Classification of Lung Cancer Patients and Controls by Chromatography of Modified Nucleosides in Serum

John E. McEntire, Kenneth C. Kuo, Mark E. Smith, David L. Stalling, Jack W. Richens, Jr., Robert W. Zumwalt, Charles W. Gehrke, and Ben W. Papermaster

ABSTRACT

A wide spectrum of modified nucleosides has been quantified by high-performance liquid chromatography in serum of 49 male lung cancer patients, 35 patients with other cancers, and 48 patients hospitalized for nonneoplastic diseases. Data for 29 modified nucleoside peaks were normalized to an internal standard and analyzed by discriminant analysis and stepwise discriminant analysis. A model based on peaks selected by a stepwise discriminant procedure correctly classified 79% of the cancer and 75% of the noncancer subjects. It also demonstrated 84% sensitivity and 79% specificity when comparing lung cancer to noncancer subjects, and 80% specificity and 55% specificity in comparing lung cancer to other cancers. The nucleoside peaks having the greatest influence on the models varied dependent on the subgroups compared, confirming the importance of quantifying a wide array of nucleosides. These data support and expand previous studies which reported the utility of measuring modified nucleoside levels in serum and show that precise measurement of an array of 29 modified nucleosides in serum by high-performance liquid chromatography with UV scanning with subsequent data modeling may provide a clinically useful approach to patient classification in diagnosis and subsequent therapeutic monitoring.

INTRODUCTION

Biochemical substances which may serve as tumor markers in body fluids have been the subject of several reviews (1-7). Numerous molecules found in urine and serum have been studied as potential lung cancer markers, including carcinoembryonic antigen, neuron-specific enolase, creatine kinase BB, phosphohexose isomerase, lipid-bound and total sialic acid, ferritin, β₂-microglobulin, peptide hormones, tissue polypeptide antigen, lung tumor-associated antigens, polynucleotides, and nucleobases and nucleosides (8-10). Combinations of these have been utilized in "multiple markers" approaches for classification of patients with lung cancer, benign respiratory disease, and normal individuals, and to distinguish lung cancer types and establish correlations with disease stage (9, 10).

Modified ribonucleosides, derived predominantly from tRNA and rRNA, are known to be excreted in abnormal amounts in the urine of cancer patients (8, 11-18; for reviews see Refs. 1 and 19). These excretion products include specific methylated nucleosides and pseudouridine (8, 16, 20, 21). Five urinary modified nucleosides were evaluated by Waalkes et al. (8) as potential tumor markers for patients with small cell carcinoma of the lung, and the total frequency of elevated values for the five nucleosides were reported to be directly related to stage of disease. In 1986, Tamura et al. (17) reported elevated urinary pseudouridine in 33% of patients with limited small cell lung cancer and elevations in 77% of patients with extensive disease. There was a clear indication that the pseudouridine levels followed the course of the disease and the authors concluded that although pseudouridine is not a specific marker for SCC it does reflect tumor burden and indicates the clinical status of the patients.

Gehrke and Kuo (22, 23) recently reported methods for serum and urine nucleoside analysis by HPLC-UV. They have applied these methods to cancer patients and normal individuals, and discussed the advantages and disadvantages of determining serum nucleoside levels as compared to urinary levels. This paper describes the application of their method for serum modified nucleoside analysis to a case-control study of cancer patients and hospitalized controls and the subsequent data analysis techniques for classification of patients.

MATERIALS AND METHODS

Sample Acquisition. The study sample consisted entirely of male patients who were part of an experimental program involving case-control analysis of a lung cancer cohort, from a population in a region of high lung cancer mortality in southwestern Missouri. Epidemiological data were obtained by interview from lung cancer patients and age and sex matched controls, and several biological markers in blood were analyzed from the same study group. Of the controls, one group consisted of patients hospitalized for cancers other than lung. The second control group consisted of age and sex matched patients hospitalized for nonneoplastic diseases. Blood samples were allowed to clot and the serum was decanted and stored at −70°C until used for analysis. Samples from patients having genitourinary cancers with potential impaired renal function or urinary obstruction were not used in this analysis. The remaining patients had no apparent renal dysfunction as determined by analysis of clinical records and had normal serum creatinine levels. Most of the patients also had 24-h urine collected upon admission and there were no apparent extremes in the volumes recorded.

