Regular paper

Pre-analytical-related variability influencing serum peptide profiles demonstrated in a mass spectrometry-based search for colorectal and prostate cancer biomarkers

Jakub Karczmarski[1#], Tymon Rubel[2#], Michal Mikula[1], Jan Wolski[3], Andrzej Rutkowski[4], Edyta Zagorowicz[1,5], Michal Dadlez[6] and Jerzy Ostrowski[1,5✉]

[1]Department of Genetics, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology; Warsaw, Poland; [2]Institute of Radio-electronics, Warsaw University of Technology, Warsaw, Poland; [3]Department of Urology, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland; [4]Department of Gastrointestinal Cancer, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland; [5]Department of Gastroenterology and Hepatology, Medical Center for Postgraduate Education, Warsaw, Poland; [6]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

Although the degradome, which comprises proteolytic fragments of blood proteins, presents a potential source of diagnostic biomarkers, studies on cancer peptide biomarkers have provided inconsistent conclusions. In the present study, we reevaluated the usefulness of serum degradome analyses for searching peptide cancer biomarker candidates. Particular attention was paid to pre-analytical factors influencing the variability of determined peptide levels, including clotting time and control group selection. Studies were conducted on 44 and 86 serum samples collected from cancer patients and healthy individuals, respectively, using liquid chromatography electrospray ionization mass spectrometry (LC-ESI-MS)-based analyses. We identified 1373 unique peptides, nearly 40% of which originated from five blood proteins: fibrinogen alpha chain, apolipoprotein A-IV (APOA4), complement C3, apolipoprotein A-I, and alpha-1-antitrypsin. A set of 118 and 88 peptides exhibited highly significant differences (adjusted p-value ≤ 0.01 and fold change ≥ 2) in pair-wise comparisons of control vs. prostate cancer and control vs. colorectal cancer, respectively, with 37 peptides displaying a consistent direction of change for these pair-wise comparisons. The levels of 67 peptides differed significantly in serum samples collected from healthy individuals immediately prior to colonoscopy and those who underwent colonoscopic examination at least four weeks earlier. Of them, 49 peptides originated from APOA4. Whereas earlier studies, including ours, have utilized fragments of fibrinopeptide A (FPA) to distinguish cancer from healthy cases, here we show that their absolute abundance is a sensitive indicator of clotting time. These observations may have implications for future serum peptidome studies since these issues have not previously been recognized.

INTRODUCTION

Proteolytic cleavage of proteins is one of the key mechanisms that regulates protein expression and/or activity (Shen *et al.*, 2010). It was initially thought that blood biomarker peptides might represent tissue leakage fragments of cancer-specific proteins. However, further studies have revealed that the entire peptidomic fraction is mostly derived from abundant blood plasma/serum proteins.

Although the concept of peptidome and degradome can be used interchangeably, according to Shen and coworkers (Shen *et al.*, 2010), the degradome refers to the low molecular weight (LMW) fractions of the plasma/serum proteome composed of proteolytic fragments of blood proteins, generated by endoproteases. Endoproteases initially cleave serum proteins to biologically active peptides, including components of the complement and coagulation cascades, which may be further degraded to sets of smaller fragments by exopeptidases. Therefore, the peptidome dominates the plasma/serum peptide spectrum being versions of the same endogenous blood peptides shortened by one or more amino acids on both the N-terminal and the C-terminal end (Koomen *et al.*, 2005).

Native or processed peptides can be detected by mass spectrometry (MS)-based analyses. Until now, most analyses of LMW peptidome peptides in cancer research have utilized matrix-assisted laser desorption time of flight (MALDI-TOF) MS or surface-enhanced laser desorption/ionization-TOF (SELDI-TOF) MS platforms, enabling the identification of about 40 human blood degradome substrates and hundreds of degradome/peptidome peptides, as summarized by Shen et al. (Shen *et al.*, 2010). Some of them have been proposed as cancer biomarkers, and subsequently, selective and specific quantitative liquid chromatography tandem MS (LC-MS/MS) methods have been developed to simplify their analysis. Although the diagnostic value of degradome/peptidome peptides in clinical practices is still unresolved, the plasma/serum peptide profiles remain a potential source of diagnostic biomarkers.

Recently, we have successfully applied the liquid chromatography electrospray ionization MS (LC-ESI-MS) analytical platform optimized for "bottom-up" global anal-

ysis of the blood degradome/peptidome for the identification of peptide profiles that could distinguish colorectal cancer (CRC) patients from cancer-free controls (Bakun *et al.*, 2009). The main advantage of LC-ESI-MS over the widely used 1D MALDI spectra analysis is that LC-ESI allows a better separation of peptide MS signals in a complex mixture, thus reliably quantifying the peptide levels both in a relative and absolute way. In the present study, we reevaluated the utility of serum peptide analysis conducted by means of the LC-ESI-MS technique for searching peptide cancer biomarker candidates, with particular attention paid to pre-analytical factors that would significantly influence the variability of determined peptide levels.

MATERIALS AND METHODS

Ethics Statement. The study protocol was approved by the Bioethical Committee of the Medical Centre for Postgraduate Education, and all patients signed informed consent before inclusion.

Sample collection. Forty-four serum samples were prospectively collected at the Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology between April and September 2008 from 16 CRC patients (nine women and seven men, aged 42–80 (mean: 64 years); prognostic group II and B stage according to Jass and Astler-Coller classifications) and 28 prostate cancer (CaP) patients. All CRCs were newly diagnosed with an indication for surgical treatment, and the blood samples were collected prior to any therapy. In contrast, locally advanced or metastatic CaPs were diagnosed at least two years earlier (between 2 and 10 years; mean, 4.8 years), and at the initial diagnosis all CaP patients received radical prostatectomy and/or external beam radiation therapy and/or androgen deprivation therapy. At the time of serum sample collection, pathologic and biochemical progression-free outcomes were recognized in 13 patients (CaP-negative) (age 56–85; mean, 73 years), while cancer progression was noted in 15 patients (CaP-positive) (age 53–80; mean, 72 years). The control group consisted of healthy individuals with no illnesses or any known history of malignancy, normal results of screening colonoscopy, and normal mammography or Prostate-Specific Antigen (PSA) levels. Blood samples were collected from 30 control individuals (15 women and 15 men, aged 50–65; mean, 59 years) who underwent standard bowel preparation by polyethylene glycol (Fortrans) followed by a cleansing phosphate enema 30 minutes prior to the colonoscopy examination, and from another 26 controls (14 women and 12 men, aged 50–65; mean, 61 years) at least a month after the colonoscopy. All prospective patients in the CRC and CaP groups and the 26 control colonoscopy screening patients were not controlled for overnight fasting. In addition, on the same day blood samples were collected twice from 15 healthy laboratory staff; the first blood sample was withdrawn after 12 h overnight fasting and the second, 2 h after a breakfast meal.

