Proteomic analysis for detecting serum biomarkers related to smoking in humans

XIAO Dan, ZHAO Li-juan, CHU Shui-lian, JING Hang and WANG Chen

Keywords: smoking; serum; biological markers; proteomics

Background  Smoking is the leading cause of death in the world. This study focused on the difference of the serum proteomic profiling between healthy smokers and nonsmokers in order to find smoking-specific serum biomarkers.

Methods  Pattern-based proteomic profiling of 100 serum samples (from 50 Chinese male smokers and 50 matched nonsmokers) was performed through magnetic bead fractionation coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis (MALDI-TOF-MS) and resulting data were statistically analyzed by Ciphergen ProteinChip software 3.0.2.

Results  We found 72 serum peaks were significantly different between smokers and nonsmokers ($P <0.05$). Marker peaks of mass-to-charge ratio ($m/z$) 3159.13, 7561.03 and 9407.32 were smoking-specific.

Conclusion  The preliminary data suggested that smoking-specific serum biomarkers could be detected in humans.

METHODS

Subjects  This study had been approved by the Ethics Committee of Beijing Chao-Yang Hospital, Capital Medical University, and written informed consents were obtained from all subjects. Inclusion criteria were Chinese male smokers or nonsmokers, aged from 18 to 75 years. Exclusion criteria included: diseases that included mental illness, drug or...
alcohol addiction, any types of cancer, chronic obstructive pulmonary disease, asthma, other lung diseases, diabetes, uncontrolled hypertension, stroke, coronary heart disease (angina or heart attack), or other heart diseases.

A total of 100 subjects were recruited from January 2007 to January 2009 at Beijing Chao-Yang Hospital. Smokers were from smoking cessation clinics, and nonsmokers were from routine medical clinics. The subjects were assigned to two groups based on their smoking status: (1) nonsmokers: those who reported never smoking in their lifetime (n=50), verified by a carbon monoxide test (≤10 ppm); and (2) smokers: those who had smoked at least 100 cigarettes and continued to smoke at the time of the study (n=50), also verified by a carbon monoxide test (>10 ppm).

All enrolled subjects had completed a standardized baseline questionnaire, including age, smoking history, and medical history for exclusion criteria. Characteristics of the subjects are listed in Table 1. Furthermore, all subjects were asked to take the carbon monoxide breath testing immediately to verify their smoking status and 4 ml of venous blood was taken for serum proteomic analysis.

**Materials**

Serum was prepared by drawing fasting blood into a 4 ml BD Vacutainer (Becton Dickinson Vacutainer Systems, USA) without anti-coagulant and was left to clot at room temperature for up to one hour. Serum was isolated by centrifugation at 2000 × g for five minutes at room temperature. Finally, serum was frozen and stored at −80°C.

The matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis (MALDI-TOF-MS) (PBS IIC) and WCX2 chips were provided by Ciphergen Biosystems (USA). The WCX magnetic beads were provided by Beijing SED Science and Technology (China). Acetonitrile (ACN), trifluoroacetic acid (TFA), urea, N’-2-hydroxyethylpiperazine-N0-2-ethanesulfonic acid (Hepes), water (HPLC grade), Tris-HCl (pH 9.0), DL-dithiothreitol (DTT), sinapic acid (SPA), WCX2 binding buffer (50 mmol/L sodium acetate, pH 4.0), 3-((3-Cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS) were provided by Sigma-Aldrich (USA).

**MALDI-TOF-MS**

WCX magnetic beads were used for serum preparation. First, WCX magnetic beads needed to be activated. Briefly, 50 μl WCX magnetic beads (50 mg/ml) was transferred to a PCR tube that was placed in a 2×8 well magnetic bead separator (MBS device) for one minute for magnetic fixation of the WCX particles. The supernatant was aspirated and the tubes were removed from the MBS device and 100 μl binding buffer (50 mmol/L sodium acetate, pH 4.0) was added and carefully mixed with the magnetic beads. The tube was placed in the MBS device and moved back and forth sequentially between adjacent wells on each side of the magnetic bar in the MBS device, which facilitated washing of the magnetic particles. After fixation of the magnetic beads for one minute in the MBS device, the supernatant was aspirated. This washing procedure was repeated twice.

Before reacting with magnetic beads, serum samples needed to be diluted. First, 10 μl serum and 20 μl U9 solution (9 mmol/L urea, 20 g/L CHAPS, 10 g/L DTT, 50 mmol/L Tris-HCl, pH 9.0) were transferred to a 0.5 ml centrifuge tube. The serum and solution were mixed and then incubated for 30 minutes at 4°C. After that 370 μl binding buffer was added to make 40-fold dilution of the serum. Then, 100 μl diluted serum was added to the activated magnetic beads and mixed carefully, and incubated for one hour at 4°C before washing twice with 100 μl binding buffer. After the final washing step, bound molecules were eluted by incubation with 10 μl 0.5% (v/v) TFA. The elute (5 μl) was diluted 2-fold by adding 5 μl SPA (50% CAN + 0.5% TFA). The diluted elute (1 μl) was aspirated and spotted onto an 8-spot pre-structured sample chip (Au-chip, Ciphergen Biosystems) and left to dry at room temperature. Finally, protein crystals on the chip were detected by MALDI-TOF-MS.

