Importance of Serum SELDI-TOF-MS Analysis in the Diagnosis of Early Lung Cancer

Cebrail Şimşek¹, Özlem Sönmez¹*, Ahmet Selim Yurdakul², Füsun Özmen³, Nurullah Zengin⁴, Atilla İhsan Keyfi⁴, Dilek Kubilay⁵, Özlem Gülbahtar⁶, Senem Ceren Karataylı⁷, Mithat Bozdayi⁷, Can Öztürk²

Abstract

**Background:** Different methods of diagnosis have been found to be inefficient in terms of screening and early diagnosis of lung cancer. Cancer cells produce proteins whose serum levels may be elevated during the early stages of cancer development. Therefore, those proteins may be recognized as potential cancer markers. The aim of this study was to differentiate healthy individuals and lung cancer cases by analyzing their serum protein profiles and evaluate the efficacy of this method in the early diagnosis of lung cancer. **Materials and Methods:** 170 patients with lung cancer, 53 under high risk of lung cancer, and 47 healthy people were included in our study. Proteomic analysis of the samples was performed with the SELDI-TOF-MS approach. **Results:** The most discriminatory peak of the high risk group was 8141. When tree classification analysis was performed between lung cancer and the healthy control group, 11547 was determined as the most discriminatory peak, with a sensitivity of 85.5%, a specificity of 89.4%, a positive predictive value (PPV) of 96.7% and a negative predictive value (NPV) of 62.7%. **Conclusions:** We determined three different protein peaks 11480, 11547 and 11679 were only present in the lung cancer group. The 8141 peak was found in the high-risk group, but not in the lung cancer and control groups. These peaks may prove to be markers of lung cancer which suggests that they may be used in the early diagnosis of lung cancer.

Keywords: Lung cancer - biomarkers - proteomic - SELDI - diagnostic tools - bioinformatics

Introduction

Lung cancer is one of the common tumors that lead to millions of deaths worldwide (Parkin et al., 1999). Despite all researches, the prognosis of lung cancer is still poor. Although surgical therapy is applied at the early period, 5-year-survival rate is known to be 67% for Stage IA and 57% for Stage IB. Mountain (2000). In general, the average 5-year-survival rate for patients with lung cancer is only 15% (Spira andEttinger, 2004). Many of the lung cancer cases are locally advanced or remotely metastasized at the time of the diagnosis (Han et al., 2008).

In light of these data, it is obvious that establishing an early diagnosis bears importance in terms of providing a cure and extending the survival time. Currently, chest X-ray, low-dose computed tomography, PET/CT, bronchoscopy, sputum cytology, as well as tumor markers such as carcinoembryogenic antigen (CEA), neuron-specific enolase (NSE), cyto keratin 19 (Cyfra 21-1), and squamous cell carcinoma antigen are employed in the diagnosis of lung cancer (Kulpa et al., 2002; Zhong et al., 2004; Yang et al., 2005; Huang et al., 2006). However, these methods have been found to be inefficient in terms of screening and early diagnosis (Fontana et al., 1984; Melamed et al., 1984; Jett, 2002). Using CT and PET/CT for screening and early diagnosis is impossible due to costly nature and high radiation doses. Bronchoscopy is not appropriate either because of its invasive character. The reported specificities and sensitivities of tumor markers are not adequate for early diagnosis, as well (Schneider et al., 2000; 2002). Therefore, there is a need for novel diagnostic tests with higher sensitivity and/or specificity in the early diagnosis of lung cancer.

Recently, the advances in the molecular and genetic fields have enabled several noninvasive analyses on blood, sputum, and similar specimens for the early diagnosis of cancer. Cancer cells, like other cells, produce proteins and secrete them into the blood stream. The serum levels of...
these proteins may be elevated during the early stages of the cancer. Therefore, those proteins may be recognized as potential cancer markers and used in the screening, follow-up, and prognosis of cancer cases (Wu et al., 2005; Han et al., 2008).

The notion of proteome was first expressed by Marc Wilkins at the beginning of 1990s which was identified as the entire range of proteins secreted by a tissue or a genome (Clarke, 2003; Huber, 2003). Similarly, it is also described as the whole spectrum of proteins produced in a cell, tissue, or an organism. On the other hand, proteome can be explained as the systematic analysis of those proteins with regard to their functions, numbers, and types. (Pen and Gyg, 2001). Apart from genome, proteome is a dynamic set of proteins varying relative to cell types, variants of the same disease, and individuals.

