Lung Cancer 99 (2016) 46–52

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Serum mass profile signature as a biomarker of early lung cancer

Piotr Widlak[a,*], Monika Pietrowska[a], Joanna Polanska[b], Michal Marczyk[b], Malgorzata Ros-Mazurczyk[a], Rafał Dziadziuszko[c], Jacek Jassem[c], Witold Rzyman[c]

[a] Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, ul. Wybrzeże Armii Krajowej 15, 44-100 Gliwice, Poland
[b] Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland
[c] Medical University of Gdańsk, ul. Dębinki 7, 80-211 Gdańsk, Poland

ARTICLE INFO

Article history:
Received 19 February 2016
Received in revised form 12 May 2016
Accepted 11 June 2016

Keywords:
Early detection
Lung cancer screening
Mass spectrometry
Proteomics
Serum biomarkers

ABSTRACT

Objectives: Circulating molecular biomarkers of lung cancer may allow the pre-selection of candidates for computed tomography screening or increase its efficacy. We aimed to identify features of serum mass profile distinguishing individuals with early lung cancer from healthy participants of the lung cancer screening program.
Methods: Blood samples were collected during a low-dose computed tomography (LD-CT) screening program performed by one institution (Medical University of Gdansk, Poland). MALDI-ToF mass spectrometry was used to characterize the low-molecular-weight (1000–14,000 Da) serum fraction. The analysis comprised 95 patients with early stage lung cancer (including 30 screen-detected cases) and a matched group of 285 healthy controls. The cases were split into two independent cohorts (discovery and validation), analyzed separately 6 months apart.
Results: Several molecular components of serum (putatively components of endogenous peptidome) discriminating patients with early lung cancer from controls were identified in a discovery cohort. This allowed building an effective cancer classifier as a model tuned to maximize negative predictive value, with an area under the curve (AUC) of 0.88, a negative predictive value of 100%, and a positive predictive value of 48%. However, the classifier performed worse in a validation cohort including independent sample sets (AUC 0.73, NPV 88% and PPV 30%).
Conclusions: We developed a serum mass profile-based signature identifying patients with early lung cancer. Although this marker has insufficient value as a stand-alone preselecting tool for LD-CT screening, its potential clinical usefulness in evaluation of indeterminate pulmonary nodules deserves further investigation.

1. Introduction

Lung cancer is a major health problem and the leading cause of cancer mortality, responsible for about 18% of cancer-related death worldwide. The majority of lung cancer cases are diagnosed at advanced stages, resulting in an average 5-year survival of merely 10–15%. However, the corresponding survival probability for early-detected disease is above 70%. Hence, in line with tobacco-control efforts (primary prevention), lung cancer screening for the detection of early stages of the disease may become the major strategy to reduce lung cancer mortality [1–3]. Several diagnostic tools allowing early lung cancer detection have been investigated within the past decades, but none have found their routine clinical application. The results of the recent National Lung Screening Trial (NLST) showed a 20% reduction in lung cancer-specific mortality with low-dose computed tomography (LD-CT) screening in a high-risk group, compared with a conventional chest X-ray examination [4]. Thus, lung cancer screening based on LD-CT is a new promising strategy for lung cancer detection, with the potential of reducing its mortality. However, in the NLST the overall positive predictive value (PPV) of the test was only 3.8% [4]. In consequence, the vast majority of patients with screen-detected pulmonary nodules are subjected to further diagnostic procedures including positron emission tomography, transthoracic needle aspiration, bronchoscopic biopsy and/or surgery. In our own experience, 75% of patients with screen-detected lung abnormalities underwent unnecessary further diagnostic work-up [5]. Hence, there is a constant need for

* Corresponding author.
E-mail addresses: widlak@io.gliwice.pl (P. Widlak), m_pietrowska@io.gliwice.pl (M. Pietrowska), joanna.polanska@polsl.pl (J. Polanska), michal.marczyk@polsl.pl (M. Marczyk), malgorzata.ros@io.gliwice.pl (M. Ros-Mazurczyk), rafald@gumed.edu.pl (R. Dziadziuszko), jjassem@gumed.edu.pl (J. Jassem), wrzyman@gumed.edu.pl (W. Rzyman).

http://dx.doi.org/10.1016/j.lungcan.2016.06.011
0169-5002/

clinical tests supporting CT-based screening for the detection of lung cancer. These tests are expected to pre-select individuals for LD-CT examination or better discrimination between benign and malignant nodules detected by LD-CT [6,7].

Blood is among the most available sources of biomarkers potentially enhancing the power of early lung cancer detection. Several blood components have been analyzed in the search for such biomarkers, including circulating tumor cells, circulating tumor DNA, micro RNA, autoantibodies and specific serum or plasma proteins [8–10]. Of those, particularly promising are multicomponent panels of serum proteins, which were found to discriminate between healthy controls and patients with lung cancer in early and advanced stages with 80–90% sensitivity and specificity [11–15]. A promising potential of multicomponent signatures based on features of blood proteome in the detection of early stage lung cancer was demonstrated in several studies [16–21].

