Comparison of peptide cancer signatures identified by mass spectrometry in serum of patients with head and neck, lung and colorectal cancers: Association with tumor progression

MONIKA PIETROWSKA[1], JOANNA POLAŃSKA[2], RAFAŁ SUWIŃSKI[1], MACIEJ WIDEŁ[1], TOMASZ RUTKOWSKI[1], MICHAŁ MARCZYK[2], IWONA DOMIŃCZYK[1], LUCYNA PONGE[1], ŁUKASZ MARCZAK[3], ANDRZEJ POLAŃSKI[2,4] and PIOTR WIDŁAK[1]

[1]Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology; [2]Silesian University of Technology, Gliwice; [3]Polish Academy of Science, Institute of Bioorganic Chemistry, Poznan; [4]Polish-Japanese Institute of Information Technology, Bytom, Poland

Received July 4, 2011; Accepted August 17, 2011

DOI: 10.3892/ijo.2011.1186

Abstract. Mass spectrometry-based analyses of the low-molecular-weight fraction of serum proteome allow identifying proteome profiles (signatures) that are potentially useful in detection and diagnostics of cancer. Here we compared serum proteome profiles of healthy donors and patients with three different types of cancer aiming to identify peptide signatures that were either common for all cancer samples or specific for cancer type. Blood samples were collected before start of the therapy from patients with head and neck squamous cell cancer, colorectal adenocarcinoma and non-small cell lung cancer, and from a corresponding group of healthy volunteers. Mass profiles of the serum proteome were recorded in the range between 2 and 13 kDa using MALDI-ToF spectrometry and 131 identified peptide ions were used for statistical analyses. Similar degrees of overall similarities were observed in all intra-group and inter-group analyses when general features of serum proteome profiles were compared between individual samples. However, classifiers built of selected spectral components allowed differentiation between healthy donors and three groups of cancer patients with 69-74% sensitivity and 82-84% specificity. There were two common peptide species (3766 and 5867 Da) with increased levels in all cancer samples. Several spectral components permitted differentiation between lung cancer samples and either head and neck cancer or colorectal cancer samples, but two latter types of samples could not be properly discriminated. Abundance of spectral components that putatively corresponded to fragments of serum amyloid A (11511 and 11667 Da) was highest in lung cancer samples, yet increased levels of these peptides appeared to generally associate with more advanced cancer cases. We concluded that certain components of serum peptide signatures are common for different cancer signatures and putatively reflect general response of organism to the disease, yet other components of such signatures are more specific and most likely correspond to clinical stage of the malignancy.

Correspondence to: Dr Piotr Widłak, Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice, Poland
E-mail: widlak@io.gliwice.pl

Key words: cancer signatures, mass spectrometry, molecular classifiers, pattern analysis, serum proteomics

Introduction

The low-molecular-weight fraction of the blood proteome appears to be a promising source of novel biomarkers of human diseases. Thus, mass spectrometry (MS) methods, which allow characterization of this particular component of human proteome, emerge as a valuable tool of clinical proteomics and disease diagnostics (1-5). The proteomics approach that takes into consideration characteristic features of the whole proteome but does not rely on particular protein, is called proteome pattern analysis or proteome profiling. In this approach specific proteome signatures are built based on several peptide/protein components, which can be exemplified by ions registered at defined m/z values in the mass spectrum. Such signatures can be used for sample identification and classification even though their particular components may lack differentiating power when analyzed separately (6-9). Matrix-assisted laser desorption/ionization spectrometry (MALDI), and its derivative surface-enhanced laser desorption/ionization spectrometry (SELDI), coupled to a time-of-flight (ToF) type of analyzers, appear to be particularly suitable for such proteome pattern analysis. Several works have been published that explored the applicability of MALDI/SELDI-based analysis of the low-molecular-weight fraction of serum/plasma proteome for cancer diagnostics since the milestone report published by Petricoin and coworkers in 2002 (10). These studies have shown that multi-peptide signatures selected in numerical tests have potential values for diagnostics of different types of cancer (5,11-15). Because of apparent problems with standardization of methodological details, both experimental and computational, none of proposed serum peptide signatures analyzed directly by mass spectrometry has been approved for routine diagnostics. However, this approach is generally accepted for identification

of marker candidates, the first stage of a biomarker's discovery pipeline.

Several components of proposed cancer signatures, especially those characteristic for advanced cancer, were identified as fragments of blood proteins involved in the acute phase and inflammatory response (16,17). This indicated that among cancer biomarker candidates to be found by the MS-based serum proteome profiling were those reflecting overall influence of a disease upon the human organism. However, the contribution in proposed cancer signatures of components specific for particular types of malignancies and 'non-specific' components related to general response of the organism has not been established yet, hence comparison of serum proteome signatures specific for different types of cancer would help to clarify this issue. Several peptides reported to discriminate serum (or plasma) samples collected from healthy persons and patients with different cancers appeared repetitively in published studies, but only a few works addressed systematically the question of specificity of identified peptide signatures for different types of cancer. Comparative analysis of serum peptides detected in samples from healthy persons and breast, bladder or prostate cancer patients allowed identification of cancer-specific signatures, yet some differentiating peptides were common for all three cancer signatures (e.g. fragments of fibrinopeptide A) (18). A few serum peptides that discriminated colon cancer patients from healthy donors were also characteristic for patients with lung or prostate cancers but not for patients with breast cancer (e.g., fragments of apolipoprotein C1) (19). Here we aimed to characterize similarities of proteome profiles registered by MALDI-ToF spectrometry in serum of healthy donors and patients with three different types of malignancies - head and neck squamous cell cancer, colorectal adenocarcinoma and non-small cell lung cancer, and verify whether cancer type-specific signatures could be built based on such proteome profiles.