The basis of early diagnosis by protein analysis in lung cancers is the evaluation of sputum, serum, pleural fluid, bronchial lavage, and tissue biopsy samples for proteins secreted by tumor cells. Proteomic method uses mass spectrometry, protein microarray, and electrophoresis for protein analysis. However, the main technique that should reveal the cancer biomarkers is mass spectrometry (Alessandro et al., 2005; Yıldız et al., 2007; Sung and Cho, 2008).

There are two types of mass spectrometry techniques: Matrix-Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry (MALDI-TOF-MS) and Surface-Enhanced Laser Desorption / Ionization Time-of-Flight Mass Spectrometry (SELDI-TOF-MS). These techniques have contributed significantly in the diagnosis of breast, ovarian, prostate, pancreatic, lung, and hepatic cancers since 2000s (Adam et al., 2002; Petricoin et al., 2005; Yıldız et al., 2007; Gamez-Pozo et al., 2009). Phosphoproteomics is a promising technology for the identification of biomarkers and novel therapeutic targets for cancer (López et al., 2012).

In the early diagnosis of lung cancer; detection of molecular changes, prediction of individuals under high risk for cancer development, and identification of tumor cells are all important steps. The next step is to establish a reliable screening test.

The aim of this study is to differentiate healthy individuals and lung cancer cases by analyzing their serum protein profiles via SELDI-TOF-MS method, and evaluate the efficacy of this method in the early diagnosis of lung cancer.

Materials and Methods

In total, 270 individuals (170 patients diagnosed as lung cancer, 53 individuals under high risk of lung cancer, and 47 healthy people) who presented to the, Gazi University School of Medicine, Department of pulmonary medicine, Atatürk Chest Diseases and Thoracic Surgery Education and Research Hospital, and Numune Hospital, were included in our study.

High-risk patients were defined as having a 40 pack-year history of smoking or family history of cancer or personal history involving cancer in other organs. The control group was consisted of volunteered individuals who were not smokers and had no history of cancer in other organs. Each participant provided a written informed consent prior to the study. Our study was approved by the local ethics committee.

Prior to the treatment, a venous blood sample of 10cc was obtained from each case. These samples were centrifuged at 2500 rpm for 13 minutes and stored at -80°C until the time of analysis.

The proteomic analysis of the samples were performed by SELDI-TOF-MS method.

Furthermore, tumor markers such as CEA (Immunospec-CA, USA, EI-207),NSE (Immunospec-CA, USA, EI-214) and Cyfra 21-1 (DRG instruments, Marburg, Germany, EIA-3943) were studied by ELISA method in the lung cancer and high-risk groups at the Biochemistry Department of Gazi University.

SELDI-TOF analysis

Regarding to number and resolution of the protein peaks, three different ProteinChip surfaces (cationic, anionic, and Cu metal binding, Ciphergen Biosystems, Fremont, CA, USA) were tested. The IMAC30 (Cu metal binding surface) protein chip, which displayed the best serum profile, was selected for further analysis. Briefly, 5 µl of each serum sample was denatured by addition of U9 solution (9M urea, 2% CHAPS and 150 mM Tris-HCl, pH 9) in a 1:3 ratio and mixed on a platform shaker for 35 min at 4°C. The array spots were pre-equilibrated twice with 100 mM CuSO₄ for 5 min at room temperature, followed by three washes with 1 mM Heps (pH 7) for 2 min. Arrays were then incubated three times with 150 µl of binding buffer (PBS 1X, 0.25 M NaCl, 0.1% Triton X-100) for 5 min. Ten microliters of each diluted serum samples was randomly added on preactivated spots with 120 µl of binding buffer for 1 hr. Each array was then washed two times with binding buffer (5 min for each), then without Triton-binding buffer and washed twice with 1 mM HEPES. The air-dried arrays were saturated with sinapinic acid matrix (Ciphergen Biosystems) in 0.5% trifluoroacetic acid and 50% acetonitrile before being read on the instrument SELDI ProteinChip System 4000 Mass spectrometer (Ciphergen Biosystems). The data was analyzed with Ciphergen Express software, version 3.0 (Ciphergen Biosystems). All the raw data with protein peaks ranging from 2000-35000 Da was further normalized to total ion current and aligned.