Chromatographic Methods. Chromatography was performed as described previously (22, 23). Nucleosides were isolated from serum with the use of phenylboronate gel columns. Serum (1.0 ml) was aliquoted into a 1.5-ml polypropylene centrifuge tube. The internal standard m²Urd was added at a concentration of 0.5 nmol/ml of serum; the solution was vortexed and filtered through a M, 25,000-30,000 cutoff membrane (YMT type, Amicon Corp.) by using a Centricon-10 microconcentrator (Amicon Corp.) and a 30-degree fixed-angle rotor centrifuge at 3500 × g. Ammonium acetate (250 μl of a 2.5 m solution, pH 9.0) was added to the ultrafiltrate and mixed well. The sample was transferred onto a washed, conditioned, and preequilibrated boronate gel column. The gel column was washed with 3 ml of 0.25 M ammonium

Received 7/26/88; revised 11/3/88; accepted 11/10/88.

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.

1 Supported in part by NIH Grant 1R43CA-45844-01 and by State of Missouri Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' Contract C86719. Additional private funds were provided by the Cancer Research Center, the Missouri Order of Eastern Star, and the Fraternal Order of Eagles' 2 To whom requests for reprints should be addressed, at Tektagen, Inc., 358 Technology Dr., Malvern, PA 19355.
acetate, pH 8.8, then washed with 300 μl of 50% methanol/water (v/v). The nucleosides were eluted with 4 ml of 0.01 N formic acid in 50% methanol/water (v/v), and collected in a polypropylene tube. Methanol was removed from the nucleoside sample by using a centrifugal evaporator, and the sample was then frozen and lyophilized to dryness. The sample was then redissolved in 200 μl of HPLC water, and 180 μl was injected into HPLC.

**Apparatus.** The HPLC system used was an HP 1090M with a LC chromatography chemical station (Hewlett-Packard, San Diego, CA) composed of an autosampler and autoinjector, ternary gradient pump, temperature-controlled column oven with circulating refrigerated coil, HP 1040A photodiode array UV detector, and data storage system.

Reagents. HPLC water was obtained through reverse osmosis, ion exchange, and organic adsorption. The methanol and acetonitrile used were distilled-in-glass grade with UV cutoff below 190 nm (Burdick and Jackson Laboratories, Muskegon, MI). Other analytical reagent grade chemicals used were: ammonium phosphate, zinc sulfate, (J. T. Baker Chemical Co., Phillipsburg, NJ); ammonium hydroxide, phosphoric acid, (Malinnckrodt Co., St. Louis, MO); Tris base (Sigma Chemical Co., St. Louis, MO).

Nucleosides were identified by comparison to standards which were obtained from several sources (Sigma Chemical Co.; Mann Research Laboratories, New York, NY; P-L Biochemicals, Milwaukee, WI; and Vega Biochemicals, Tucson, AZ).

**HPLC Conditions.** The HPLC buffers used are: (a) 2.5% methanol/0.01 M NH₄H₂PO₄, pH 5.3; (b) 20% methanol/0.01 M NH₄H₂PO₄, pH 5.1; and (c) 35% acetonitrile/0.01 M NH₄H₂PO₄, pH 4.9. The column used was a Supelco LC-18S, 150 x 4.6 mm. Chromatography was performed at 26°C. Columns were washed for 20 min with 70% methanol/water and equilibrated with starting buffer for 1 min/cm of column length. The chromatography was according to a multistep gradient which has been previously described by Gehrke and Kuo (23).