Sample preparation. Serum harvesting, handling and processing were performed as described previously (Bakun *et al.*, 2009). Briefly, blood was obtained by venepuncture using Serum Gel S/7.5 ml collection tubes (Sarstedt S-Monovette), and then allowed to clot for 60 min at 25°C and centrifuged at $1\,300 \times g$ for 10 min at 4°C. Two hundred microliter aliquots of the serum were kept in 1.5 ml siliconized, polypropylene microtubes (Sigma-Aldrich, T4816) at –80°C until further use.

To assess the influence of clotting time on peptide levels, two blood aliquots were collected separately from 12 healthy volunteers. Blood aliquot #1 was allowed to clot for 1 h and aliquot #2 for 2 h at room temperature (water bath at 26ºC). Afterwards, aliquots #1 and #2 were immediately centrifuged and sera were transferred to –80ºC.

A mix of 100 µl serum + 300 µl 25% acetonitrile was centrifuged through the 5 kDa cutoff filtration membrane (Millipore Ultrafree-MC, UFC3LCC00), washed twice with 25% acetonitrile prior to use, and centrifuged at 5 000 rpm, at 15°C, for 4–5 h. The < 5 kDa fraction was lyophilized (Speed Vac) in siliconized microtubes (Sigma-Aldrich) and stored at –80°C for further use. Before LC-MS analysis, samples were dissolved in 0.1% trifluoroacetic acid.

For the acquisition of MS/MS data, pooled sera samples were prepared by combining 10 µl of serum from all the samples enrolled in the study.

Peptide synthesis. For absolute quantitation of fibrinopeptide A (FPA) ladder peptides, eight isotopically labelled N-terminally truncated versions of the peptide (from ADSGEGDFLAEGG*G*VR to FLAEGG*GVR, where G* indicates deuterated glycine) were used. Peptides were obtained by chemical synthesis as described previously (Bakun *et al.*, 2009).

LC-MS settings. Analysis was performed using the LTQ FT (Thermo Fisher Scientific, Waltham, MA, USA) mass spectrometer coupled to a nanoAcquity (Waters Corporation, Milford, MA, USA) LC system. Spectrometer parameters were as follows: capillary voltage, 2.5 kV; cone, 40 V; N_2 gas flow, 0; and m/z range, 300–2000. Samples were loaded from the autosampler tray (cooled to 10ºC) to the precolumn (Symmetry C18, 180 µm × 20 mm, 5 µm (Waters)) using a mobile phase of 100% MilliQ water acidified with 0.1% formic acid. Peptides were transferred to the nano-UPLC column (BEH130 C18, 75 µm × 250 mm, 1.7 mm (Waters)) by a gradient of 5–30% ACN in 0.1% formic acid in 45 min (250 nl/min), and then directly eluted to the ion source of the mass spectrometer. A blank run to ensure the lack of cross-contamination from previous samples preceded each analysis. Qualitative analyses were performed on pooled blood serum samples in data-dependent acquisition mode. Up to ten MS/MS processes were allowed for each MS scan. To increase the number of peptide identifications, three LC-MS/MS runs were performed per pooled sample, each covering one of three ranges of *m/z* values: 300–700, 600–900 and 800–2000. Quantitative analyses were carried out for individual samples in a separate profile type survey scan LC-MS run with an *m/z* measurement range of 300–2000.

Database search. MS/MS raw data files were processed to peak lists with the Mascot Distiller software (version 2.2.1, Matrix Science, London, UK). The resulting ion lists were searched using the Mascot search engine (version 2.2.03, Matrix Science, London, UK) against a composite database containing human protein entries from the Swissprot database and reversed versions of these entries. The search parameters were set as follows: enzyme specificity, none; fixed modification, none; variable modifications, oxidation (M); protein mass, unrestricted.

The peptide and fragment ions mass tolerances were established separately for individual LC-MS/MS runs after mass measurement error correction, as previously described (Mikula *et al.*, 2010). The statistical significance of peptide identifications was assessed using a joined target/decoy database search approach and a procedure

that provides q-value estimates for each peptide spectrum match (PSMs) in the dataset (Käll *et al.*, 2008; Mikula *et al.*, 2010). Only PSMs with q-values less than or equal to 0.01 were regarded as confidently identified. Mascot search results processing was carried out using a proprietary software tool implemented in Java programming language (MScan, available at http://proteom.ibb. waw.pl/).

Quantitative feature extraction. Peptide identifications from all the LC-MS/MS runs were merged into a common list, which was subsequently overlaid onto 2-D maps generated from the LC-MS profile data. This list was used to tag the corresponding peptide-related ion spectra on the basis of m/z difference, the deviation from the predicted elution time, and the match between the theoretical and observed isotopic envelopes. The relative abundance of each peptide was determined as the volume of a 2-D fit to the two most prominent peaks of the tagged isotopic envelope. A more detailed description of the employed feature extraction procedure implemented in MSparky software was previously presented in (Bakun *et al.*, 2009).

In the quality control step, samples with less than 60% reliably detected features were excluded from further analysis. Next, features with missing quantitative values in more than 20% of the samples were filtered out from the dataset. Finally, a modified k-nearest neighbours (Troyanskaya *et al.*, 2001) procedure was used for the imputation of the remaining missing values. The applied modifications included the following: the usage of correlation-based metric instead of Euclidean distance and variable neighbourhood, with the parameter k being the maximum size.

To minimize the effects of non-biological sources of variation, log-transformed peptide abundances were normalized by fitting a robust locally weighted regression smoother (LOESS) between the individual samples and a median pseudo-sample. The parameters of the fit were established using a set of features exhibiting low variance in the unnormalized data and then applied to the whole data set. After normalization different charge states of the same peptide were combined by summing their relative abundances.

