DOI: 10.15386/cjmed-880



# PLASMA PROTEIN ABSOLUTE QUANTIFICATION BY NANO-LC Q-TOF UDMS$^E$ FOR CLINICAL BIOMARKER VERIFICATION

**MARIA ILIES[1,2], CRISTINA ADELA IUGA[1,3], FELICIA LOGHIN[4], VISHNU MUKUND DHOPLE[2], ELKE HAMMER[2,5]**

**[1]Department of Pharmaceutical Analysis, Faculty of Pharmacy, Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania**
**[2]Department of Functional Genomics, Interfaculty Institute of Genetics and Functional Genomics, University Medicine Greifswald, Germany**
**[3]Department of Proteomics and Metabolomics, MedFuture Research Center for Advanced Medicine, Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania**
**[4]Department of Toxicology, Faculty of Pharmacy Iuliu Haţieganu University of Medicine and PharmacyCluj-Napoca, Romania**
**[5]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany**

***Abstract***

***Background and aims.*** *Proteome-based biomarker studies are targeting proteins that could serve as diagnostic, prognosis, and prediction molecules. In the clinical routine, immunoassays are currently used for the absolute quantification of such biomarkers, with the major limitation that only one molecule can be targeted per assay. The aim of our study was to test a mass spectrometry based absolute quantification method for the verification of plasma protein sets which might serve as reliable biomarker panels for the clinical practice.*

***Methods.*** *Six EDTA plasma samples were analyzed after tryptic digestion using a high throughput data independent acquisition nano-LC Q-TOF UDMS$^E$ proteomics approach. Synthetic Escherichia coli standard peptides were spiked in each sample for the absolute quantification. Data analysis was performed using ProgenesisQI v2.0 software (Waters Corporation).*

***Results.*** *Our method ensured absolute quantification of 242 non redundant plasma proteins in a single run analysis. The dynamic range covered was 105. 86% were represented by classical plasma proteins. The overall median coefficient of variation was 0.36, while a set of 63 proteins was found to be highly stable. Absolute protein concentrations strongly correlated with values reviewed in the literature.*

***Conclusions.*** *Nano-LC Q-TOF UDMS$^E$ proteomic analysis can be used for a simple and rapid determination of absolute amounts of plasma proteins. A large number of plasma proteins could be analyzed, while a wide dynamic range was covered with low coefficient of variation at protein level. The method proved to be a reliable tool for the quantification of protein panel for biomarker verification in the clinical practice.*

**Keywords:** plasma, proteomics, absolute quantification, nano-LC Q-TOF UDMS$^E$

Manuscript received: 25.07.2017
Accepted: 31.08.2017
Address for correspondence: ilies.maria@umfcluj.ro

## Background and aims

Proteomic based biomarker studies are targeting proteins that could serve as screening, diagnostic, staging, prognosis, prediction and monitoring molecules in the clinical practice. Several factors have contributed to the expansion of biomarker studies worldwide. First of all, there is an ever increasing number of biobanks offering high quality biospecimens for investigations [1]. The sample type most commonly used for protein biomarker research is blood. Blood samples, such as plasma, which are collected in a non-invasive manner by using well established and low cost techniques, are considered "ideal fluids" for research. Moreover, plasma encloses proteins from the whole body and a single sample offers a wide range of information, which has also recently been reviewed with regard to cancer biomarkers [2]. Secondly, mass spectrometry (MS) has emerged as a high resolution technique offering a wide range of applications in proteomic research. A single run analysis allows for protein profiling, targeted analysis, as well as quantitative measurement of a high number of proteins. Recently, Sabbagh et al. [3] reviewed MS based protein quantification methods and showed that multiple reaction monitoring (MRM) and data-independent acquisition (DIA) appear to be the most suitable techniques for the clinical practice. Nevertheless, determining the absolute concentration of plasma proteins and implementing such a method for the clinical practice remain a challenge.

