MAGE-A expression status and survival regardless of whether cases were stratified by tumor histology, stage, or by weak or strong (labeling score >100) nuclear staining, weak or strong (labeling score >100) cytoplasmic staining, or in those cases where there was discordance the positive staining was weak. There was also strong correlation between tumor histology and MAGE-A expression status. Squamous carcinomas were more frequently positive than adenocarcinomas, large cell carcinomas, or bronchioloalveolar carcinomas (Table 5; χ², P < 0.00002).

DISCUSSION

The schema created in this study for biomarker discovery included three steps: first, a few lung tumor cell lines were analyzed for gene expression using high-density oligonucleotide arrays. This step had the advantages that pure tumor cells could be tested without microdissection and that usage of microarrays was efficient and minimal. At this step, a stringent filtering algorithm was used incorporating both cultured epithelial cells as well as lung tissue homogenates as normalizing controls to limit the list of highly overexpressed genes to just 20. Remarkably, 30% of the genes that survived the filtration were
CTAG genes. The second step consisted of confirmation by RT-PCR of overexpression of selected candidate biomarkers in an expanded set of cell lines. This provided an independent and economical way for assessing expression profiles in cell lines of specific histological type. It also proved to be highly predictive of protein expression level. The third step was the application of specific monoclonal antibody to a large TMA that was linked to clinical follow-up. This provided a rapid way to assess prognostic importance of expression of the candidate biomarker. It also provided an opportunity to verify the distribution of protein at a cellular level. A step beyond this discovery schema would consist of testing blood, urine, and sputum for the presence of specific RNA and protein in clinical cohorts with and without invasive carcinoma or preinvasive lesions. Such clinical testing is expensive and time consuming, and before being launched requires the maximum rigor in estimating specificity of the putative biomarker. The schema used in this study efficiently answers many of the preliminary questions that should be answered before proceeding to larger clinical trials.

A surprising feature of the list of biomarkers that emerged from this schema was strong representation of CTAG genes. This gene group comprised 30% of the 20 biomarker genes identified. Expression of CTAG genes is restricted to normal testicular (and ovarian) germ cells and tumors of a number of cell types (26, 27). The first of these genes to be identified were the MAGE-A genes, which were originally named MAGE-1 through 3. They were discovered because they elicit an HLA I-dependent cytotoxic response in sensitized lymphocytes against the melanoma cell line M22-MEL (28). It is of interest that this first report indicated that MAGE expression could be demonstrated not only in melanoma cell lines but also in SCLC and NSCLC as well. The original 3 MAGE genes were soon supplemented by 9 additional MAGE genes discovered by screening cosmid libraries temporarily bringing the total number of MAGE genes to 12 (29), all encoded at chromosome Xq28 (29, 30). Sequencing of chromosomal region Xp21.3 led to the identification of a second subfamily of MAGE genes named MAGE-B (31–33). In recent years, the list of MAGE family genes has continued to increase, and the MAGE family now is thought to contain 55 homologous members divided into 9 subfamilies (34). Although structurally homologous, some recently described MAGE subfamilies are ubiquitously expressed (35–37) and are not members of the CTAG gene group. Early reports indicating that MAGE genes may be expressed by tumors of many types (28) have been confirmed in many different laboratories for many different types of tumors including brain (38, 39), skeletal muscle (40), esophageal and stomach (41), Reed-Sternberg cells (42), bladder (43), biliary tract (44), and breast (45).

A second CTAG gene family, NY-ESO-1, was identified by autologous screening of a cDNA expression library constructed from a case of esophageal carcinoma (26). A similar if not identical gene was reported a short time later as LAGE-1 (46). Like the MAGE-A gene family, NY-ESO-1/LAGE-1 maps to chromosome Xq28 (46).