Materials and methods

Characteristics of patient and control groups. One hundred and twenty male cancer patients was enrolled into this study: 35 patients with squamous cell cancer located in head and neck region (40% in larynx, 29% in pharynx and 31% in oral cavity; called collectively head and neck cancers), 35 patient with colorectal cancer (adenocarcinoma type) and 50 patients with non-small cell lung cancer (hereafter called lung cancer). For comparison, the head and neck cancer group consisted of patients with the least advanced cancers while the lung cancer group consisted of patients with the most advanced cancers, on average; Table I shows more detailed information on the analyzed groups. Forty-five healthy male volunteers were included in the study as a control group. All participants were Caucasians; there was a similar proportion of smokers (~73%) and alcohol consumers (~71%) in all groups (to further reduce heterogeneity of the analyzed group only male donors were recruited). The study was approved by the local Ethics Committee at the Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, and all participants provided informed consent indicating their conscious and voluntary participation.

Preparation of serum samples. Samples were collected before the start of a therapy of patients and processed following a standardized protocol. Blood was collected into a 5 ml Vacutainer Tube (Becton-Dickinson), 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. Directly before analysis, samples were diluted 1:5 with 20% acetonitrile (ACN) and 25 mM ammonium bicarbonate.

Registration of mass spectra. Samples were analyzed using an Autoflex MALDI-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany); the analyzer worked in the linear mode and positive ions were recorded in the mass range between 2 and 13 kDa. Mass calibration was performed after every four samples using standards in the range of 2.8-16.9 kDa. Prior to analysis each sample was loaded onto a ZipTip C18 tip-microcolumn by passing it through repeatedly 10 times, column was washed with water and then eluted with 1 μl of matrix solution (30 mg/ml sinapinic acid in 50% ACN/H_2O and 0.1% TFA with addition of 1 mM n-octyl glucopyranoside) directly onto the 600 μm AnchorChip (Bruker Daltonics) plate. ZipTip extraction/loading was repeated twice for each sample and for each spot on the plate two spectra were acquired (i.e., four spectra were recorded for each sample). Randomization in blocks was used in spectra registration to avoid a possible batch effect.

Spectral data processing. The preprocessing of spectral data that included alignment and averaging of technical repeats, binning of neighboring points to reduce data complexity, removal of the spectral area below baseline and normalization of the total ion current (TIC), was performed according to procedures considered to be standard in the field (20,21). In the second step spectral components, which reflected $[M+H]^+$ peptide ions recorded at defined m/z values, were identified using decomposition of mass spectra into a mixture of Gaussian components as described elsewhere (22) followed by several post-processing steps. The average spectrum was decomposed into a sum of 300 Gaussian bell-shaped curves by using a variant of the expectation maximization (EM) algorithm (23) and Bayesian Information Criterion (BIC) for model selection (24). The initial set of 300 Gaussian components, defined by their mean values and standard deviations was further divided into two subsets according to their coefficients of variation (CV); a threshold value was obtained by decomposing the CV density function into a mixture of two normal density functions and employing the maximum probability criterion for their classification. Components with standard deviations bigger than the threshold value (i.e., 0.17% of the m/z value), which presumably represented the residual baseline, were excluded from analyses. Additionally, considering the real resolution of registered spectra, overlaping components were merged if their mean values were closer than 0.1% of the m/z value and were homogeneous in variances (verified by F test). The post-processing procedure resulted in dimension reduction from 300 to 131 Gaussian components. The final Gaussian components were used to compute features of registered spectra (termed spectral components afterward) for all samples by the operations of convolutions with Gaussian masks. The knowledge base EPO-KB (Empirical Proteomic Ontology Knowledge Base) (25), which annotates registered m/z values to known peptide/proteins, was used for hypothetical identification of spectral components assuming their mono-protonation and allowing for a 0.5% mass accuracy limit.

Table I. Characteristics of analyzed groups of blood donors.

Group		n	Age (median and range)	T1+T2 score	T3 score	T4 score	N^+	M1
Healthy volunteers	**Ctr.**	45	49 (26-72)	-	-	-	-	-
Head and neck cancer patients	**HNC**	35	54 (40-75)	42%	29%	29%	46%	0%
Colorectal cancer patients	**CRC**	35	61 (36-78)	20%	65%	15%	58%	23%
Lung cancer patients	**LC**	50	66 (52-78)	28%	18%	54%	90%	34%

Statistical analyses. The similarity measure proposed by Frank and coworkers (26) was used for evaluation of similarities of mass profiles within and between groups; a modified measure that based on all 131 spectral components was used. The intra-group similarity was defined as the mass profile similarity calculated pairwise between the group members, while the inter-group similarity was calculated by the pairwise comparison of individuals from different groups. To provide optimal statistical tools for selection of discriminative spectral components the Lilliefors's test was applied to assess for distribution normality and Bartlett's test was employed to check homogeneity of variances. Because the statistically significant heteroscedasticity of spectral components was observed the modified Welch statistics (MWT) (27) was chosen for selection of discriminating components. Due to numerous spectral components analyzed a correction for multiple testing was applied; Storey's q-values (28) with thresholds for FDR (false discovery rate) equal to 0.05 were used. Support vector machine (SVM) technique with a linear kernel function (29) combined with the feature ranking based on the MWT values was used for sample classification. In all experiments predictive performance of the classifier was measured by the average error rate together with sensitivity and specificity level; multiple random validation procedure (i.e., Monte Carlo cross-validation) was used for the classifier validation. To demonstrate the structure and determinant power of the information included in spectral components an unsupervised hierarchical clustering was performed; the Euclidean metric distance between mass profiles with the criterion of complete linkage was applied.

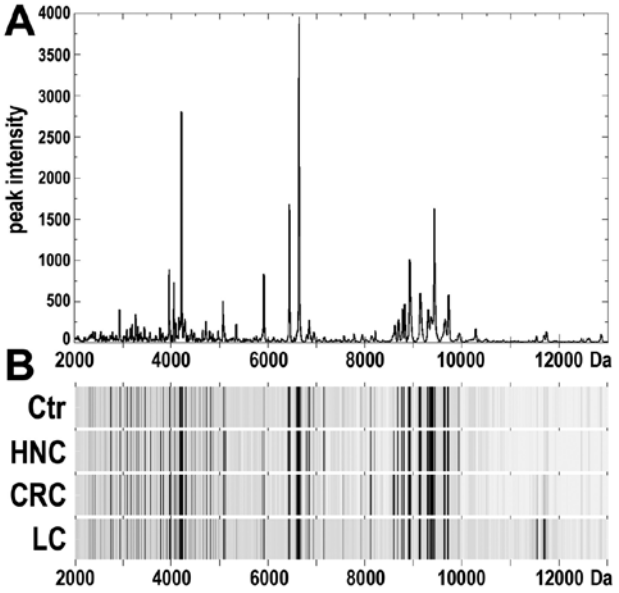

Figure 1. Mass profile of the low-molecular-weight fraction of serum proteome. (A) An average mass spectrum registered in the range of 2,000-13,000 Da. (B) Average spectra from control healthy donors (Ctr) and patients with head and neck (HNC), colorectal (CRC) and lung (LC) cancers.