**Data Analysis.** In analyzing the data our primary interests were to determine which, if any, nucleosides correlated with lung cancer, or other cancers, and were different in noncancer controls. To accomplish this, discriminant analysis was performed by using the procedures DISCRIM and STEPDISC of the SAS. Discriminant analysis is a method for analyzing a set of data containing one or more quantitative variables and a classification variable that defines different groups of observations. During the analysis, a function is developed that is used to determine the value of the classification variable (lung cancer, other cancer, and noncancer), using the quantitative variables (nucleosides). The observations are then reclassified by using the discriminant function. The accuracy of the model can be measured by its sensitivity (the number of true positives classified as positive), and by its specificity (the number of true negatives classified as negatives). The function can then be used to classify another set of observations containing variables of the same type. Discriminant analysis was carried out along three different lines. The three approaches differed in the selection of the nucleosides to be used in constructing the model. (a) Construct a model with 24 nucleosides (5 nucleosides were eliminated from consideration because of their sparseness); (b) construct a model with 1 of 3 nucleosides, pseudouridine m²Gguo, or t²Ado as is often reported in the literature; and (c) construct a model by using nucleosides selected by a stepwise discriminant procedure.

In each approach split samples were utilized to test the accuracy of the discriminant model. The data were divided into two sets of approximately equal size: the calibration set and the test set on the basis of uniform random deviates. The calibration set was used to derive the discriminant function. The test set was then used to determine the accuracy of the resulting function.

**RESULTS AND DISCUSSION**

**Sample Acquisition and Description.** Initial evaluation of data with all patients included identified a group of outliers by application of principal component analysis. Upon examination, these outlying data patterns were from patients having impaired renal function, so they were eliminated as a group.

The remaining patients had no apparent renal dysfunction. Samples were collected and analyzed from the remaining 49 lung cancer patients, 35 patients with nonsmoking-related cancer, and 48 hospitalized control patients. All sera were from males. The three groups were of comparable age distribution as determined by contingency table analysis (P > 0.27).

**Chromatographic Analysis of Nucleosides from Sera.** Sera were analyzed as described in “Materials and Methods.” Peak heights were determined and normalized to an internal standard (m²Urd). A number of the chromatographic peaks which were used in this analysis are structurally unidentified nucleosides, and therefore, there were no molar response factors to allow peak height to be converted to concentration. Therefore, normalized peak heights were used as units for further analysis of all known and unknown peaks. A typical chromatographic analysis with peaks identified is presented in Fig. 1. Data plots of representative nucleosides are presented in Fig. 2, which shows the distribution of normalized peak heights of selected nucleosides for case and control groups. For nucleosides ac²Cyd, m²²Guo, t²Ado, and pseudouridine, the mean peak heights were elevated in both lung cancer and other cancers as compared to noncancer controls (P < 0.05). The mean peak height of unknown number 9 was also elevated in lung cancer compared to the control cancer group and noncancer controls (P < 0.04).

As discussed by Kuo et al. (22), there are several potential advantages of serum over urine as a source of nucleosides for studies of this type. Serum volume is directly related to total surface area of the body. This allows for direct comparison of data in terms of concentration rather than normalizing on the basis of another molecule such as creatinine as is usually required in studies of urine. Additionally, serum nucleosides may be subject to fewer structural alterations than urinary nucleosides. This may account for higher serum levels of some modified nucleosides. Finally, availability of serum samples, ease of collection, and physician preference for serum values of other biological markers favor use of serum over urine and bring serum nucleoside measures a step nearer the goal of a clinical assay.

**Data Analysis.** Results of the data analysis are presented in Tables 1–6. The ability of a model constructed from all 24 evaluative nucleosides to classify subjects into either cancer or noncancer groups is shown in Table 1. In comparing cancers to noncancer controls, this procedure correctly classifies 79% of the cancer patients but only 62% of the noncancer patients. Similar results are obtained when lung cancer is compared with noncancer and when cancers other than lung are compared with noncancer controls. In comparing lung cancer with other cancers, 76% of the lung cancers are correctly classified but only one-half of the other cancers are distinguishable from lung cancer.