Statistical analysis of quantitative MS results. For differentially expressed peptide selection, a non-parametric resampling-based test with t-statistics was used, and the resulting p-values were next adjusted for multiple hypothesis testing using a two step Benjamini–Hochberg procedure that controls for false discovery rate (FDR) (Benjamini & Hochberg, 1995). Peptides abundances with corrected p-values ≤ 0.01 and fold change (FC) values ≥ 2 in pair wise comparisons were considered as changed significantly. Principal component analysis (PCA) was used for graphical summarization and evaluation of the relationships among the studied samples. All statistical analyses were performed using proprietary software developed in MATLAB environment (MathWorks, Natick, MA, USA) (MStat, available at http://proteom. ibb.waw.pl/).

RESULTS

Peptide identification

LC-MS/MS analyses on pooled serum samples were performed to establish a MS/MS database of the LMW fraction for further quantitative comparisons between the samples (Fig. 1). In total, 14 LC-MS/MS runs of pooled serum samples were performed, resulting in 229 969 measured MS/MS spectra. A database search performed with the Mascot engine led to the identification of 1373 unique peptides with an estimated FDR of 0.01 (Table S1, all supplementary materials designed as S available on www.actabp.pl). Of them, 1304 and 22 peptides were represented by only their unmodified and modified (methionine oxidized) forms, respectively. For 47 peptides both the modified and unmodified sequences were observed.

The identified peptides originated from 196 proteins, of which only 89 were represented by at least two peptides (Table S2 on www.actabp.pl, proteins matching the same set of peptides were grouped into single clusters). A query of the Plasma Proteome Database (Muthusamy *et al.*, 2005) and PRIDE (Vizcaíno *et al.*, 2009) databases determined 195 of the proteins as blood constituents (Table S2 on www.actabp.pl). Of them, 104 (53.1%) were previously classified as substrates of serum proteases. Cleavage at the arginine and lysine residues of parent peptides at the split sites could be strongly attributed to trypsin-like activity of endoproteases (Fig. S1 on www.actabp.pl). While the specificity of trypsin-like protease cleavage observed in this study is similar to that observed by others (Villanueva *et al.*, 2006; Song *et al.*, 2006; van den Broek *et al.*, 2010b), our LC-ESI-MS−based analyses allowed for the detection of a substantial number of potential new protease targets. Most of them represented endoprotease-derived substrate peptides further cleaved to smaller fragments, which in turn form the serum peptidome, as illustrated by apolipoprotein A-IV (APOA4) (Fig. 2).

Nearly 40% of the identified peptides originated from five of the most prominently degraded proteins (Table S2 on www.actabp.pl): fibrinogen alpha chain (FGA), APOA4, complement C3 (C3), apolipoprotein A-I (APOA1) and alpha-1-antitrypsin (SERPINA1), while 63% peptides represented less than 10% (20 of 196) of the identified proteins. At the same time, only a small fraction of peptides stemmed from the most abundant blood proteins (Hortin, 2006), particularly albumin and immunoglobulins. These observations are in agreement with previous studies of the LMW blood degradome/proteome utilizing the MALDI/SELDI and ESI approach (Koomen *et al.*, 2005; Villanueva *et al.*, 2006; Ueda *et al.*, 2011) and indicate that the observed peptides originate from specific proteolysis.

A crosscheck with the data acquired using the LC-ESI-MS platform by Ueda *et al.* (Ueda *et al.*, 2011) revealed that peptides from our study tended to supplement each other in the protein coverage as opposed to overlapping. Together, the two studies report 1616 peptides, 154 of which (9.5%) are common between the two (Fig. 3A). On the other hand, the proportion of shared proteins is higher and equals 50 (20%) and 27 (27.7%) for all proteins (250) and two-peptide proteins (98), respectively (Fig. 3B, C). Thus, the bulk of peptide incompatibility may be attributed both to variations in sample handling procedures influencing the peptide ladder formation and to an inherent limitation of MS/MS, where peptide identification is selective.

Impact of clotting time on blood samples from healthy donors

Several pre-analytical factors may influence the level of peptide concentration variability in serum. Of them, most serum sample preparation procedures, including the time and temperature of blood clotting, as well as
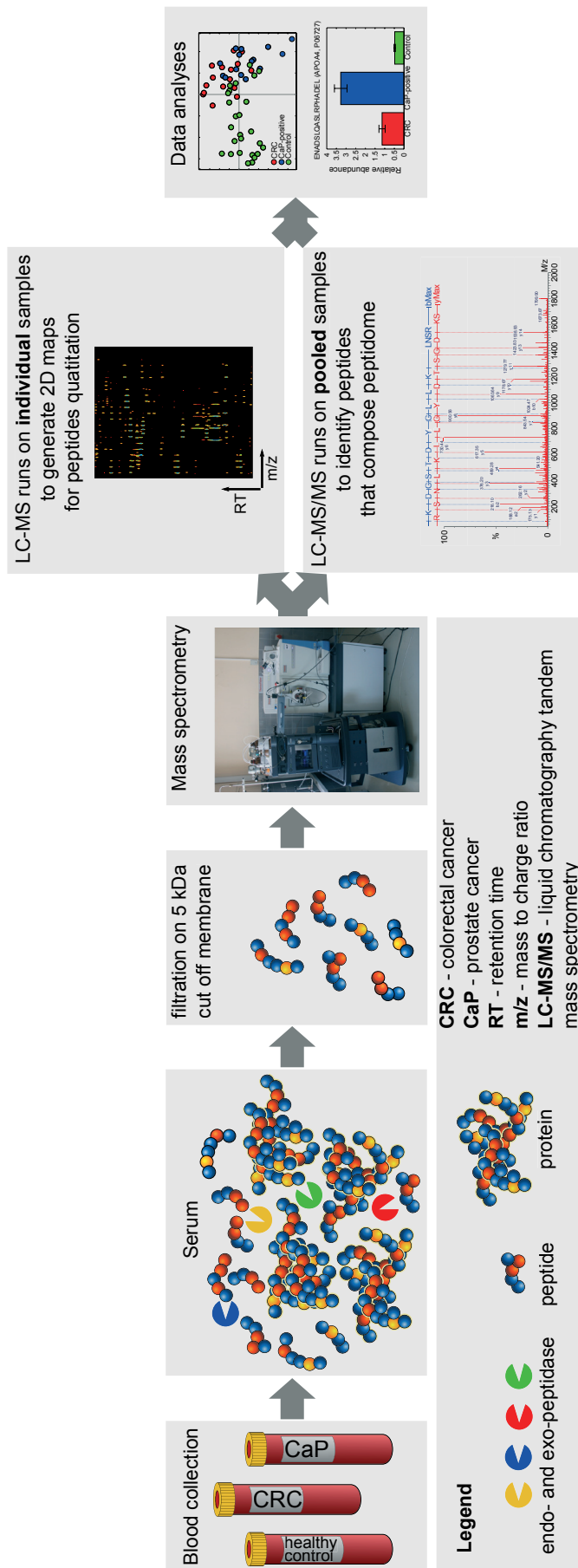
Figure 1. Schematic diagram of the experimental procedures used in profiling serum peptidomes from cancer patients.

