Components of serum peptidome can differentiate between healthy controls and patients with early stage lung cancer

**Piotr Widłak¹, Monika Pietrowska¹, Joanna Polańska², Michał Marczyk², Rafał Dziadziuszko³, Witold Rzyman³**

¹ Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, 44-101 Gliwice, Poland; ² Silesian University of Technology, 44-100 Gliwice, Poland; ³ Medical University of Gdańsk, 80-210 Gdańsk, Poland

The project was supported by NCBiR Grant POIG.01.01.02-00-080/09 (MOLTEST_2013)

**OBJECTIVE**

**Background:** Screening with low-dose computed tomography of high-risk group for lung cancer development allows for early detection of malignancy in a minor proportion of subjects and leads to improved outcomes. Implementation of complementary minimally-invasive molecular markers for more efficient pre-selection of candidates for imaging tests or help to further define detected changes is a rational way to further improve efficacy of such screening.

**Aim:** To identify features of serum peptidome that could be used for differentiation of individuals with early lung cancer from other lung cancer screening program.

**Keywords:** biomarkers; early lung cancer; mass spectrometry; screening test; serum proteomics

**MATERIALS AND METHODS**

**Material:** Blood samples were collected during lung cancer screening program performed in Pomerania district (Poland); initial material was collected from about 6500 participants of the program subjected to LD-CT. The analysis was performed in a group of 100 lung cancer patients (with early stage lung cancer diagnosed without clinical symptoms during the screening program or through routine diagnostic procedures) and a matched group of 300 controls (participants of the screening without malignancy).

**Proteomic analysis:** MALDI-ToF mass spectrometry was used to characterize the low-molecular-weight fraction of serum proteome (so called endogenous serum peptidome). Spectra were registered in the 800-14,000 Da range (Fig. 1). Spectral components, which reflected [M+H]+ peptide ions recorded at defined m/z values, were detected using decomposition of mass spectra into their Gaussian components. The initial set of Gaussian components was further processed to merge overlapping components and to remove the residual baseline; this resulted in dimension reduction to 238 Gaussian components. The final 238 Gaussian components were used to compute features of registered spectra for all samples by the operations of convolutions with Gaussian masks.

**RESULTS**

**Building and training of cancer classifier**

Group of 50 donors with early lung cancer and matched group of 150 healthy controls (participants of the screening program) were used as the training group. Differentiating spectral components (serum peptide ions) were detected using parametric T-test (tuned by test of homogeneity of variances). Cancer classifier was constructed using logistic regression model (the likelihood ratio test was applied for model selection). Logistic classifier threshold was set to maximize NPV (Negative Predictive Value). Resulting classifier built of 5 components was characterized by: 94.1% sensitivity, 39.6% specificity, 95.2% NPV and 34.3% PPV.

**Validation of biomarker signature**

Independent group of 50 donors with early lung cancer and matched group of 150 healthy controls (participants of the screening program) were used as the validation group. Mass spectra were registered for this group separately 6 months after the training group. Pre-processing of spectra in the validation data set was performed independently of the training data set to avoid information leakage (spectra were only normalized to the reference TIC from the training data set). The optimal classifier built on the training data set was applied to predict the outcome in the validation data set. The obtained classifier performance indices were as follows: 76.1% sensitivity, 24.9% specificity, 77.6% NPV and 23.43% PPV (values comparable to widely used molecular tests for cancer screening, e.g. PSA).

**SUMMARY AND CONCLUSIONS**

MALDI-based profiling of serum peptidome allowed identification of components differentiating patients with early stage lung cancer from healthy individuals. Hence, biomarker based on serum peptide signature might have a potential applicability to support CT-based screening tests. Results of this hypothetical molecular test could be used for pre-selection of candidates for CT-screening: negative result would preclude cancer while positive result could be verified during CT scan.