The data analysis methods used in this study were selected to take advantage of the high resolving ability of the chromatographic method. In the present study, for example, the data base consists of 132 samples and 29 nucleosides, and thus contains some 4524 elements. The ability to classify sera from cancerous and noncancerous patients correctly is clearly possible with discriminant analysis. Additionally, classification of lung cancer versus other cancers and controls, as well as specific types of lung cancer such as small cell cancer, are possible with these techniques. The relative success of discriminant analysis in constructing appropriate models with this limited set of samples points out the need to apply these methods to larger
sets of data. An increased number of patient samples may lead to more accurate predictive models.

Classification by using a model based on only one nucleoside is shown in Table 2. The nucleoside selected in this case was m22Guo. Two other nucleosides were also examined in this manner (pseudouridine and t6Ado) without any significant differences being observed. This comparison does not allow useful classification of groups. Although the noncancers are correctly classified at 75% specificity, cancer, lung cancer, and other cancers are poorly differentiated (39-42% sensitivity). Again, lung cancer is not readily distinguished from other cancers by using a one nucleoside model.

Nucleosides selected with a stepwise discriminant method were used to construct an additional model. Each comparison resulted in a different set of nucleosides being selected. The selected nucleosides and the resulting sensitivity and specificity of the model are reported in Table 3. The stepwise discriminant technique most accurately classifies the various types of cancer and noncancer. It also correctly classifies 80% of the lung cancer compared to other cancers, but because of the heterogeneity of both the lung cancer and the other cancer groups, can still only correctly classify 56% of non-lung cancer group. Different nucleosides are shown to be important in each classification, suggesting that each of the heterogeneous cancer groups may be made up of distinguishable smaller subsets.

To test this possibility that the heterogeneous lung cancer case group was comprised of smaller homogeneous groups, small cell cancer was compared to other lung cancers. Table 4 shows similar discriminant analysis applied to SCC versus other lung cancer. As in the previous analysis, a stepwise discriminant procedure was applied first to select a set of nucleosides with the most predictive power. A full discriminant analysis was then conducted with a model containing the nucleosides indicated by stepwise analysis. Since only 9 SCC patients were available for this study, the split-sample technique was not used. This method very accurately classifies SCC and other lung cancer with 78% sensitivity and 85% specificity, suggesting that small cell lung cancer may compose a group distinguishable from other lung cancers in its modified nucleoside pattern. The peaks selected by the stepwise procedure for use in discriminant analysis were m1Ado, acCyd, and unknown peaks 7, 9, 21, 23. The heterogeneity of the lung cancer group as well as that of the cancer group in general may be responsible for the observed overlap in the data sets. The original heterogeneous group may be made up of more homogeneous segments constituting histologically similar tumors. If so, patterns of modified nucleosides in sera may be specific for histological types and of potential value in future diagnostic batteries. However, other factors such as extent of disease were not considered, and in fact, may vary considerably between histological types. Thus, no definitive statements relating to SCC can be made until more sera are analyzed. We are accumulating additional sera from patients with histologically defined lung cancers.

The advantage of a larger data base of modified nucleosides for various analytical techniques is best observed in comparing the classification of groups by using a single nucleoside level as is usually reported in the literature. The problem is readily illustrated in Fig. 2. Although several nucleoside means are elevated in cancer, the range of values is widely distributed and not useful for accurate classification. Using data obtained in this study, classification of the cancer versus normal groups is not very successful when the classification was based on only one nucleoside (e.g., m22Guo). Only 42% of the cancers were classified correctly with 75% specificity. The other group comparisons were similarly of low sensitivity. A more accurate model was generated by using all 24 of the evaluable nucleosides. This model was 72-79% sensitive with specificity ranging from 50 to 75%, depending on the groups compared.

When comparing lung cancer to noncancer, the model is 84% sensitive and 79% specific. Similar results are obtained when cancer excluding lung are compared to controls. Using the same procedures, 80% of the lung cancer patients could be placed in the correct group when compared to other cancers, although the other cancers still overlapped with the lung cancer group significantly (56% specificity). Each of these comparisons selects different nucleosides for inclusion in the final analysis.