Reproducibility analysis of SELDI protein spectra

To evaluate machine reproducibility, serum sample of a healthy volunteer was applied onto different spots on every chip run as a quality control (QC). In these protein profiles several independent control peaks were identified to calculate the CV (coefficient of variance) of intensity and CV of m/z in intra-assay and inter-assay.

Statistical analysis

Protein peaks were clustered with the Ciphergen Express software, version 3.0 performing Expression Difference Mapping (EDM). In all spectra autodetect peaks labeling was achieved with 5.0 S/N (signal to noise ratio), 3
valley depth for the first pass; minimal peak threshold: 20\% of all spectra and 3.0 S/N. 1 valley depth for the second pass with 0.2\% cluster mass window and the estimated peaks were added. Discriminatory peaks, depending on peak intensity, were identified using the Mann–Whitney non-parametric test. To determine the best discriminative proteomic index receiver operating characteristics curves (AUROC) was used as a measurement. Statistically significant discriminatory peaks between groups were determined with AUROC>0.7 and p<0.05. Patients with lung cancer were compared with healthy control and high risk groups using tree classification (cross-validation) with the Biomarker Patterns Software (BPS), version 5.0 (Ciphergen Biosystems). Briefly, the classification tree split the data into two nodes using one rule at a time in the form of peak intensity. The splitting decisions were based on the normalized intensity levels of peaks. The process of splitting was continued until terminal nodes are reached.

Results

Reproducibility of SELDI system

The reproducibility of the instrument was demonstrated using 4 peaks from a QC sample. For all peaks the coefficients of variation for mass accuracy was 0.03\%. The CV calculated for intra-assay and inter-assay intensities were 25\% and 28\%, respectively.

Identification of serum proteomic features associated with lung cancer, healthy control and high risk groups

When analysis was performed between lung cancer and healthy control groups, 9 peaks, 11679, 11480, 11547, 11899, 11740, 5832, 6438, 3951, and 6646m/z were found to be discriminatory, with P values less than 10^{-4}. The same 9 peaks were also discriminatory for lung cancer and high risk groups. In addition to these peaks, 3 more peaks of 5907 m/z, 6115 m/z and 8141 m/z were found to be discriminative for lung cancer and high risk (Table 1). The comparison between high risk and healthy control groups revealed 3 discriminative proteomic features of 6115 m/z, 8597 m/z and 8141 m/z (Table 2). The peak of 8141 m/z showed higher intensity, and the peak of 6115 m/z showed lower intensity in the serum of patients with high risk when compared with serum proteomes of both lung cancers and healthy controls (Table 1 and 2). SELDI-TOF spectra of patients with lung cancer, healthy control and high risk are shown for the most discriminatory peaks of 11480 m/z, 11547 m/z and 11679 m/z with p values less than 10^{-3}.

As it is shown in Figure 1, these proteomic features were present in lung cancer group, but not in in healthy control or high risk serum samples. SELDI-TOF spectra of the most discriminatory peak of the high risk group (8141 m/z) is shown in Figure 2.

When tree classification analysis was performed between lung cancer and healthy control group, a decision tree with 4 terminal nodes was created (Figure 3).

The peak of 11547 m/z was determined as the most discriminatory peak in the root node. If the intensity of this peak was higher than 0.147 these samples split into terminal node 4, which includes 99.2\% lung cancers and 0.8\% high risk cases (Figure 4).