Expression of CTAG genes by lung tumors has been documented by IHC and RT-PCR in a limited number of studies. In two separate studies using monoclonal antibodies 57B and MA454 that react with MAGE-A protein, Junghbluth et al. (47) have found heterogeneous expression in 32% and 56% (48) of NSCLC, respectively. IHC studies are complicated by the high degree of homology among different MAGE-A proteins so that many anti-MAGE antibodies cross-react with several different MAGE-A subfamily members (49–51). The monoclonal antibody used in the present study, 6C1, reacts with an epitope in the COOH-terminal regions of MAGE-A1, -A3, -A4, -A6, -A10, -A11, and -A12 (51) and may thus be considered an anti pan-MAGE-A reagent. The use of antibody in a sensitive immunoperoxidase procedure on 187 tumors in a TMA linked to clinical and histological data allowed us to determine that 44% NSCLC express MAGE-A protein, that expression varied according to tumor histology, and that expression is unrelated to prognosis.

By RT-PCR, MAGE-A1, -A3, and B2 RNA sequences have been found recently in 70%, 85%, and 85% of a small series of NSCLC and is often accompanied by promoter hypomethylation (52). Also of interest, this report indicated that bronchial epithelium from a large proportion of 20 former smokers without lung carcinoma also frequently expressed MAGE genes and suggested that MAGE gene

### Table 5 Histology versus MAGE-A status

<table>
<thead>
<tr>
<th>Histology</th>
<th>MAGE-A status</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Squamous carcinoma</td>
<td>57</td>
<td>34</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>19</td>
<td>52</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Bronchioalveolar carcinoma</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Totals</td>
<td>80</td>
<td>107</td>
</tr>
</tbody>
</table>

Fig. 4. Graphic representation of RT-PCR results for tumor cell lines of various cell types using primers for genes found to be overexpressed at high level by OMA. Markers with specificity for SCLC include NY-ESO-1, MAGE-A10, and ASH1.

Fig. 5. TMA stained for MAGE-A proteins using the monoclonal antibody 6C1 in a sensitive immunoperoxidase method. On the left is an image of the microarray slide and on the right a high magnification of tumor cells exhibiting strong (4+) nuclear and weaker (2+) cytoplasmic staining.
expression may occur early in lung carcinogenesis and may be a suitable target for lung cancer prevention.

Whether MAGE-A protein can be found in the blood, urine, or sputum of patients with lung cancer is not known at present. The absence of signal sequences in the MAGE-A genes suggests that MAGE-A proteins are not actively secreted. However, it may not be essential that a protein be actively secreted to be useful as a biomarker because protein may be released from dying tumor cells, which are frequent in lung carcinomas. Also, tumors that occur frequently in central airways (SCLC and squamous carcinomas) are most often MAGE-A/ANY-ESO-1-positive, suggesting that these biomarkers may be particularly useful for sputum testing.

An advantage of OMA analysis for potential biomarkers is the ability to interrogate microarray data for expression patterns of all or many members of entire functional pathways. In this context it is of interest that melanoma genes other than the CTAG genes were found not to be overexpressed in lung cancer lines indicating that there are significant differences in activation of functional pathways between these two tumor types. The function of CTAG genes in general and MAGE-A genes in particular is not known. Necdin, a 325 amino acid protein with 30% homology to MAGE proteins (reviewed in Ref. 53) has been shown recently to interact with p53, inhibiting p53-induced apoptosis. Whether or not the MAGE proteins function in a similar way is unknown at present but such a function would be consistent with the frequent expression of MAGE proteins in aggressive malignancies.

Several of the remaining highly overexpressed genes have properties that suggest they may be useful biomarkers including signal peptide coding sequences. One of four protein products of genes containing signal peptides has been tested as a lung cancer biomarker. Chromogranin A has been found in the serum of 50% (14) of all of the neuroendocrine tumors and 61–70% (15, 54) of SCLC, and is generally regarded as a promising marker for the diagnosis of neuroendocrine neoplasia. A second putatively secreted protein, ASH1, has been associated previously with neuroendocrine neoplasia (55) but has not been tested as a serum or urinary biomarker. That this protein is in fact secreted is doubtful, given its role as a transcription factor and its nuclear localization. Other overexpressed genes associated with neuroendocrine differentiation in pulmonary neoplasia include IA-1 (56) and DOPA decarboxylase (57, 58).