Currently, immunoassays are employed for biomarker verification as well as quantitative measurements of proteins in the clinical practice. Although the enzyme-linked immunosorbent assay (ELISA) is currently considered the "gold standard" in the clinical laboratory practice, there are several limitations of the ELISA over MS-based approaches and their broader applicability in the clinical practice. ELISA based absolute quantification of biomarkers is affected by the method of blood sample collection [4]. Using MS based methods, proteins can be detected based on their unique mass to charge ratio precursor peptides, whereas no antibody interference or cross reactivity can occur. ELISA methods offer accurate quantitative determination for a single protein, whereas MS based methods offer quantification of hundreds of proteins in a single run [5]. Moreover, the syntheses of antibodies for every protein, as well as the entire assay production itself, are time consuming, laborious, and expensive. The introduction of ELISA based assays in the laboratory practice has therefore been delayed.

Hence, the aim of our study was to test a simple and rapid absolute quantification method for the verification of plasma protein sets which might be used as reliable biomarker panels in the clinical practice. The method uses synthetic *Escherichia coli (E. coli)* standard peptides for the absolute quantification as described earlier by Silva et al. [6] and was adapted for the analyses of blood plasma using a high throughput data independent acquisition nano-LC Q-TOF UDMS$^E$ proteomics approach.

## Methods

### *Study design*

Six healthy young volunteers (three males, three females, aged 24 to 29) were enrolled in this study. At the blood donation facility of the University Medicine of Greifswald, Germany, one venous blood sample was drawn from each subject in BD Vacutainer® tubes (Becton Dickenson, Heidelberg, Germany) containing ethylenediaminetetraacetic acid (EDTA) using standard venipuncture technique. EDTA tube characteristics and plasma preparation recommendations can be found in Ref. [7]. The plasma was obtained following the tube manufacturer's protocol, and aliquots were stored at -80 °C until processing. The study was conducted with respect to the WMA Declaration of Helsinki. Prior to the sample collection, all study participants signed a document of informed consent. The study was approved by the Ethics Committee of the University Medicine Greifswald, Germany.

### *Protein tryptic digestion*

Protein digestion was done according to Ilies et al. [8]. Bradford Assay (BioRad Laboratories, Munich, Germany) was used for the protein concentration determination [9]. Reduction (2.5 mM dithiothreitol, 1 h, 60 °C) and alkylation (10 mM iodoacetamide, 15 min at 37 °C) was employed for each of four µg of sample. Trypsin (Promega, Madison, WI, USA) was used for the enzymatic digestion (1: 25 protease: protein ratio, 16 h, 37 °C). The digestion reaction was quenched with 1% acetic acid, samples were desalted using ZipTip µC18 (Millipore Cooperation, Billerica, MA, USA), and the eluted peptides were lyophilized. Hi3 *E. coli* standard peptides (protein disaggregation chaperone ClpB, 186006012, Waters Corporation, Milford, MA, USA) were spiked in each sample (1.65 fmol/100 ng) for the absolute quantification. Samples of 0.1 µg/µL final concentration in 0.1% acetic acid: acetonitrile (98:2 v/v) were subjected to LC-MS/MS analysis.

### *Nano-LC Q-TOF UDMS$^E$ protein identification and absolute quantification*

Proteins were identified according to Ilies et al. [8]. 200 ng peptides were analyzed by reversed phase chromatography using an ACQUITY UPLC® M-Class HSS T3 column. A non-linear gradient of 7% to 60% acetonitrile in 0.1% acetic acid within 105 min at a flow rate of 400 nL/min was employed. For the detection of the eluted peptides, an on-line coupled travelling wave ion-mobility-enabled hybrid quadrupole orthogonal acceleration time-of-flight mass spectrometer (SYNAPT G2-Si HDMS, Waters Corporation) was used. Data acquisition mode was set to independent acquisition, whereas collision voltage ramping was set according to our previous study [8]. MassLynx™ Software Version 1.53.1398 (Waters Corporation) was used

for data acquisition.

LC–UDMS$^E$ raw data was processed with the Progenesis QI v2.0 (Waters Corporation) software, with automated peak picking and alignment of the ions. The built-in search engine of Progenesis was employed for the spectra search against a Uniprot/ Swissprot database (06/2016) limited to human entries (20151) and extended by the amino acid sequence of the three peptides with highest signal intensities of the spike-in standard (LPQVEGTGGDV QPSQDLVRNNPVLIGEPGVGKVTDAEIAEVLAR). Enzyme specificity was trypsin and a maximum of 1 missed cleavage was allowed. Carbamidomethylation of cysteine was set as fixed modification and oxidation of methionine as a variable modification. A false discovery rate of less than 4% was set as search tolerance parameters. Proteins were considered as significantly identified when the following ion matching requirements were fulfilled: fragments/peptide ≥2, fragments/protein ≥5 and peptides/ protein ≥1, confidence score of ≥5. Absolute protein quantification was performed on the spiked in amount of the *E. coli* standard peptide using the absolute quantitation feature of the Progenesis QI software, which is based on the method described by Silva et al. [6].