Results

The low-molecular-weight peptide component of serum samples collected before the start of an anticancer treatment from patients with head and neck squamous cell cancer (HNC), colorectal adenocarcinoma (CRC), non-small cell lung cancer (LC) and corresponding group of healthy controls (Ctr) was analyzed by mass spectrometry. One hundred and thirty-one spectral components that corresponded to $[M+H]^+$ peptide ions registered with specific m/z value and abundance defined the mass profile of each serum specimen. Fig. 1A shows a typical mass spectrum of serum proteome registered in analyzed material by MALDI-ToF spectrometer in the 2,000 to 13,000 Da range. Average profiles of mass spectra appeared comparable among analyzed groups of donors, which is shown graphically in Fig. 1B. In the first step of statistical analysis general similarities between serum proteome profiles registered for different groups of samples were computed. Based on the abundance of each spectral component the overall similarity of proteome profiles was established for each pair of serum samples collected from different donors, and then average similarity within or between groups was assessed based on all pairwise analyses. Fig. 2A shows examples of such pairwise analyses of samples with high and low level of similarity of proteome profiles (S-value). The similarities of registered serum proteome profiles between different donors were computed for both intra-group and inter-group comparisons and characterized by the median value of all pairwise analyses, which is presented in Table II (Fig. 2B shows graphic examples of such results). We found that general intra-group similarities of serum proteome profiles were comparable for all four groups of donors. The similarity of serum proteome profiles among healthy donors was slightly higher (median S-value 0.797) when compared to inter-personal similarities within each of three cancer patient groups (median S-value in a range 0.764-0.774); however, such differences were not statistically significant. Importantly, levels of similarities of serum proteome profiles identified between groups of donors (median S-value ranged from 0.751 to 0.777) were comparable to levels of similarities detected inside each group. We concluded that general features of serum proteome profiles were similar in all analyzed groups of donors, which apparently reflected a homeostatic nature of the analyzed tissue.

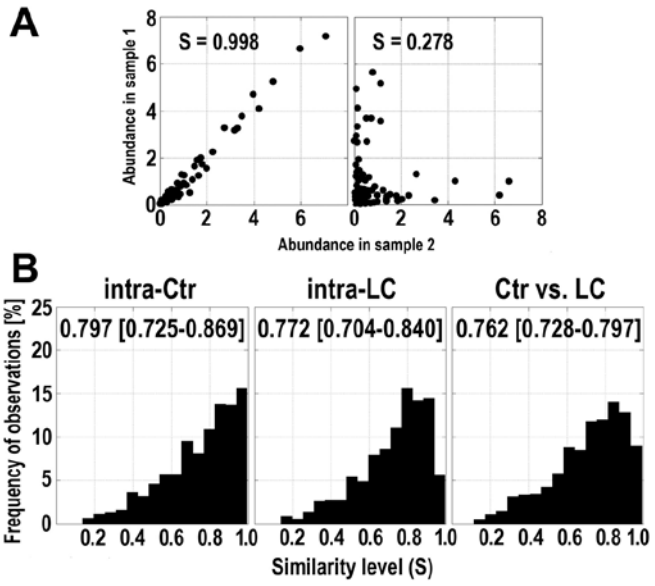

Figure 2. Assessment of the overall similarity of mass profiles. (A) Examples of pairs of serum samples with high (S=0.998, left panel) and low (S=0.278, right panel) degree of the overall similarity (each dot represents one spectral component). (B) Histograms showing examples of intra- and inter-group similarity measures (numbers show median values of similarities between samples and their 95% confidence intervals).

Table II. Similarity of serum proteome profiles within and between groups of donors.

Analyzed groups	Median	95% CI
Intra-group similarity		
Ctr.	0.797	0.725-0.869
HNC	0.774	0.696-0.852
CRC	0.764	0.669-0.858
LC	0.772	0.704-0.840
Inter-group similarity		
Ctr. vs. **HNC**	0.777	0.744-0.809
Ctr. vs. **LC**	0.762	0.728-0.797
Ctr. vs. **CRC**	0.759	0.718-0.801
HNC vs. **CRC**	0.765	0.725-0.806
HNC vs. **LC**	0.751	0.716-0.786
LC vs. **CRC**	0.775	0.745-0.806

Shown are median values of similarities between samples and their 95% confidence intervals.

Next we searched for differentiating spectral components (i.e., m/z peptide ions) and build classifiers, which allowed discriminating serum proteome profiles and identifying peptide signatures characteristic for each of four groups of donors. Table III shows numbers of spectral components that differentiated each pair of compared groups in univariate analyses (at a 0.05 and 0.001 q-value significance cut-off level), and characteristics of corresponding classifiers: number of components (features) in optimal classifiers, total errors of classification, sensitivity and specificity of optimal classifiers. We found similar reliability of classification when control samples of healthy donors were analyzed against samples from each group of cancer patients (sensitivities in a range 69-74%, specificities in a range 81-84%). However, such three cancer classifiers (signatures) were built of different numbers of different components. Table IV shows characteristics of spectral components, which were the most important for such classification and differentiation between controls and three types of cancer samples. There were only two components of classifiers (i.e., registered serum peptide ions), which at high level of statistical significance discriminated between control samples and all three types of cancer samples. Abundance of these peptides with registered m/z values 3766 and 5867 Da was higher in serum of cancer patients than in serum of healthy controls (Fig. 3). Other most frequent components of cancer classifiers appeared to be more unique for particular groups of patients, e.g., components 11,511 and 11,667 Da (hypothetically annotated as fragments of serum amyloid A, SAA1) were specific for differences between control and lung cancer samples. In the next step we searched for features of serum proteome profiles that differentiated between three groups of cancer samples; Table V shows characteristics of peptide ions, which were the most essential for such differentiation. We found that reliable classification (i.e., good discrimination) could be obtained for comparison between lung cancer (LC) and head and neck cancer (HNC) samples, and slightly worse classification for comparison between LC and colorectal cancer (CRC) samples. In addition, the same spectral components were important for differentiation between LC and HNC samples, and between LC and CRC samples. Noteworthy, spectral components that putatively corresponded to fragments of SAA1 (m/z = 11,511 and 11,667 Da) differentiated LC samples from either HNC, CRC and healthy control samples (Fig. 4A). In marked contrast, none of spectral components discriminated between HNC and CRC samples in univariate analyses, and reliable classification was not possible when these two types of samples were compared (Table III).