Many of the listed genes may be useful for detection and monitoring of NSCLC, or both SCLC and NSCLC. These include aldo-keto reductase family 1, ABCC2 (59), basic hair keratin 1, prostaglandin E synthase (60), PGP 9.5 (61, 62), Na+, and K(+)−ATPase (63–65). Additional evaluation of these candidate biomarkers either alone or in combination will be required to establish the utility of these overexpressed genes as useful biomarkers for early detection and monitoring, and to better define their biological role. Better understanding of expression profiles for these genes may also suggest novel approaches to therapeutic intervention lung carcinoma.

Several other large-scale gene expression analyses of lung cancer have been reported recently. An OMA analysis of 186 pulmonary tumors also using the HG-U95A microarray has been performed recently on tumor homogenates (22). Cluster analysis of the resulting data indicated that gene expression profiles corresponded to histological type for SCLC, squamous carcinoma, and carcinoid tumor, but adenocarcinomas were heterogeneous and could be subdivided into five categories including one for metastasis from colon. Three of the genes identified in the prior study also appear on the present list of overexpressed genes, ASH1, IA-1, and DOPA decarboxylase. Another analysis of tumor homogenates using 24,000 element cDNA microarrays has also been published recently (66). From among the several hundred genes with expression patterns that discriminate among differing histological types, only four genes were also present in the current list of overexpressed genes, PGE synthase, cMOAT, ASH-1, and IA-1. Finally, in a recent SAGE analysis (67), 115 highly differentially expressed genes were reported and among these was the aldo-keto reductase, member B10 gene. CTAG genes were not listed in any of these large-scale gene expression studies.

This small number of overlapping genes between the current and other recent studies has several possible explanations. First, the number of specimens examined is smaller in the present analysis than in the other analyses. Second, we attached only limited importance to intertumor heterogeneity of gene expression in this current study. We assumed that lung cancers, because they are heterogeneous in almost every respect, are likely to also exhibit a high degree of heterogeneity in gene expression profiles. We expected that there would be a great imbalance of many cellular pathways engendered by chromosomal and genetic instability, potentially resulting in high levels of overexpression of specific genes. Our objective was to identify these genes and to estimate the likelihood that their products could serve as tumor biomarkers. Finally, we made no distinction among tumors of various histological origins in screening for highest level overexpression. This allowed us to focus on genes that are massively overexpressed in cancer cells in comparison to normal lung regardless of the cell type, an approach specifically tailored for biomarker discovery.

We conclude that the detailed gene expression data can now be readily obtained using OMAs. Testing of even a few suitable specimens can identify potential biomarkers for lung and other cancers that can be rapidly validated by high throughput testing of TMA linked to clinical outcome. This model should be a rich source of promising new biomarkers. To exploit these new analytical tools it will be imperative that correlative biological materials be collected during large-scale and treatment trials that are currently being designed.

ACKNOWLEDGMENT

We thank Dr. Mariellea Varella-Garcia for spectral karyotype analysis of the bronchial epithelial cell culture.

REFERENCES

CANCER/TESTIS ANTIGEN EXPRESSION IN LUNG CANCER


Combined Use of Oligonucleotide and Tissue Microarrays Identifies Cancer/Testis Antigens as Biomarkers in Lung Carcinoma

Michio Sugita, Mark Geraci, Bifeng Gao, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/14/3971

Cited articles
This article cites 66 articles, 27 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/14/3971.full.html#ref-list-1

Citing articles
This article has been cited by 16 HighWire-hosted articles. Access the articles at:
/content/62/14/3971.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.