***Bioinformatics analysis***

Exported data were filtered for valid values per protein and all proteins for which the signal to noise ratio did not allow absolute quantification were excluded. The absolute amount was further calculated in µmol/L of plasma analyzed. Results on protein variation were presented as coefficient of variation (CV).

## Results

***Characterization of the blood proteins***

Using UDMS$^E$ a total number of 242 non-redundant plasma proteins were absolutely quantified in every one of the six individual samples analyzed. Each of these proteins was characterized by 18 unique peptides (median). The absolute concentration of every protein was calculated based on the 3 peptides with the highest signal intensities in comparison to the TOP3 intensities of a spike-in standard with known absolute amount. To ensure that the same 3 peptides of the 6 *E. coli* standard peptides were used throughout the runs, the runs were previously screened for the highest and most robust standard peptide intensities. Only the sequences of those 3 peptides instead of the whole *E. coli* ClpB protein sequence were added to the human database. Subsequently, absolute amounts per loaded sample as well as plasma concentrations were calculated for each plasma protein.

Based on the average concentration, 86% were represented by classical proteins, known also as secreted proteins, whereas a 7% fraction was represented by leakage proteins (Figure 1). The plasma protein concentration dynamic range covered was $10^5$ with a maximum average concentration of 839.79 µmol/L (serum albumin) and a minimum of 0.01 µmol/L (serum paraoxonase/lactonase 3).

We further investigated protein variation among the quantified protein set. The median coefficient of variation was CV= 0.36. Among these, 63 proteins showed a CV≤ 0.25 and were subsequently considered as the most stable protein set. The plasma concentration distribution of the most stable protein set in comparison to the complete set is illustrated in Figure 2.

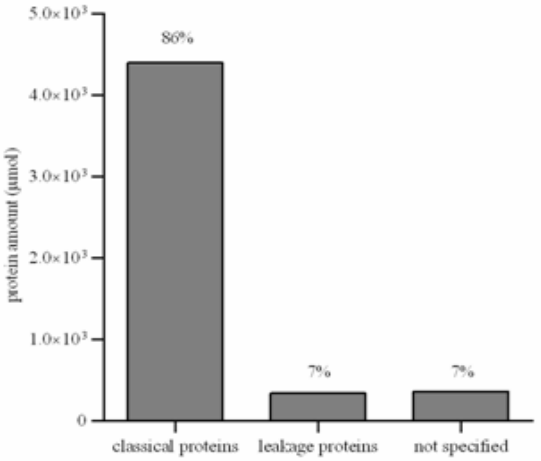


**Figure 1.** Classical and leakage protein distribution.

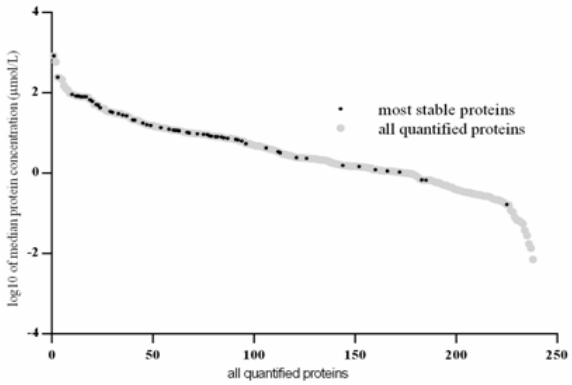


**Figure 2.** Plasma concentration distribution of the most stable protein set.

***Correlation with reviewed absolute values***

We compared the absolute concentration correlation for the proteins quantified with the absolute amount of proteins reported in the literature. A set of highly and medium abundant proteins were reviewed by Hortin et al. [10], with 94 proteins being common to our protein set. Figure 3 A shows a scatter plot for the corresponding Pearson's correlation with r=0.7664. We also compared our results to the protein panel quantified by Percy et al. [11] in a multiple reaction monitoring approach (MRM). 88 proteins of this panel were also identified in our profiling study. Figure 3 B shows the corresponding Pearson's correlation with r=0.6710.