Finally we searched for association between features of serum proteome profiles and clinical progress of malignancies (i.e., clinical stages according to the TNM system) in overall group of cancer patients. Statistical tests were applied to identify spectral components that discriminated between groups of patients differing in clinical stage of primary tumor (T), lymph node status (N) and distant metastases (M). We found several of such differences, yet none of them remained statistically significant when the correction against multiple testing was applied (data not shown). However, when we focused on spectral components essential for discrimination between lung cancer samples and other types of cancer samples, i.e., 11,511 and 11,667 Da, clear tendency for association between abundance of these peptides and the degree of progression of malignancy was noted. Fig. 4B shows that average abundance of the 11,511 Da component was higher in serum of patients with highest stage of primary tumor (T4 vs. T1+T2+T3) and with distant metastases (M1 vs. M0), while the difference between subgroups with different lymph node status (N^+ vs. N0) was less evident (the same differences were observed for the 11,667 Da component). It had to be noted that patients with lung cancer dominated in subgroups of patients with highest clinical stage of primary tumor (27 out of 41) and with distant metastases (17 out of 26).

Table III. Characteristics of peptide signatures built to differentiate groups of blood donors.

Compared groups	Differentiating components		Optimal classifiers			
	q<0.05	q<0.001	No. of components	Total error (%)	Sensitivity (%)	Specificity (%)
Ctr. HNC	9	3	2	19.9	74.1	83.6
Ctr. LC	95	16	7	22.9	71.7	82.6
Ctr. CRC	54	3	4	23.7	68.6	81.9
HNC CRC	0	0	50	40.0	63.1	56.3
HNC LC	70	12	3	21.7	78.6	77.9
LC CRC	17	4	8	27.5	73.6	71.1

Shown are numbers of differentiating components in univariate analyses (q<0.05 and q<0.001), and numbers of features (components), total errors, sensitivities and specificities in optimal (best performing) multi-component classifiers.

Table IV. Examples of spectral components that discriminated between serum samples collected from healthy controls and cancer patients.

m/z (Da)	Ctr. vs. HNC		Ctr. vs. CRC		Ctr. vs. LC		Hypothetical identity
	q-value	%	q-value	%	q-value	%	
2425	0.306	0	**1.48E-03**	**23**	0.013	1	-
2786	0.301	0	0.058	1	**2.09E-06**	**82**	HAMP (frag. 60-84)
3766	**2.59E-04**	**61**	**6.23E-05**	**91**	7.80E-05	44	-
5867	**7.77E-08**	**93**	**2.07E-05**	**84**	**5.50E-05**	**46**	-
6429	0.242	0	0.268	0	**7.03E-05**	**48**	APOC1 (frag. 27-81)
8598	0.046	1	**1.72E-04**	**74**	**5.11E-04**	**20**	C4A (frag. 680-755)
11511	0.126	0	0.221	0	**1.31E-06**	**97**	SAA1 (frag. 20-122)
11667	0.126	0	0.164	0	**2.09E-06**	**91**	SAA1 (frag. 19-122)
11717	0.046	1	6.79E-03	16	**5.50E-05**	**51**	B2M (frag. 21-119)

Shown are m/z values of components, q-values of differences in univariate analyses and frequencies in classifiers for each pairwise analysis; the most frequent components of classifiers are marked in bold characters. Hypothetical identity of fragments of peptides based on annotation of registered m/z values at the EPO-KB knowledge base: HAMP, hepcidin; APOC1, apolipoprotein C1; C4A, component C4A; SAA1, serum amyloid A; B2M, β2-microglobulin.

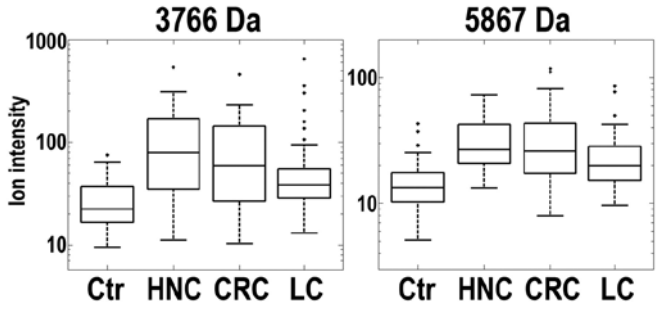

Figure 3. Abundance of spectral components that differentiated samples of control healthy donors and cancer patients (box-plots show minimum, lower quartile, median, upper quartile and maximum values; outliers are marked by asterisks).

Hence, differences observed between subgroups of patients with less-advanced and more-advanced cancer could simply reflect cancer type-specific features of peptide signatures characteristic for lung cancer. To exclude such possibility the analysis of association between abundance of spectral components and the clinical stages according to the TNM system was performed also for the group of lung cancer patients only. Fig. 4C shows that in the group of lung cancer patients average abundance of the 11,511 Da component was also higher in serum of patients with malignancies at clinical stages T4 and/or M1 (the same tendency was observed for the 11,667 Da component). In addition, other spectral components essential for discrimination between LC samples and HNC/CRC samples (i.e., 3556, 6438, 8917, 9137 Da) also showed tendency to correlate with a clinical progression of malignancy, both in the all-cancer patient and in the lung cancer patient groups (all this components were less abundant in LC samples when compared to HNC/CRC samples, and in general their lower abundance appeared to correlate with more advanced cancer cases). We concluded that differences observed between LC samples and both CRC and HNC samples at least partially

Table V. Examples of spectral components that discriminated between different groups of cancer patients.