blood collection tube type and storage conditions, affect the results of degradome/peptidome analyses (Van den Broek *et al.*, 2010a). From a practical point of view, the clotting time appears to be the most important variable that can be controlled in everyday situation. Therefore, herein we gave more attention to this factor. To this end, based on our previous work (Bakun *et al.*, 2009), we prepared a list of peptides that differentiated 1h and 2 h clotting in blood samples from 12 healthy individuals. Consequently, we used this list to filter out clotting time sensitive peptides from any other pair wise group comparisons.

Next, a more detailed investigation into the effect of clotting time was carried out by introducing known amounts of eight isotopically labeled synthetic peptides of fibrinopeptide A (FPA) to the samples drawn from 12 healthy individuals. This approach allowed us to measure the absolute concentration of each natural counterpart in serum (for details refer to: Bakun *et al.*, 2009).

The data acquired for 24 sera samples obtained from 12 healthy individuals showed significant differences between 1 h and 2 h clotting time for six out of eight peptides that comprised the FPA ladder (Table S3 on www. actabp.pl). While the levels of the four longest peptides decreased about twofold, the amounts of the two shortest peptides increased two-fold. Further analytical runs of the same 24 aliquots (technical repeat of MS run) as well as the runs performed on freshly prepared samples (technical repeat of peptides isolation from sera followed by MS run) from the same 24 sera showed consistent results. Moreover, a support vector machine (SVM) classifier (Cortes & Vapnik, 1995) built on the basis of the first measurement of labeled peptides allowed us to correctly identify samples with repeated measurements with an accuracy of 100% and 98% for re-measured aliquots and freshly prepared samples, respectively. The absolute amounts of FPA peptides were measured across all the samples in this study and their levels were compared to control samples clotted for 1 and 2 h. None of the mean levels of these peptides could differentiate the 1 h control from the CaP-positive, CaP-negative and control groups. Only the level of peptide DSGEGDFLAEGGGVR was significantly different for the CRC group as compared to 1 h; however, the observed direction of change was opposite to that which would result
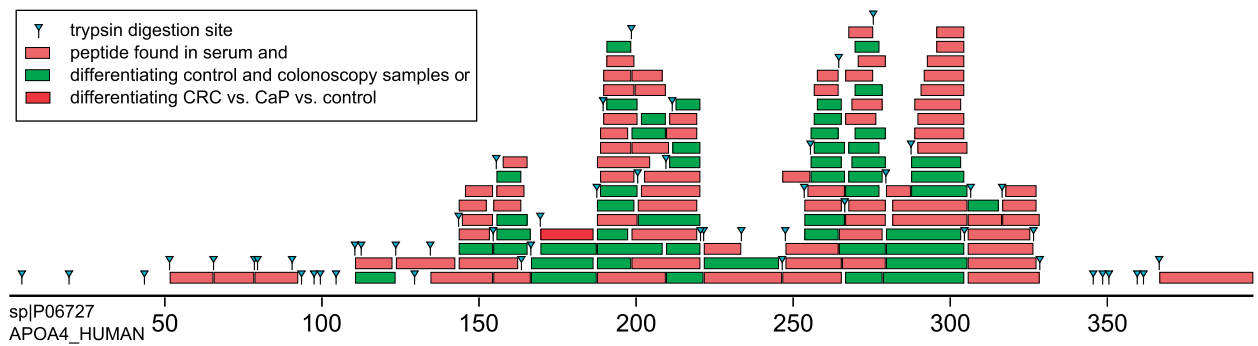
Figure 2. APOA4 peptide ladders identified in serum blood peptidome and its trypsin-like cleavage sites.
Colored boxes and blue flags represent the serum peptidome peptides identified in this work and trypsin-like digestion sites, respectively. CRC, colorectal cancer; CaP, prostate cancer.
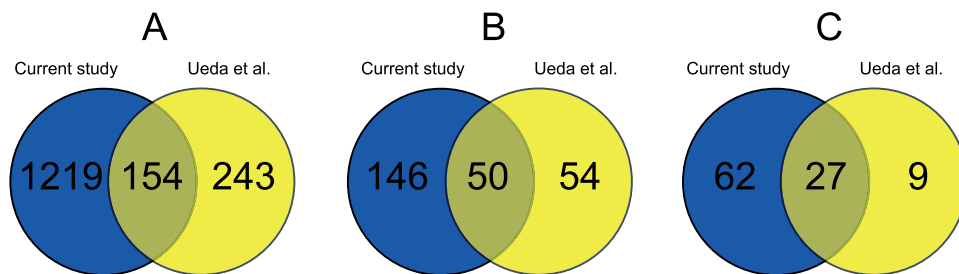

Figure 3. Comparison of peptide (A), protein (B) and two-peptide protein (C) numbers and identities with those in the study by Ueda *et al.* (Ueda *et al.*, 2011).

from the effect of the prolonged clotting time (Table S4 on www.actabp.pl).

Therefore, we could conclude that monitoring the abundance of six peptides that comprise the FPA ladder seems to be a sensitive indicator of clotting time, at least in healthy individuals.

Effect of overnight fasting

To determine the influence of overnight fasting on peptide levels, blood samples were collected from 15 healthy individuals enrolled from laboratory staff after 12 h overnight fasting and 2 h after a breakfast meal. Since no differentially abundant peptides were identified between paired serum samples (Fig. S2 on www.actabp.pl), all prospective patients in the CRC and CaP groups and the 26 control colonoscopy screening patients were not controlled for overnight fasting.