The low-molecular-weight fraction of the blood proteome, the so-called endogenous peptidome, is a potentially valuable source of novel cancer biomarkers [22]. This proteome fraction can be analyzed using a pattern analysis that considers a specific profile of the detected components, exemplified by ions registered at defined m/z values in a mass spectrum (MS) of the low-molecular-weight fraction of serum or plasma [23,24]. Numerous studies have explored the applicability of MS-based profiling for cancer evaluation, resulting in the development of several multi-peptide cancer-specific signatures [22,25–28]. Mass profiling of serum or plasma was also used for the identification of multi-peptide signatures discriminating healthy subjects from patients with different types of lung cancer [29–35]. However, none of the MS-based cancer signatures was successfully tested and validated in the context of lung cancer screening program, and none has found its clinical application in this setting. Nevertheless, such signatures were also shown to provide added value to clinical and imaging assessment of undetermined lung abnormalities [32,36].

Matrix-Assisted Laser Desorption Ionization (MALDI) mass spectrometry coupled to a Time-of-Flight (ToF) analyzer (and its derivative SELDI-ToF) is considered particularly suitable for mass profiling, and this approach was used in the above-mentioned studies. However, MALDI-ToF has several drawbacks limiting its clinical applicability, such as lack of direct actual identification of detected compounds and the requirement of very rigid standardization of analytical procedures to warrant reproducibility of results [28]. Hence, other MS tools have been employed in studies aimed at discovery of specific lung cancer biomarkers [19,20]. On the other hand, MALDI profiling is faster, easier, less laborious and less expensive than other MS-based proteomic tools, making it a preferential method for population screening programs involving thousands of participants. Of note, a certified MALDI Biotyper system (Bruker Daltonics), which uses MALDI-profiling for identification of microorganisms, is among very few MS-based diagnostic tests with direct clinical relevance. Here, we investigated the value of MALDI-based profiling of serum components as a tool for the early detection of lung cancer and enhancement of LD-CT lung cancer screening.

2. Materials and methods

2.1. Study subjects

Material for this study was collected during the Pomeranian Lung Cancer Screening Program performed by Medical University of Gdańsk between 2008 and 2010, which offered LD-CT examination for current or former smokers aged from 50 to 75 years with at least a 20 pack-year history. Out of above 8600 participants of this program, blood samples were collected from 3600 volunteers, 30 of whom were finally diagnosed with lung cancer. Additionally, blood samples were collected (by Medical University of Gdańsk) before any therapeutic intervention from 65 asymptomatic, incidentally detected early-stage lung cancer (IA-IIB) patients not participating in the screening program, with similar characteristics regarding smoking habits and age. Each cancer case was accompanied by three controls matched according to sex, age and smoking status. All controls were selected from the participants of the screening program for whom no malignancy was detected. A detailed characteristic of all groups of donors is presented in Table 1. All samples were processed according to unified, strictly designed protocol. The institutional Ethics Committee approved the study and all participants provided informed consent indicating their voluntary participation in the project.

2.2. Serum sample preparation

Peripheral blood was collected into a 5 mL BD Vacutainer Tube, incubated for 30 min at room temperature to allow clotting, and then centrifuged at 1000*g* for 10 min to remove the clot. The serum was aliquoted and stored at −70 °C.

2.3. Profiling of the low-molecular-weight fraction of serum

Serum samples were diluted 1:5 with a mixture of acetonitrile (20% ACN) and ammonium bicarbonate (250 mM) directly before analysis, and then albumin and other large-molecular-weight proteins were removed by centrifugation (30 min, 3000*g*) through Microcon spin filters with 50 kDa cut-off membrane (Millipore, Billerica, MA, USA). Each sample was loaded onto a C18 ZipTip microcolumn (Millipore) by passing it through repeatedly to achieve saturation, then the column was washed with water and eluted with 1 μL of matrix solution (saturated solution of alpha-cyjano-4-hydroxy-cinnamic acid in 50% ACN/H_2O and 0.1% trifluoroacetic acid) directly onto 800 μm AnchorChip plate (Bruker Daltonik, Bremen, Germany). Samples were analyzed using an ultrafleXtreme MALDI-ToF mass spectrometer (Bruker Daltonik); the analyzer worked in the linear mode and positive ions were recorded in the mass range between 1000 and 14,000 Da. Mass calibration was performed after every four samples using standards in the mass range of 2.8–16.9 kDa. ZipTip extraction/loading was repeated twice for each sample and two spectra were acquired for each spot (i.e. four spectra were recorded for each sample). All samples were analyzed in a random sequence to avoid a possible batch effect.

2.4. Spectral data processing

Low quality spectra were detected with the use of modeling of robust signal-to-noise ratio measure and total ion current (TIC) distribution analysis [37], and excluded from further analysis (as a result of that 2 control samples were removed). Pre-processing of spectra included global linear alignment of all spectra to the major peak located at m/z 1465.65 Da (corresponding to fibrinopeptide A in des-Ala form, FIBA fragment 21–35), baseline identification and removal, alignment and averaging of technical repeats (with outlier detection done using the Dixon procedure), and binning the neighboring points to reduce data complexity. Signal intensities were normalized using the Total Ion Current method, and then distributions of spectra intensities were normalized using the Invariant Set method [38], which chooses a subset of m/z values with small within-subset rank difference in spectra that serve as the basis for fitting a normalization curve (the fitted curve is the running median curve in the scatterplot of intensities of the baseline spectrum and the normalized spectrum). Spectrum with median overall intensity in Dataset A (see below) was chosen as the base-