	HNC vs. CRC		HNC vs. LC		CRC vs. LC		
m/z (Da)	q-value	%	q-value	%	q-value	%	Hypothetical identity
3556	0.333	8	6.01E-04	6	**5.00E-04**	**72**	-
3886	0.177	26	2.59E-03	0	**2.99E-03**	**51**	-
6438	0.302	8	3.07E-04	7	**5.00E-04**	**89**	APOC1 (frag. 29-83)
8676	0.302	8	2.59E-03	1	**2.33E-03**	**53**	-
8917	0.251	17	1.03E-03	3	**5.00E-04**	**81**	C3 (frag. 672-747)
8928	0.145	33	1.03E-02	0	**2.47E-03**	**52**	-
9137	0.061	82	**2.77E-05**	**30**	5.40E-02	5	HP (frag. 79-160)
9414	**0.061**	**97**	8.31E-05	1	0.293	0	-
11511	0.101	45	**2.39E-08**	**92**	**5.00E-04**	**82**	SAA1 (frag. 20-122)
11667	0.081	59	**4.78E-08**	**86**	**2.33E-03**	**56**	SAA1 (frag. 19-122)

Shown are m/z values of components, q-values of differences in univariate analyses and frequencies in classifiers for each pairwise analysis; the most frequent components of classifiers are marked in bold characters. Hypothetical identity of fragments of peptides based on annotation of registered m/z values at the EPO-KB knowledge base: APOC1, apolipoprotein C1; C3, component C3; HP, haptoglobin; SAA1, serum amyloid A.

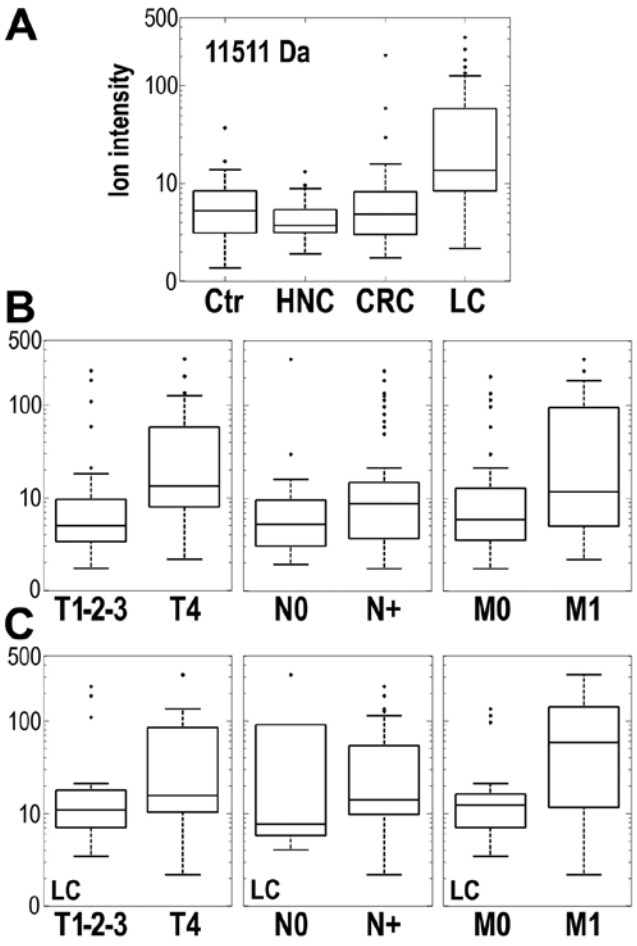

Figure 4. Abundance of the 11,511 Da spectral component. (A) Comparison of abundance of the 11,511 Da component in samples of healthy donors and cancer patients. (B) Abundance of the 11,511 Da component in samples of all cancer patient groups differing in clinical stage of primary tumor (T), lymph node status (N) and distant metastases (M). (C) Abundance of the 11,511 Da component in samples of lung cancer patients with different TNM stages (box-plots show minimum, lower quartile, median, upper quartile and maximum values; outliers are marked by asterisks).

reflected the presence of serum peptides characteristic for more advanced stages of a disease, which cases were most frequent in group of patients with lung cancer.

To further characterize differences and similarities between serum proteome profiles characteristic for compared groups of donors an unsupervised hierarchical cluster analysis was performed (Fig. 5). Apparently we did not observe any clear clustering of four analyzed groups of donors. However, when two major clusters were compared it appeared that the majorities of samples from healthy donors (together with samples from HNC patients) and samples from lung cancer patients (together with samples from CRC patients) were present in separate clusters: the former ones were over-represented in cluster 1 (71 and 62%, respectively) while the latter ones in cluster 2 (64 and 60%, respectively). Furthermore, it appeared that patients with more advanced malignancies were more frequent in cluster 2 (e.g., distant metastases were observed in 18 and 27% of cancer patients in cluster 1 and cluster 2, respectively). Thus, in spite of large heterogeneity of samples, the unsupervised analyses indicated that differences between peptide signatures characteristic for healthy donors and lung cancer patients were the most distinct, and that features of serum proteome profiles were associated with a disease progression.