Selection of control group

Our previous work revealed that, the fold changes of the peptides that differentiated cancer from non-cancer samples were relatively small and did not exceed a three-fold decrease or increase (Bakun *et al.*, 2009). This indicates the need for the selection of a proper control group for the identification of cancer markers by the MS signal quantitation methods, especially because of the limited specificity of degradome/peptidome-based assays, leading to both false-positive and false-negative results. Technical difficulties with sample preparation reproducibility, signal processing artifacts, and statistical data analysis methods have been highlighted (Van den Broek *et al.*, 2010a).

Although not assessed, peptidome composition could be influenced by bowel clean-up prior to colonoscopy. This issue seems to be important if realized that due to

the facilitated access to blood samples of patients enrolled in screening colonoscopy programs, studies of the LMW degradome often utilize such samples as controls (Van den Broek *et al.*, 2010b). Our set of control blood samples initially consisted of specimens obtained from 30 healthy individuals immediately prior to colonoscopy and 26 samples drawn from individuals who underwent colonoscopic examination at least four weeks earlier. Unsupervised PCA evaluation effectively separated these control groups (Fig. 4) and a pair-wise comparison revealed a set of 67 peptides exhibiting highly significant differences between the two groups of healthy controls (a threshold of adjusted p-value ≤ 0.01 and FC ≥ 2; see
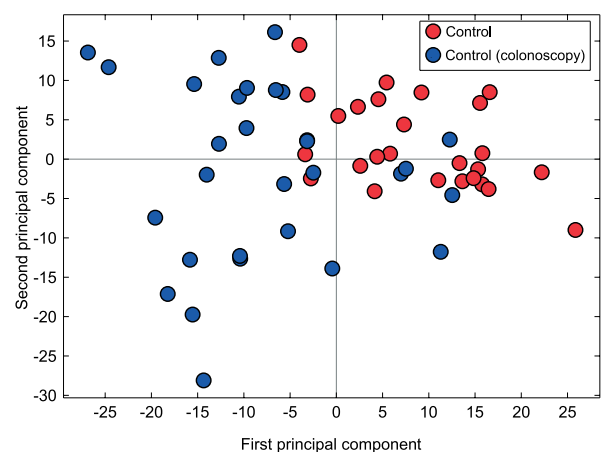

Figure 4. Control samples obtained from 30 and 26 healthy individuals immediately prior to colonoscopy and those who underwent colonoscopic examination at least four weeks earlier, respectively, visualized on the plane of the two first principal components calculated from all the peptides in the data set.
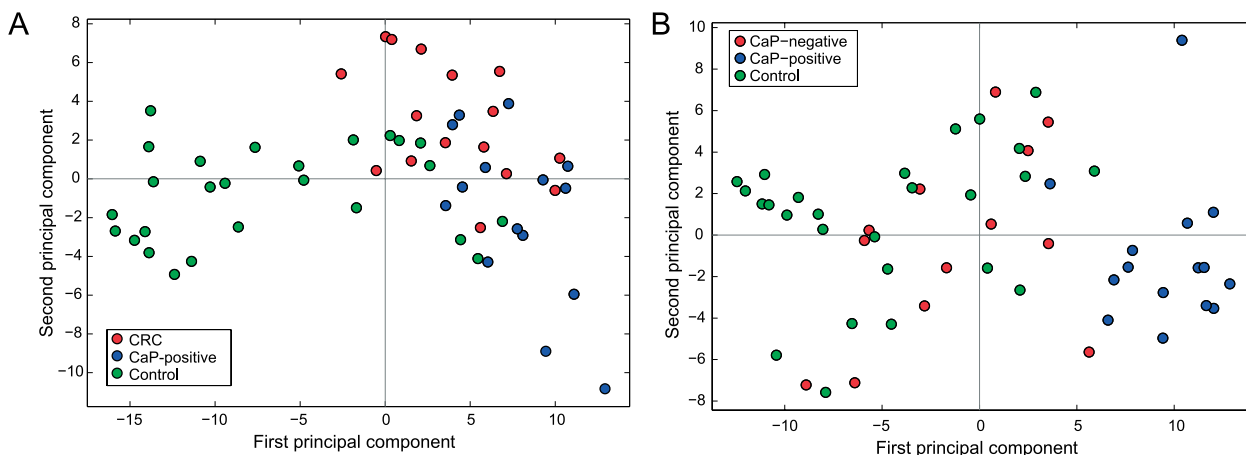
Figure 5. Distribution of samples on the plane spanned by the first two principal components of a set of peptides showing statistical significance (FDR ≤ 0.01; FC ≥ 2) in at least one of the three pair-wise comparisons
(**A**) CaP *vs.* Control, CRC *vs.* Control, and CaP *vs.* CRC. (B) Principal components analysis based on a set of 118 peptides showing statistically significant abundance changes (FDR ≤ 0.01; FC ≥ 2) in a pair-wise comparison of CaP *vs.* Control.

Statistical analysis of quantitative MS results Section in Materials and Methods for details). Of them, 49 peptides originated from APOA4 (Table S5 on www.actabp.pl and Fig. 2). Concentrations of 61 peptides, including all representing APOA4, which differentiated healthy individuals from these two control groups, were lower in serum samples obtained prior to colonoscopy. This result clearly indicated that the preparation for colonoscopy influences serum peptidome composition in healthy patients. Therefore, for a LMW peptidome quantitative healthy - cancer comparisons we chose 26 controls that had colonoscopic examination at least four weeks before blood collection.

Quantitative comparison of CRC–CaP–Control

For the quantitative comparisons, we selected 16, 15 and 26 samples from patients with colorectal (CRC) and prostate cancer-positive (CaP-positive) and from controls, respectively.

Using a threshold of adjusted *p*-value ≤ 0.01 and FC ≥ 2, a set of 169 peptides exhibiting significant differences in pair-wise comparisons of control *vs.* CRC and control *vs.* CaP was identified (Table S6 on www.actabp.pl). Among them, 118 and 88 peptides could differentiate CaP and CRC from the control, respectively. Interestingly, 37 peptides were common between the two pair-wise comparisons and exhibited the same directionality of change. It is worth mentioning that the actual similarity of these comparisons may be greater than would appear from the specified number of common peptides. Although in most cases the differences in both comparisons showed the same direction of change, they did not reach the established criteria of statistical significance. The linear correlation coefficient between the log-ratios of the mean values of groups in the pair-wise comparison of control vs. CRC and control vs. CaP for these 169 peptides was 0.809 (Fig. S3 on www.actabp.pl). Further PCA graphical overview showed that cancer samples tend to group together (Fig. 5A), suggesting the presence of certain peptide profiles inherent to cancer samples.