If the intensity of peak 11547 m/z was higher than 0.147 and the intensity of peak 5832 m/z was higher than 1.221 samples split into terminal node 3, which includes 100\% samples with lung cancer. If the intensity of peak 11547 m/z was lower or equal to 0.147 and the intensity of peak 9192 m/z was higher than 13.574 samples split into terminal node 2, which includes 100\% samples with lung cancer. This model correctly identified 149 of 173 lung cancer (86\%) and 46 of 47 healthy controls. The test data yielded a sensitivity of 85.5\% (148/173), a specificity of

<p>| Table 1. Discriminatory Proteomic Features between Groups with Lung Cancer and High Risk |
|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Proteomic feature (m/z)</th>
<th>**p value</th>
<th>*AUROC</th>
<th>**Average intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>11679</td>
<td>1.88-16</td>
<td>0.86</td>
<td>1.24</td>
</tr>
<tr>
<td>11899</td>
<td>4.90-16</td>
<td>0.85</td>
<td>0.40</td>
</tr>
<tr>
<td>11547</td>
<td>2.23-15</td>
<td>0.85</td>
<td>1.03</td>
</tr>
<tr>
<td>8141</td>
<td>3.19-14</td>
<td>0.84</td>
<td>3.11</td>
</tr>
<tr>
<td>11740</td>
<td>2.46-14</td>
<td>0.84</td>
<td>1.31</td>
</tr>
<tr>
<td>11480</td>
<td>4.89-14</td>
<td>0.83</td>
<td>0.88</td>
</tr>
<tr>
<td>6438</td>
<td>2.17-12</td>
<td>0.81</td>
<td>1.49</td>
</tr>
<tr>
<td>5832</td>
<td>9.32-11</td>
<td>0.78</td>
<td>1.18</td>
</tr>
<tr>
<td>3951</td>
<td>8.37-9</td>
<td>0.76</td>
<td>1.81</td>
</tr>
<tr>
<td>6646</td>
<td>2.90-9</td>
<td>0.75</td>
<td>3.33</td>
</tr>
<tr>
<td>6115</td>
<td>8.61-7</td>
<td>0.72</td>
<td>2.05</td>
</tr>
<tr>
<td>5907</td>
<td>3.81-6</td>
<td>0.70</td>
<td>18.40</td>
</tr>
</tbody>
</table>

*The area under the ROC curve; **Mann-Whitney U test

<p>| Table 2. Discriminatory Proteomic Features between Healthy Controls and High Risk Cases |
|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Proteomic feature (m/z)</th>
<th>**p value</th>
<th>*AUROC</th>
<th>**Average intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8141</td>
<td>9.03-15</td>
<td>0.93</td>
<td>1.66</td>
</tr>
<tr>
<td>8597</td>
<td>1.82-5</td>
<td>0.72</td>
<td>1.75</td>
</tr>
<tr>
<td>6115</td>
<td>1.79-4</td>
<td>0.70</td>
<td>1.88</td>
</tr>
</tbody>
</table>

*The area under the ROC curve; **Mann-Whitney U test

![Figure 1. SELDI-TOF Mass Spectra.](image-url) Protein peaks of 11480 m/z, 11547 m/z, and 11679 m/z that are positively correlated with lung cancer.
89.4% (42/47), a positive predictive value (ppv) of 96.7% and a negative predictive value (npv) of 62.7%. The ROC analysis gave an AUROC of 0.92.

In addition to Mann–Whitney non-parametric test the comparison between patients with lung cancer and high risk was performed using tree classification to find a pattern of protein markers. With the help of Biomarker Patterns Software (BPS), a decision tree with 3 terminal nodes was created (Figure 4). The root node contains the most discriminatory protein peak of 8141 m/z, which was also determined as a statistically significant single peak in the present study. If the intensity of the peak 8141 m/z was lower or equal to 2.422 these samples split into terminal node 1, which includes 95.8% lung cancers and 4.2% high risk cases (Figure 4). If the intensity of peak m/z 8141 was higher than 2.422 and the intensity of peak m/z 6438 was lower or equal to 1.549 samples split into terminal node 2, which includes 88.2% samples with lung cancer and 11.8% samples high risk cases. Terminal node 3 contains cases with higher intensity of peak 6438 m/z, which includes 89.6% patients with high risk. This model correctly identified 168 of 173 lung cancer and 43 of 53 high risk patients. The test data yielded a sensitivity of 96% (166/173), a specificity of 77.4% (41/53), a positive predictive value (ppv) of 93.3% and a negative predictive value (npv) of 85.4% The ROC analysis gave an AUROC of 0.91.

In addition, when the sensitivity and specificity of those tumor markers were evaluated, the sensitivity was found to be very low. (Table 4). In the lung cancer group, sensitivity of CEA, Cyfra 21, NSE were 81, 56, 35 respectively and in the high risk group, sensitivity were 34, 2, 4 respectively.