Prepared Au-chips were placed on the Protein Biological System IIC mass spectrometer reader (PBS IIC, Ciphergen Biosystems), and time-of-flight spectra were generated by averaging 80 laser shots collected on each spot at laser intensity 205 and detector sensitivity 8. Mass

<table>
<thead>
<tr>
<th>Variables</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>100</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>18–30</td>
<td>30</td>
</tr>
<tr>
<td>31–40</td>
<td>32</td>
</tr>
<tr>
<td>41–50</td>
<td>21</td>
</tr>
<tr>
<td>51–60</td>
<td>13</td>
</tr>
<tr>
<td>61–70</td>
<td>4</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
</tr>
<tr>
<td>Smokers</td>
<td>50</td>
</tr>
<tr>
<td>Nonsmokers</td>
<td>50</td>
</tr>
<tr>
<td>Cigarettes consumption (per day)</td>
<td></td>
</tr>
<tr>
<td>≤10</td>
<td>9</td>
</tr>
<tr>
<td>11–20</td>
<td>34</td>
</tr>
<tr>
<td>21–30</td>
<td>3</td>
</tr>
<tr>
<td>≥31</td>
<td>4</td>
</tr>
<tr>
<td>Smoking index (cigarettes consumption per day × smoking years)</td>
<td></td>
</tr>
<tr>
<td>&lt;400</td>
<td>26</td>
</tr>
<tr>
<td>≥400</td>
<td>24</td>
</tr>
<tr>
<td>FTND score</td>
<td></td>
</tr>
<tr>
<td>0–3</td>
<td>5</td>
</tr>
<tr>
<td>4–6</td>
<td>25</td>
</tr>
<tr>
<td>7–10</td>
<td>20</td>
</tr>
</tbody>
</table>
accuracy was calibrated externally by standard procedures using the all-in-one peptide molecular mass standards (Ciphergen Biosystems).

Data analysis
The data analysis involved three stages: (1) peak detection and alignment, (2) selection of differently expressed peaks that may represent potential biomarkers among the different groups, and (3) data analysis using a decision tree algorithm.

Peak detection was performed by Ciphergen ProteinChip software 3.0.2 (Ciphergen Biosystems). The protein or peptide peaks with mass-to-charge ratio (m/z) ranging from 2500 to 15 000 were selected for analysis. Peak detection involved (1) baseline subtraction, (2) mass accuracy calibration, and (3) automatic peak detection. Using Biomarker Wizard Version 3.1.0 (Ciphergen Biosystems), biomarkers were generated that represent consistent protein or peptide peak sets across multiple spectra. Baseline subtraction was performed on all spectra. The settings for auto-detect peaks to cluster were as follows: signal-to-noise ratio was five and minimum peak threshold was 10% for the first pass.

Differently expressed peaks were selected by the differences in protein or peptide peak intensities between groups using the Biomarker Wizard Version 3.1.0, which used the non-parametric Kruskal-Wallis test and Mann-Whitney test. Construction of the decision tree classification algorithm was performed by Biomarker Patterns Software 5.0 (BPS, Ciphergen Biosystems). BPS is an implementation of the Classification and Regression Trees (CART) decision tree system. BPS uses the peak information generated by the training set of known samples to build a binary decision tree algorithm.

The spectra of 100 serum samples were separated by a stratified random sampling into “training set” data and “blinded testing set” data. Our aim was to develop a boosting decision classification using the training set data, for which both the clinical diagnosis and the proteomic spectrum profile data pairs of each sample were available. The resulting boosting decision tree classification was then applied to the blinded testing set data for estimating the accuracy of the decision tree model. Ciphergen ProteinChip Software 3.0.2, Biomarker Wizard software 3.1.0 and BPS 5.0 (Ciphergen Biosystems) were the available supporting software provided by Ciphergen Company for data analysis. All the results were expressed as mean ± standard deviation (SD), and a P <0.05 was considered statistically significant.

RESULTS

Identification of serum proteomic profiles
Protemic spectra of 100 serum samples were generated by MALDI-TOF-MS combined with WCX magnetic beads. The combination was particularly effective in resolving low molecular weight proteins and peptides (Figure 1).

Comparison of serum proteomic spectra between smokers and nonsmokers
A total of 110 peaks were detected that were different between smokers and nonsmokers in the training set, among which 72 peaks were significantly different (P <0.05).