There were 30 peptides differentiating CRC from CaP (Table S7 on www.actabp.pl) but only two of them discriminated cancer types from control patients. Under the applied threshold of significance, a peptide ENAD-SLQASLRPHADEL derived from APOA4 (Figs. 2 and 6) and a peptide SSIKEKTVGR from inter-alpha-trypsin inhibitor heavy chain H2 (ITIH2) could be regarded as markers of specific types of cancer (Fig. 6). Moreover, the abundance of both peptides was not changed in a comparison of control group samples and was not sensitive to clotting time.

The comparison of 13 CaP-negative samples with control samples revealed no peptides with statistically significant differences between these two groups. The comparison of CaP-negative with CaP-positive patients' samples showed 36 differentiating peptides (Table S8 on www.actabp.pl); among them, 34 were common with the set of 118 peptides identified in the comparison of CaP-positive samples and the control. The small number of peptides is probably due to the lower number of CaP-negative samples as compared to the control group: 13 *vs.* 26, respectively. Consistent with this assumption was the observation that the values of peptide FCs within the set of 118 peptides identified in the CaP vs. control comparison, when compared with those identified in the CaP-negative *vs.* control comparison, showed concordant directionality of change for 115 out of 118 peptides, with a coefficient of correlation of 0.969 (Fig. S4 on www.actabp.pl). An additional confirmation of the relative lack of difference between CaP-negative and control groups was their relative positions on the PCA plot, where all but one of the CaP-negative samples clustered with the control samples (Fig. 5B). This could indicate that the similarities shown in Fig. 5A for CRC and CaP samples are due to intrinsic cancer features, in this case differences in peptide abundance, which are not observed or are less pronounced when no cancerous tissue is present.

DISCUSSION

Blood peptide composition may vary between normal and diseased phenotypes as a result of the activity of more than 500 proteases responsible for protein degradation (Shen *et al.*, 2010). Some peptides are generated *in vivo* while others are generated *ex vivo*, after blood collection. Our studies identified 1373 unique peptides; different sets of peptides differentiated CaP and/or CRC

from controls (Fig. 5A), while no peptide set exhibited differences between progression-free CaP patients and control samples in pair-wise comparisons (Fig. 5B).

When our peptide markers were compared to peptides reported by already published MALDI/SELDI (Ebert *et al.*, 2005; Koomen *et al.*, 2005; Fung *et al.*, 2005; Cheng *et al.*, 2005; Villanueva *et al.*, 2006; Song *et al.*, 2006; Shi *et al.*, 2006) and ESI (Bakun *et al.*, 2009; Ueda *et al.*, 2011) studies, for 59 of 119 literature-extracted peptides, we were able to obtain reliable quantitative values in our dataset (Table S9 on www.actabp.pl). In this comparison, we considered only those markers for which identified MS/MS sequences and evident directional abundance changes were found, and which were recognized by the authors of these studies as relevant. All of them were derived from a limited set of proteins, often from a particular part of a protein. Moreover, these proteins are abundant in LMW serum; therefore, the peptides obtained from them are inherently more convenient for MS identification.

In view of numerous reports, the existence of cancer-related protease cleavage of serum proteins seems to be undisputable. However, the results of peptidome analyses are prone to variations in sample collection, handling procedures and limitations of MS measurements (Koomen *et al.*, 2005; van den Broek *et al.*, 2010a). As we highlighted in a comparison with Ueda et al. data (Ueda *et al.*, 2011) (Fig. 3) both MS/MS inherent analytical features and differences in sample handling might be a cause of bias in peptides identification. This fact alone makes it difficult to obtain reliable collection of peptides that should be further examined as potential biomarkers. Furthermore, the influence of pre-analytical factors can also be noticed in quantitative analyses of MS data. For example, in most of the previously published studies on the peptidome, including ours (Ebert *et al.*, 2006; Bakun *et al.*, 2009), the peptides composing FPA could differentiate cancer from healthy samples. The present study has established, using sera from healthy individuals, that FPA peptide levels are mostly sensitive to clotting time. However, since blood clotting could differ under inflammatory conditions, such as inflammatory bowel disease (Danese *et al.*, 2007) or liver dysfunctions (Amitrano *et al.*, 2002), it should be stressed that it may be a potential source of pre-analytical bias during the analysis of serum peptides. Additionally, the degradome profiling discriminating cancer patients from healthy controls can also

reflect disturbances in coagulation affected by neoplastic processes. Recurrent thromboembolic events in cancer patients result from the expression of procoagulant proteins that lead to platelet activation and fibrin clot formation (Ahmad & Ansari, 2011), and altered blood coagulation is commonly seen in cancer patients both by clinical observation and as a result of blood coagulation tests. However, since altered levels of FPA peptides were also reported as markers of diabetic nephropathy (Gianazza *et al.*, 2010) and coronary artery disease (Zito *et al.*, 2000), other variables may also account for observed differences among FPA peptide levels.

In search for potential peptide biomarkers, the FPA peptides example underscores the need to minimize pre-analytical variation and proper selection of control groups. To address the first point, we focused on clotting time as it is a parameter that could be strictly controlled under laboratory conditions, however, it may be difficult to control in everyday clinical conditions. Therefore, future serum peptidome biomarker research should include scrutinization of clotting time sensitive peptides since their diagnostic usefulness seems to be very limited. With regard to the second issue we show that a standard bowel preparation prior to colonoscopy examination affects the serum peptide profile (Fig. 4), significantly decreasing the levels of 67 peptides, including 49 peptides originated from APOA4. This information should be taken into consideration in further studies when blood samples collected during screening colonoscopy programs will be utilized as control samples (Van den Broek *et al.*, 2010b).