Discussion

There are several studies published where MALDI/SELDI-based analyses of the low-molecular-weight fraction of serum/plasma proteome were applied for identification of peptide signatures specific for patients with head and neck cancer (30-34), colorectal cancer (19,35-38) and lung cancer (39-44). All these works proposed differentiating peptides and multi-peptide signatures that discriminated between specimens collected from cancer patients and healthy donors, yet such cancer signatures were built of apparently different spectral components. Some peptide ions appeared repetitively in several signatures (based on similarities

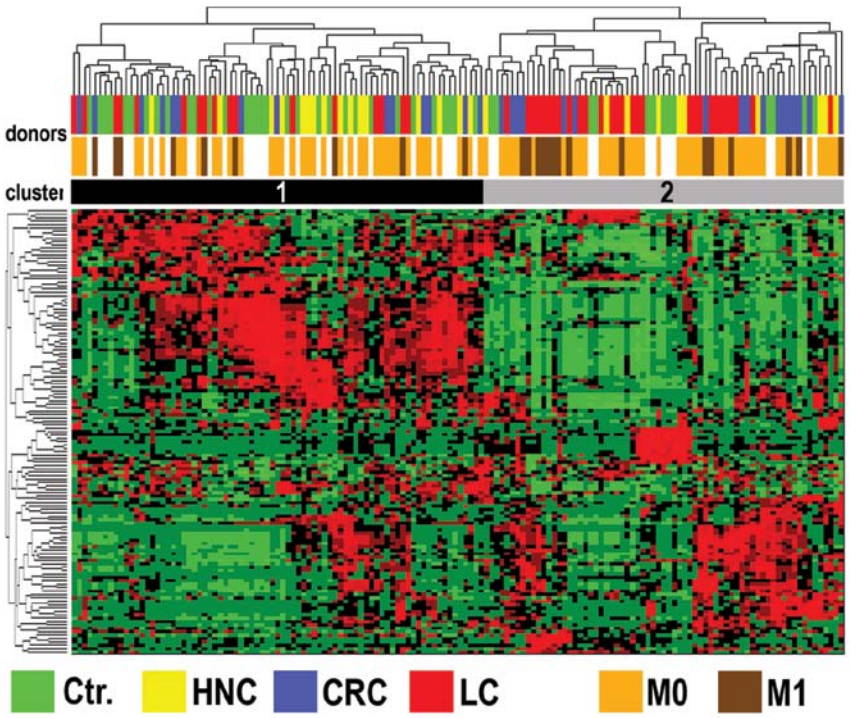

Figure 5. Unsupervised hierarchical cluster analysis of serum samples of healthy donors and cancer patients. Two major clusters of samples are separated; color bars represent different groups of donors (upper panel), and cancer patients with or without distant metastases (bottom panel) (increased abundance of a spectral component is marked in red, decreased abundance is marked in green).

of their registered m/z values), but only few of them (e.g., ~5,900 and ~11,700 Da) were reported as important for discrimination between healthy controls and all these three types of cancer. However, because of different methodological approaches implemented in these works conclusions on real similarities and specificities of proposed signatures were impossible. Here we performed simultaneous analysis of four groups of donors, which allowed direct comparison of specific peptide signatures. We found that only two spectral components, i.e., 3766 and 5867 Da, were present in all three cancer signatures, while other components discriminating between control and cancer samples were more unique for analyzed groups of patients. Peptide ion with registered m/z value 5867 Da putatively corresponded to peptide ions ~5900 Da reported earlier as discriminating between control and cancer samples (19,31,32,40), yet verification of this component's resemblance would have required its direct sequence identification. Nevertheless, based on literature reports and presented data we concluded that peptide signatures differentiating healthy donors from head and neck, colorectal and lung cancers patients were most apparently built of two types of components - common for all cancer cases and specific for patient groups.

In the second part of analyses we searched for serum proteome features that were putatively cancer type-specific and might allow discrimination between groups of cancer patients. We found that abundance of several spectral components was different in samples of lung cancer patients and in samples of patients with either colorectal or head and neck cancer. The latter two types of samples could not be discriminated based on features of serum proteome profiles even though compared cancers represented different histological and molecular types of malignancies (i.e., adenocarcinomas and squamous cell carcinomas). The analyzed group of lung cancer consisted of patients with more advanced tumors as compared to groups of head and neck or colorectal cancer patients, which suggested existence of a correlation between features of serum proteome and clinical stage of cancer. In fact, we observed a clear tendency toward association between stage of the cancer and abundance of serum peptides essential for discrimination between lung cancer samples and other cancer samples (i.e., 3556, 6438, 8917, 9137, 11511 and 11667 Da components). Most importantly, this association was noted not only when overall group of cancer patients was analyzed (which could reflect overrepresentation of lung cancer patients in a subgroup of patient with more advanced cancers) but also when more homogeneous group of lung cancer patients was analyzed separately. All these observations suggested that serum proteome features that differentiated analyzed groups of cancer patients were rather cancer stage-specific than cancer type-specific. Such possibility could be rationally justified if we assumed that the most characteristic changes observed in the serum proteome profiles reflected general response of human organism to malignancy, and that intensity of changes depended on overall escalation of the disease rather than specific histological or molecular features of a cancer type.

Among serum components with elevated levels in the group of patients with lung cancer and associated with more advanced cancer cases there were two components, i.e. 11,511 and 11,667 Da, which putatively corresponded to fragments of serum amyloid A1 (hypothetical identification based on annotation at the EPO-KB database). SAA1 is an acute phase apolipoprotein typically induced in liver in response to inflammatory stimuli. However, increased expression of SAA1 was also observed during tumorigenesis and elevated serum level of this protein was a general feature of progressive and metastatic cancer cases, hence

being proposed as a prognostic cancer marker (17). Importantly, increased levels of ~11,5 and/or ~11,7 kDa serum peptides, which corresponded to fragments of SAA1, were already detected by MALDI/SELDI-based proteomics approaches in serum of patients with different types of advanced cancers, e.g., renal cancer (45,46), ovarian cancer (47), pancreatic cancer (48), prostate cancer (49), osteosarcoma (50), neuroblastoma (51), also including malignancies addressed in this work - head and neck cancer (52), colon cancer (53) and lung cancer (44,54,55). Consequently, all these reports indicated that SAA1 fragments were indeed cancer stage-specific but not cancer type-specific components, which validated our hypothesis on the nature of major features of cancer signatures that based on serum proteome mass profiles.

Finally, we concluded that peptides putatively involved in the systemic response of human organism to malignancy, whose abundances were associated with a general progression of the disease, predominated in identified cancer signatures, which apparently reduced the possibility to build serum proteome-based markers specific for different molecular and histological types of cancer.

Acknowledgements

This work was supported by the Polish Ministry of Science and Higher Education, Grant N402 3506 38.