Using the HPLC technique described, it is technically no more difficult, nor time consuming to quantify 24 nucleosides from a single chromatographic run than it would be to quantify only a few. We are currently attempting to determine the minimum set of nucleosides required for similar classification studies which use less technically demanding assay techniques such as immunoassays.

Because of the varied clearance rates of the modified nucleosides, it is difficult to compare our data obtained from serum directly with previously published studies of urine nucleosides and cancer. As in studies of modified nucleosides in urine, our
Fig. 2. Distribution of normalized peak heights of selected nucleosides for case and control groups. Peak heights were normalized to the height of the internal standard and this ratio was plotted by using the "sunflower" plot available on Statview 512+ (BrainPower, Inc. Calabasas, CA). In this graphic presentation of the data, each petal of the sunflower represents a point counted in the region. Horizontal bars, means for each group. For nucleosides ac'Cyd, m22Guo, t'Ado, and pseudouridine, the mean peak heights were elevated in both lung cancer and other cancers as compared to noncancer controls (*P < 0.05). The mean peak height of unknown number 9 was also elevated in lung cancer compared to the control cancer group and noncancer controls (*P < 0.04).

Table 1: Sensitivity and specificity of models constructed by using a profile of 24 nucleosides

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer versus noncancer</td>
<td>79.1</td>
<td>62.5</td>
</tr>
<tr>
<td>Lung cancer versus noncancer</td>
<td>72.0</td>
<td>66.7</td>
</tr>
<tr>
<td>Other cancer versus noncancer</td>
<td>72.2</td>
<td>75.0</td>
</tr>
<tr>
<td>Lung cancer versus other cancer</td>
<td>76.0</td>
<td>50.0</td>
</tr>
</tbody>
</table>

Table 2: Sensitivity and specificity of models constructed with m22Guo

Data were collected, normalized, and analyzed as described in "Materials and Methods" and Table 1. One half the data from a single nucleoside peak, m22Guo, was used to build a model using discriminant analysis and the other half was used to test the model.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer versus noncancer</td>
<td>41.9</td>
<td>75.0</td>
</tr>
<tr>
<td>Lung cancer versus noncancer</td>
<td>40.0</td>
<td>75.0</td>
</tr>
<tr>
<td>Other cancer versus noncancer</td>
<td>38.9</td>
<td>75.0</td>
</tr>
<tr>
<td>Lung cancer versus other cancer</td>
<td>60.0</td>
<td>33.3</td>
</tr>
</tbody>
</table>

The data show the elevation of most nucleosides in cancer patients as compared to controls.

Scoble et al. (24, 25) have demonstrated the applicability of nucleoside HPLC and discriminate analysis of data for classification of leukemia patients. In these studies, acute lymphocytic leukemia plasma samples and controls were classified with 100% sensitivity and specificity while chronic leukemia samples...
niques can accurately classify sera from cancer patients and controls. There was adequate sensitivity and specificity to demonstrate the feasibility of applying these techniques in clinical studies of cancer patients. This study differs from most of the modified nucleoside analyses in the literature in several areas. First, it is a case-control study involving 132 patients and carefully matched controls. Second, it measures not just 1 or 2 modified nucleosides, but quantifies 29 nucleoside peaks from serum, 24 of which are useful in various classification models. Third, it utilizes data analysis techniques which can take advantage of the large data set generated by the study. As the number of patient samples increases and the data models become more clearly defined, these methods may very well have application in diagnostic test batteries, distinguishing tumor sites or types, or in monitoring the therapeutic course of patients.

ACKNOWLEDGMENTS

The authors wish to thank cooperating physicians in the Joplin, Missouri area, especially Drs. Mark J. Skelly, Robert L. Carter, James H. Habermann, and their nursing and office staffs for their help in obtaining patient sera, medical records, and interviews. We also wish to thank the administrators and staff of St. John’s Hospital, Joplin, Missouri, for access to patients and use of facilities.

REFERENCES

MODIFIED NUCLEOSIDES IN LUNG CANCER

Classification of Lung Cancer Patients and Controls by Chromatography of Modified Nucleosides in Serum

John E. McEntire, Kenneth C. Kuo, Mark E. Smith, et al.