It is also worth emphasizing the potential role of peptides originated from APOA4. In humans, APOA4 is secreted by enterocytes; however, its mRNA expression has also been demonstrated in human liver (Nilsson *et al.*, 2007) and in the rat hypothalamus (Shen *et al.*, 2008). A number of functions for APOA4 have been proposed, including a pivotal role in food intake and body weight regulation (Stan *et al.*, 2003). Interestingly, as shown before in humans, 24 h fasting decreases APOA4 concentration in plasma by ~50%, while typical 10–12 h overnight fasting before blood drawing did not significantly change it (Rafat *et al.*, 2004). This observation is in concordance with the abundance of APOA4 peptides measured in the sera of patients prior to colonoscopy (Fig. 4) and the outcome of the experiment where serum was collected from the same individuals following overnight fasting and 2 h after
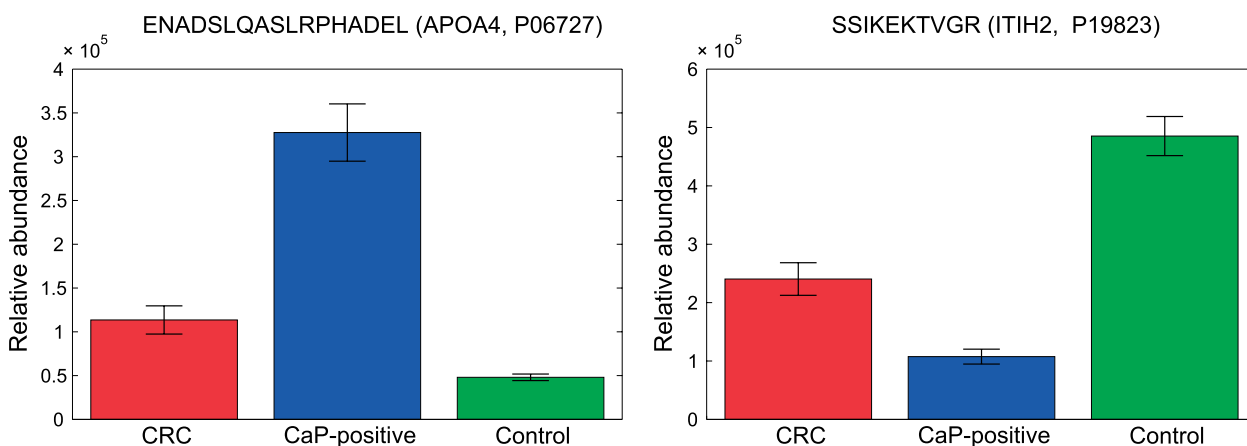

Figure 6. Levels of peptides differentiating simultaneously in the pair-wise comparisons of colorectal cancer (CRC)–control, prostate cancer (CaP)-positive–control, and CRC–CaP-positive.
Bars represent mean values and whiskers are for standard error of mean values.

the first meal (Fig. S2 B on www.actabp.pl). While both cancer-associated endo- and exo-peptidases are believed to be responsible for the cleavage of APOA4 (Ueda *et al.*, 2011), physiological functions of endogenously-processed APOA4-derived peptides are not known. Interestingly, Bertile et al. hypothesized that lower levels of blood APOA4 protein observed during prolonged fasting states could contribute to triggering of feeding behavior (Bertile *et al.*, 2009). Accordingly, it is tempting to speculate the involvement of APOA4-derived peptides in stimulating such physiological responses.

The results on CRC and CaP peptidome presented herein suggest that the vast majority of discriminating peptides are not specific to a particular cancer type, rather their abundance results from general state of health of cancer patients. This observation is supported by the concordance of differentiating peptides lists in CRC and CaP *vs.* control sera comparisons (Table S6 on www.actabp.pl) and a high correlation of FC values in both comparisons (Fig. S3 on www.actabp.pl). Future work is needed to include other cancer types to determine whether or not changes in serum peptides reflect a specific tumor type. It is noteworthy that in our study only two APOA4- and ITIH2-derived peptides out of the initial list of 169 peptides appeared to be differentiating simultaneously in the pair-wise comparisons of CRC–Control, CaP–Control and CRC–CaP and, therefore, could be regarded as markers of specific types of cancer (Fig. 6). The diagnostic value of these peptides requires verification using alternative methods and including higher number of samples. Nevertheless, it is worth emphasizing that the analysis of serum peptidome, bearing in mind pre-analytical factors as well as validation peptide specificity, should be useful to focus on the most promising biomarkers. Despite improvements biomarker validation remains a challenging time- and resource-consuming task (Whiteaker *et al.*, 2011). Thus, careful consideration is needed to reduce the number of false identification at the exploratory stage of research using LC-ESI-MS.

CONCLUSIONS

Our results show that bowel clean up may alter blood peptidome composition. The significance of this information is underlined by the fact that control/healthy blood samples are often collected within the frames of screening colonoscopy programs. Furthermore, we propose a method to control clotting time variable through monitoring FPA peptides as well as eliminating the clotting sensitive peptides. FPA peptides, under strictly controlled clotting conditions undergo extensive and consistent changes. Although substantial efforts being invested since the initial publication on LMW proteome a decade ago (Petricoin *et al.*, 2002), none of the peptides differing in abundance between healthy control and cancer samples appeared to be diagnostically applicable biomarkers for clinical practices. Since there is no consensus on appropriate methods for blood handling and processing, and experimental protocols described in many studies are largely speculative, the resultant variations in degradome profiles are the biggest obstacle to their practical application.

Acknowledgements

REFERENCES

Ahmad S, Ansari AA (2011) Therapeutic roles of heparin anticoagulants in cancer and related disorders. *Med Chem* **7**: 504–517.

Amitrano L, Guardascione MA, Brancaccio V, Balzano A (2002) Coagulation Disorders in Liver Disease. *Seminars in Liver Disease* **22**: 083–096.

Bakun M, Karczmarski J, Poznanski J, Rubel T, Rozga M, Malinowska A, Sands D, Hennig E, Oledzki J, Ostrowski J, Dadlez M (2009) An integrated LC-ESI-MS platform for quantitation of serum peptide ladders. Application for colon carcinoma study. *Proteomics Clin Appl* **3**: 932–946.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Statistic Soc Series B (Methodological)* **57**: 289–300.

Bertile F, Schaeffer C, Le Maho Y, Raclot T & Van Dorsselaer A (2009) A proteomic approach to identify differentially expressed plasma proteins between the fed and prolonged fasted states. *Proteomics* **9**: 148–158.