Construction of boosting decision tree of smoking
Of the classification trees analyzed by BPS software, the most optimal classification tree with the lowest error cost was eventually established, that was, marker peaks of m/z 3159.13, 7561.03 and 9407.32 were selected to build the classification tree. The protein or peptide peaks of m/z 3159.13 and 7561.03 were down-regulated and m/z 9407.32 was up-regulated in smokers compared with nonsmokers (Figure 2). All training set serum samples were differentiated into four terminal nodes (Figure 3). Samples differentiated into terminal nodes 1 and 4 were assigned to nonsmokers, terminal nodes 2 and 3 to smokers. For example, if an unknown sample had a low intensity of peak m/z 3159.13 (intensity ≤9.83) and low intensity of m/z 9407.32 (intensity ≤3.16), it will be placed in terminal node 1. The corresponding receiver operating characteristics curve (ROC) of the optimal decision tree was 0.984 supplied by the BPS 5.0 (Figure 4).

The decision tree of smoking could differentiate samples of smokers from nonsmokers. In the training set, the decision tree model could accurately recognize 38 of 40 smokers and 39 of 40 nonsmokers. Validation on the blinded testing set indicated that the decision tree could differentiate 8 of 10 smokers, sensitivity was 80%, and all nonsmokers, specificity was 100% (Table 2).

DISCUSSION
This study has provided a picture of the serum proteome related to smoking in Chinese males. There were 72 serum peaks found to be different between smokers and
nonsmokers ($P < 0.05$). Among these, marker peaks of $m/z$ 3159.13, 7561.03 and 9407.32 were smoking-specific and could be used to construct the diagnostic pattern of smoking. Alterations related to smoking have been repeatedly reported in peripheral tissues.3-9 Many studies also provided evidence that smoking-specific proteins or peptides could be detected by proteomics analysis in different samples, such as sputum,13 bronchoalveolar lavage,14 exhaled breath,15 exhaled breath condensate,16 and urine.17 However, these studies only assessed the alterations related to smoking in a microenvironment rather than systemic environment.
A serum proteomic analysis was performed by 2D gel electrophoresis (2-DE) on rats during nicotine self-administration, extinction and relapse. It was found that C reactive protein and haemopexin displayed a significant reduction after nicotine administration, which suggested that nicotine intake could change the serum protein profile. Tewari et al utilized 2-D DIGE/MS to characterize the proteomic changes of cigarette smoke exposed and control mice. The proteins identified were involved in vascular function, coagulation, metabolism and immune function, which could potentially serve as markers predicting the onset and progression of cardiovascular and pulmonary diseases.

Della et al analyzed the platelet proteome of healthy smokers and nonsmokers by 2-DE and compared by Decyder software and identified by mass spectrometry analysis (nano-LC-MS/MS). It was found that three proteins (Factor XIII-A subunit, platelet glycoprotein IIb and beta-actin) were significantly up-regulated, whereas WDR1 protein and chaperone HSP60 were down-regulated in smokers, and these proteins were related to inflammation and apoptosis.

Compared to 2-DE, MALDI-TOF-MS is a novel and useful tool for proteomic analysis. It can achieve high-throughput protein analysis of biological samples with high sensitivity, reproducibility, and resolving capability. Protein profiling by MALDI-TOF-MS after proteome fractionation with magnetic beads is a robust, precise and rapid technique for the investigation of complex blood samples and is appropriate for preliminary biomarker discovery, especially for the detection of low concentrations of proteins and peptides in serum. Protein profiling has been applied in the study of various diseases.

Musharraf et al used MALDI-TOF-MS to focus on the correlation between the proteomic profiling of patients with lung cancer, healthy nonsmokers and smokers. It was found that marker peaks at m/z 1760, 5773, 5851, 2940, and 7172 could serve with high sensitivity and specificity as a differentiation tool for lung cancer patients. However, this study did not analyze the proteomic difference between smokers and nonsmokers.

Our study focused on the difference in serum proteomic profiling between healthy smokers and nonsmokers. Existence of 72 differential marker peaks between smokers and nonsmokers indicated that broad pathological changes in the serum proteome related to smoking might be involved. Instead of a single protein or peptide type, a group of proteins or peptides might be related to smoking. Based on the differential peaks, potential smoking-specific serum biomarkers were found. The initial decision tree classification for distinguishing smokers from nonsmokers was generated by analyzing “training set” data by a machine learning algorithm derived from CART. Three potential biomarkers were involved in the decision tree. The decision tree could classify serum samples into two groups: smokers and nonsmokers. The validation with “blinded testing set” data indicated that the decision tree could correctly recognize 8/10 smokers and all nonsmokers. The high sensitivity and specificity of the results is encouraging.

Although the 2010 Surgeon General’s Report, How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease, focused on the pathological mechanism of smoking and nicotine dependence, there were still many issues that had not been completely revealed. The results from our study suggested that smoking-specific serum biomarkers could be detected in humans, which indicated that it would be a chance for us to detect serum biomarkers specific to nicotine dependence in humans. In the future, determining the properties of these smoking-specific serum biomarkers and detecting the serum biomarkers specific to nicotine dependence will provide more evidence of the pathological mechanism of smoking and nicotine dependence.

REFERENCES