References

1. Hanash S: Disease proteomics. Nature 422: 226-232, 2003.
2. Wulfkuhle JD, Liotta LA and Petricoin EF: Proteomic applications for the early detection of cancer. Nat Rev Cancer 3: 267-275, 2003.
3. Aebersold R and Mann M: Mass spectrometry-based proteomics. Nature 422: 198-207, 2003.
4. Liotta LA, Ferrari M and Petricoin EF: Clinical proteomics: written in blood. Nature 425: 905, 2003.
5. Palmblad M, Tiss A and Cramer R: Mass spectrometry in clinical proteomics - from the present to the future. Proteomics Clin Appl 3: 6-17, 2009.
6. Li L, Tang H, Wu Z, Gong J, Gruidl M, Zou J, Tockman M and Clark RA: Data mining techniques for cancer detection using serum proteomic profiling. Artif Intel Med 32: 71-83, 2004.
7. Conrads TP, Hood BL, Issaq HJ and Veenstra TD: Proteomic patterns as a diagnostic tool for early-stage cancer: a review of its progress to a clinically relevant tool. Mol Diagn 8: 77-85, 2004.
8. Dworzanski JP and Snyder AP: Classification and identification of bacteria using mass spectrometry-based proteomics. Expert Rev Proteomics 2: 863-878, 2005.
9. Somorjai RL: Pattern recognition approaches for classifying proteomic mass spectra of biofluids. Methods Mol Biol 428: 383-396, 2008.
10. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC and Liotta LA: Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359: 572-577, 2002.
11. Posadas EM, Simpkins F, Liotta LA, MacDonald C and Kohn EC: Proteomic analysis for the early detection and rational treatment of cancer-realistic hope? Ann Oncol 16: 16-22, 2005.
12. Azad NS, Rasool N, Annunziata CM, Minasian L, Whiteley G and Kohn EC: Proteomics in clinical trials and practice. Mol Cell Proteomics 5: 1819-1829, 2006.
13. Solassol J, Jacot W, Lhermitte L, Boulle N, Maudelonde T and Mangé A: Clinical proteomics and mass spectrometry profiling for cancer detection. Expert Rev Proteomics 3: 311-320, 2006.
14. Cho WC: Proteomics technologies and challenges. Genomics Proteomics Bioinformatics 5: 77-85, 2007.
15. Cho WC and Cheng CH: Oncoproteomics: current trends and future perspectives. Expert Rev Proteomics 4: 401-410, 2007.
16. Hortin GL: The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. Clin Chem 52: 1223-1237, 2006.
17. Malle E, Sodin-Semrlb S and Kovacevica A: Serum amyloid A: an acute-phase protein involved in tumour pathogenesis. Cell Mol Life Sci 66: 9-26, 2009.
18. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI and Tempst P: Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 116: 271-284, 2006.
19. Engwegen JY, Helgason HH, Cats A, Harris N, Bonfrer JMG, Schellens JHM and Beijnen JH: Identification of serum proteins discriminating colorectal cancer patients and healthy controls using surface-enhanced laser desorption ionisation-time of flight mass spectrometry. World J Gastroenterol 12: 1536-1544, 2006.
20. Karpievitch YV, Hill EG, Smolka AJ, Morris JS, Coombes KR, Baggerly KA and Almeida JS: PrepMS: TOF MS Data Graphical Preprocessing Tool. Bioinformatics 23: 264-265, 2007.
21. Hilario M, Kalousis A, Pellegrini C and Müller M: Processing and classification of protein mass spectra. Mass Spectrometry Rev 25: 409-449, 2006.
22. Pietrowska M, Marczak Ł, Polańska J, Behrendt K, Nowicka E, Walaszczyk A, Chmura A, Deja R, Stobiecki M, Polański A, Tarnawski R and Widłak P: Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early stage breast cancer. J Translat Med 7: e60, 2009.
23. Hastie T, Tibshirani R and Friedman JH: The Elements of Statistical Learning. Springer Verlag, 2001.
24. Schwarz GE: Estimating the dimension of a model. Ann Statistics 6: 461-464, 1978.
25. Lustgarten JL, Kimmel C, Ryberg H and Hogan W: EPO-KB: a searchable knowledge base of biomarker to protein links. Bioinformatics 24: 1418-1419, 2008.
26. Frank AM, Bandeira N, Shen Z, Tanner S, Briggs SP, Smith RD and Pevzner PA: Clustering millions of tandem mass spectra. J Proteome Res 7: 113-122, 2008.
27. Demissie M, Mascialino B, Calza S and Pawitan Y: Unequal group variances in microarray data analyses. Bioinformatics 24: 1168-1174, 2008.
28. Storey JD: A direct approach to false discovery rates. J Roy Statis Soc B 64: 79-49, 2002.
29. Vapnik V: The Nature of Statistical Learning Theory. Springer, 1995.
30. Cheng AJ, Chen LC, Chien KY, Chen YJ, Chang JTC, Wang HM, Liao CT and Chen IH: Oral cancer plasma tumor marker identified with bead-based affinity-fractionated proteomic technology. Clin Chem 51: 2236-2244, 2004.
31. Soltys SG, Le QT, Shi G, Tibshirani R, Giaccia AJ and Koong AC: The use of plasma surface-enhanced laser desorption/ionization time-of-flight mass spectrometry proteomic patterns for detection of head and neck squamous cell cancers. Clin Cancer Res 10: 4806-4812, 2004.
32. Wadsworth JT, Somers KD, Cazares LH, Malik G, Adam B-L, Stack BC, Wright GL and Semmes OJ: Serum protein profiles to identify head and neck cancer. Clin Cancer Res 10: 1625-1632, 2004.
33. Wei Y-S, Zheng Y-H, Liang W-B, Zhang J-Z, Yang Z-H, Lv M-L, Jia J and Zhan L: Identification of serum biomarkers for nasopharyngeal carcinoma by proteomic analysis. Cancer 112: 544-551, 2008.
34. Huang Y-J, Xuan C, Zhang B-B, Liao M, Deng K-F, He M and Zhao J-M: SELDI-TOF MS profiling of serum for detection of nasopharyngeal carcinoma. J Exp Clin Cancer Res 28: e85, 2009.
35. Ward DG, Suggett N, Cheng Y, Wei W, Johnson H, Billingham LJ, Ismail T, Wakelam MJO, Johnson PJ and Martin A: Identification of serum biomarkers for colon cancer by proteomic analysis. Br J Cancer 94: 1898-1905, 2006.
36. Wang Q, Shen J, Li Z, Jie J, Wang W, Wang J, Zhang Z, Li Z, Yan L and Gu J: Limitations in SELDI-TOF MS whole serum proteomic profiling with IMAC surface to specifically detect colorectal cancer. BMC Cancer 9: e287, 2009.
37. Liu M, Li C-F, Chen H-S, Lin L-Q, Zhang C-P, Zhao J-L, Liu Y, Zhang S-J, Jin J-C, Wang L and Liu J-R: Differential expression of proteomics models of colorectal cancer, colorectal benign disease and healthy controls. Proteome Sci 8: e16, 2010.
38. Liao CCL, Mehta A, Ward NJ, Marsh S, Arulampalam T and Norton JD: Analysis of post-operative changes in serum protein expression profiles from colorectal cancer patients by MALDI-TOF mass spectrometry: a pilot methodological study. World J Surg Oncol 8: e33, 2010.