Van den Broek I, Sparidans RW, Schellens JHM, Beijnen JH (2010a) Specific investigation of sample handling effects on protease activities and absolute serum concentrations of various putative peptidome biomarkers. *Clin Proteomics* **6**: 115–127.

Van den Broek I, Sparidans RW, Van Winden AWJ, Gast M-CW, Van Dulken EJ, Schellens JHM, Beijnen JH (2010b) The absolute quantification of eight inter-α-trypsin inhibitor heavy chain 4 (ITIH4)-derived peptides in serum from breast cancer patients. *Proteomics Clin Appl* **4**: 931–939.

Cheng A-J, Chen L-C, Chien K-Y, Chen Y-J, Chang JT-C, Wang H-M, Liao C-T, Chen I-H (2005) Oral cancer plasma tumor marker identified with bead-based affinity-fractionated proteomic technology. *Clin Chem* **51**: 2236–2244.

Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* **20**: 273–297.

Danese S, Papa A, Saibeni S, Repici A, Malesci A, Vecchi M (2007) Inflammation and coagulation in inflammatory bowel disease: The clot thickens. *Am J Gastroenterol* **102**: 174–186.

Ebert MPA, Lamer S, Meuer J, Malfertheiner P, Reymond M, Buschmann T, Röcken C, Seibert V (2005) Identification of the thrombin light chain a as the single best mass for differentiation of gastric cancer patients from individuals with dyspepsia by proteome analysis. *J Proteome Res* **4**: 586–590.

Ebert MPA, Niemeyer D, Deininger SO, Wex T, Knippig C, Hoffmann J, Sauer J, Albrecht W, Malfertheiner P, Röcken C (2006) Identification and confirmation of increased fibrinopeptide a serum protein levels in gastric cancer sera by magnet bead assisted MALDI-TOF mass spectrometry. *J Proteome Res* **5**: 2152–2158.

Fung ET, Yip T-T, Lomas L, Wang Z, Yip C, Meng X-Y, Lin S, Zhang F, Zhang Z, Chan DW, Weinberger SR (2005) Classification of cancer types by measuring variants of host response proteins using SELDI serum assays. *Int J Cancer* **115**: 783–789.

Gianazza E, Mainini V, Castoldi G, Chinello C, Zerbini G, Bianchi C, Galbusera C, Stella A, Mauri G, Zoppis I, Magni F, Kienle MG (2010) Different expression of fibrinopeptide A and related fragments in serum of type 1 diabetic patients with nephropathy. *J Proteomics* **73**: 593–601.

Hortin GL (2006) The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin Chem* **52**: 1223–1237.

Käll L, Storey JD, MacCoss MJ, Noble WS (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res* **7**: 29–34.

Koomen JM, Li D, Xiao L, Liu TC, Coombes KR, Abbruzzese J, Kobayashi R (2005) Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. *J Proteome Res* **4**: 972–981.

Mikula M, Gaj P, Dzwonek K, Rubel T, Karczmarski J, Paziewska A, Dzwonek A, Bragoszewski P, Dadlez M, Ostrowski J (2010) Comprehensive analysis of the palindromic motif TCTCGCGAGA: a regulatory element of the HNRNPK promoter. *DNA Res* **17**: 245–260.

Muthusamy B *et al.* (2005) Plasma Proteome Database as a resource for proteomics research. *Proteomics* **5**: 3531–3536.

Nilsson L-M, Abrahamsson A, Sahlin S, Gustafsson U, Angelin B, Parini P, Einarsson C (2007) Bile acids and lipoprotein metabolism: effects of cholestyramine and chenodeoxycholic acid on human hepatic mRNA expression. *Biochem Biophys Res Commun* **357**: 707–711.

Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA

(2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**: 572–577.

Rafat M, Sattler AM, Hackler R, Soufi M, Steinmetz A, Maisch B, Schaefer JR (2004) Apolipoprotein A-IV in the Fed and Fasting States. *Clin Chem* **50**: 1270–1271.

Shen L, Pearson KJ, Xiong Y, Lo C-M, Tso P, Woods SC, Davidson WS, Liu M (2008) Characterization of apolipoprotein A-IV in brain areas involved in energy homeostasis. *Physiol Behav* **95**: 161–167.

Shen Y, Liu T, Tolić N, Petritis BO, Zhao R, Moore RJ, Purvine SO, Camp DG, Smith RD (2010) Strategy for degradomic-peptidomic analysis of human blood plasma. *J Proteome Res* **9**: 2339–2346.

Shi Q, Harris LN, Lu X, Li X, Hwang J, Gentleman R, Iglehart JD, Miron A (2006) Declining plasma fibrinogen alpha fragment identifies HER2-positive breast cancer patients and reverts to normal levels after surgery. *J Proteome Res* **5**: 2947–2955.

Song J, Patel M, Rosenzweig CN, Chan-Li Y, Sokoll LJ, Fung ET, Choi-Miura N-H, Goggins M, Chan DW, Zhang Z (2006) Quantification of fragments of human serum inter-alpha-trypsin inhibitor heavy chain 4 by a surface-enhanced laser desorption/ionization-based immunoassay. *Clin Chem* **52**: 1045–1053.

Stan S, Delvin E, Lambert M, Seidman E, Levy E (2003) Apo A-IV: an update on regulation and physiologic functions. *Biochim Biophys Acta* **1631**: 177–187.

Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**: 520–525.

Ueda K, Saichi N, Takami S, Kang D, Toyama A, Daigo Y, Ishikawa N, Kohno N, Tamura K, Shuin T, Nakayama M, Sato T-A, Nakamura Y, Nakagawa H (2011) A comprehensive peptidome profiling technology for the identification of early detection biomarkers for lung adenocarcinoma. *PLoS ONE* **6**: e18567.

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, Tempst P (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* **116**: 271–284.

Vizcaíno JA, Côté R, Reisinger F, Foster JM, Mueller M, Rameseder J, Hermjakob H, Martens L (2009) A guide to the Proteomics Identifications Database proteomics data repository. *Proteomics* **9**: 4276–4283.

Whiteaker JR *et al.* (2011) A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotechnol* **29**: 625–634.

Zito F, Drummond F, Bujac SR, Esnouf MP, Morrissey JH, Humphries SE, Miller GJ (2000) Epidemiological and genetic associations of activated factor XII concentration with factor VII activity, fibrinopeptide A concentration, and risk of coronary heart disease in men. *Circulation* **102**: 2058–2062.