Discussion

The methods used in the early diagnosis and screening of lung cancer have resulted in unsatisfactory outcomes (Yang et al., 2005). Even the efficacy of low-dose computed tomography, considered as a screening test among high-risk groups, remains to be contentious (Bellomi et al., 2006; Hicks et al., 2007). Therefore, there is a need for novel methods in the early and accurate diagnosis of lung cancer.

Proteomic studies using SELDI-TOF-MS analysis for lung cancer appear to be inadequate. Xiao et al. (2003) used two SELDI-TOF-MS chips (IAMC3, WCX29) and found 5 protein peaks that can be used for the differentiation of lung cancer cases and healthy individuals: 4353, 4466, 15120, 15880, and 15962 (each was determined by both of the chips). The sensitivity of this method was 44.8% and 91.3%, whereas the specificity was 85% and 94.4% (Xiao et al., 2003). Yang et al. employed WCX 2 chip and reported five protein peaks that differentiate lung cancer cases from healthy individuals: 11493, 6429, 8245, 5335, 2538. Its sensitivity and specificity were 86.9% and 80%, respectively (Yang et al., 2005).

Jacot et al. (2008) evaluated serum protein profiles by fingerprint method in lung cancer cases and patients with benign lung cancer, and determined 88 different protein peaks. Among those, 4628 Da protein was found to have
a statistical significance in the prognosis of advanced stage lung cancer.

In our study, we determined three different protein peaks (11480 m/z, 11547 m/z, and 11679 m/z) that were present in the lung cancer group, while being absent in the high-risk and control groups. These peaks may prove to be markers of lung cancer which suggests that they may be used in the early diagnosis of lung cancer. Furthermore, 8141 m/z peak was found in the high-risk group, but not in the lung cancer and control groups.

Lin et al. (2010) conducted a study including 35 lung cancer, 46 benign lung diseases, and 44 healthy cases in 2010 wherein 4053.88, 4209.57 and 3883.33 peaks were determined to be distinctive for the differentiation of lung cancer and benign groups, while 2951.83, 4209.73 peaks were found to be distinctive for differentiation of the lung cancer and control groups (Lin et al., 2010). Yildiz et al. (2007) determined 7 proteins by MALDI-TOF-MS analysis in lung cancer cases in 2007 and described this profile, although it was not a biomarker, as a factor that could differentiate lung cancer from other diseases (Yildiz et al., 2007).

Xiao et al. (2003). used two separate chips and SELDI-TOF analysis, and found their sensitivity as 44.8% and 91.3%, and specificity as 85% and 94.4%. Yang et al. (2005). employed WCX 2 chip and reported 5 protein peaks differentiating lung cancer cases from healthy controls; the sensitivity and specificity were 86.9% and 80%, respectively (Xiao et al., 2003). In our study, the sensitivity, specificity, positive predictive value (ppv), and negative predictive value (npv) for differentiating lung cancer from healthy controls were 85.5%, 89.4%, 96.7%, and 62.7%, respectively. For identification of lung cancer and high-risk groups, the sensitivity, specificity, positive predictive value (ppv), and negative predictive value (npv) were 96%, 77.4%, 93.3%, 85.4%.

In 2008, Han et al. (2008). performed a study on a population comprised of 151 lung cancer and 102 healthy individuals, and found the level of 5808 and 5971 peaks 6 times higher in the lung cancer group. The sensitivity and specificity for those peaks were 89% and 91% (Han et al., 2008). We believe that detection of different proteins in our study and the above noted trials may be stemming from selection of different Da range, use of different chips, and varying sample sizes.

Since the tumor markers appear to have low sensitivities despite showing high specificities, they may not be beneficial. In previous studies, tumor markers have shown low sensitivity and specificity, as well (Schneider et al., 2000; Kulpa et al., 2002; Hang et al., 2011).

In conclusion, we believe that protein peaks detected by SELDI-TOF analyses may be helpful in differentiating healthy individuals from lung cancer cases in a non-invasive method, and also SELDI-TOF-MS analysis may be used for early diagnosis of lung cancer as well as its screening in the future. Further studies with larger sample sizes are required.

References

Cebrail Şimşek et al