39. Howard BA, Wang MZ, Campa MJ, Corro C, Fitzgerald MC and Patz EF Jr: Identification and validation of a potential lung cancer serum biomarker detected by matrix-assisted laser desorption/ionization-time of flight spectra analysis. Proteomics 3: 1720-1724, 2003.
40. Xiao X, Liu D, Tang Y, Guo F, Xia L, Liu J and He D: Development of proteomic patterns for detecting lung cancer. Dis Markers 19: 33-39, 2004.
41. Yang S, Xiao X, Zhang W, Zhang L, Zhang W, Zhou B, Chen G and He D: Application of serum SELDI proteomic patterns in diagnosis of lung cancer. BMC Cancer 5: e83, 2005.
42. Han KQ, Huang G, Gao CF, Wang XL, Ma B, Sun LQ and Wei ZJ: Identification of lung cancer patients by serum protein profiling using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. Am J Clin Oncol 31: 133-139, 2008.
43. Du J, Yang S, Lin X, Bu L, Nan Y, Huo S and Shang W: Use of anchorchip-time-of-flight spectrometry technology to screen tumor biomarker proteins in serum for small cell lung cancer Diagn Pathol 5: e60, 2010.
44. Cremona M, Calabro E, Randi G, De Bortoli M, Mondellini P, Verri C, Sozzi G, Pierotti MA, La Vecchia C, Pastorino U and Bongarzone I: Elevated levels of the acute-phase serum amyloid are associated with heightened lung cancer risk. Cancer 116: 1326-1335, 2010.
45. Tolson J, Bogumil R, Brunst E, Beck H, Elsner R, Humeny A, Kratzin H, Deeg M, Kuczyk M, Mueller GA, Mueller CA and Flad T: Serum protein profiling by SELDI mass spectrometry: detection of multiple variants of serum amyloid alpha in renal cancer patients. Lab Invest 84: 1220-1221, 2004.
46. Engwegen JY, Mehra N, Haanen JB, Bonfrer JM, Schellens JH, Voest EE and Beijnen JH: Validation of SELDI-TOF MS serum protein profiles for renal cell carcinoma in new populations. Lab Invest 87: 161-172, 2007.
47. Moshkovskii SA, Serebryakova MV, Kuteykin-Teplyakov KB, Tikhonova OV, Goufman EI, Zgoda VG, Taranets IN, Makarov OV and Archakov AI: Ovarian cancer marker of 11.7 kDa detected by proteomics is a serum amyloid A1. Proteomics 5: 3790-3797, 2005.
48. Koomen JM, Shih LN, Coombes KR, Li D, Xiao LC, Fidler IJ, Abbruzzese JL and Kobayashi R: Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins. Clin Cancer Res 11: 1110-1118, 2005.
49. Le L, Chi K, Tyldesley S, Flibotte S, Diamond DL, Kuzyk MA and Sadar MD: Identification of serum amyloid A as a biomarker to distinguish prostate cancer patients with bone lesions. Clin Chem 51: 695-707, 2005.
50. Li Y, Dang TA, Shen J, Perlaky L, Hicks J, Murray J, Meyer W, Chintagumpala M, Lau CC and Man TK: Identification of a plasma proteomic signature to distinguish pediatric osteosarcoma from benign osteochondroma. Proteomics 6: 3426-3435, 2006.
51. Combaret V, Bergeron C, Brejon S, Iacono I, Perol D, Negrier S and Puisieux A: Protein chip array profiling analysis of sera from neuroblastoma patients. Cancer Lett 228: 91-96, 2005.
52. Cho WC, Yip TT, Ngan RK, Yip TT, Podust VN, Yip C, Yiu HH, Yip V, Cheng WW, Ma VW and Law SC: ProteinChip array profiling for identification of disease- and chemotherapy-associated biomarkers of nasopharyngeal carcinoma. Clin Chem 53: 241-250, 2007.
53. Roelofsen H, Alvarez-Llamas G, Dijkstra M, Breitling R, Havenga K, Bijzet J, Zandbergen W, De Vries MP, Ploeg RJ and Vonk RJ: Analyses of intricate kinetics of the serum proteome during and after colon surgery by protein expression time series. Proteomics 7: 3219-3228, 2007.
54. Yildiz PB, Shyr Y, Rahman JS, Wardwell NR, Zimmerman LJ, Shakhtour B, Gray WH, Chen S, Li M, Roder H, Liebler DC, Bigbee WL, Siegfried JM, Weissfeld JL, Gonzalez AL, Ninan M, Johnson DH, Carbone DP, Caprioli RM and Massion PP: Diagnostic accuracy of MALDI mass spectrometric analysis of unfractionated serum in lung cancer. J. Thorac Oncol 2: 893-901, 2007.
55. Cho WC, Yip TT, Cheng WW and Au JSK: Serum amyloid A is elevated in the serum of lung cancer patients with poor prognosis. Br J Cancer 102: 1731-1735, 2010.