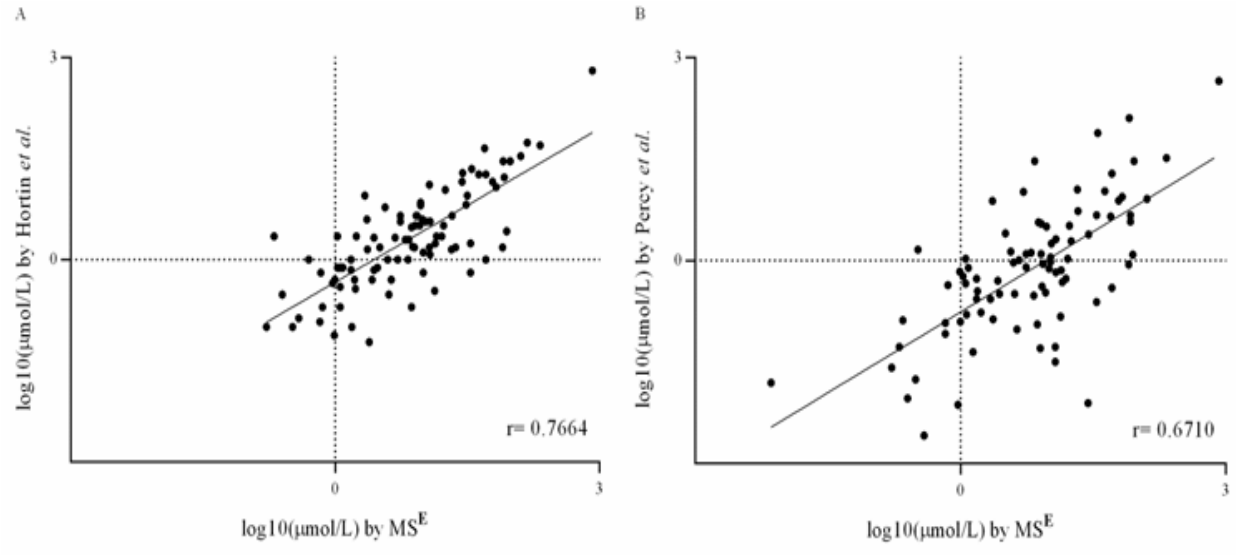


**Figure 3.** Correlation of plasma concentration with reviewed absolute values.

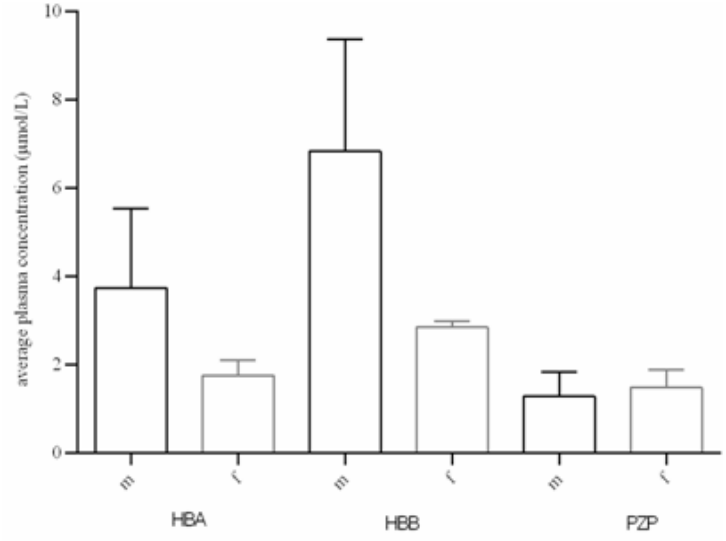


**Figure 4.** Sex specific plasma protein concentration as a proof of concept.

As a further proof of concept, plasma concentrations of hemoglobin subunits alpha (HBA) and beta (HBB) were analyzed in relation to their sex specific plasma concentration. HBA and HBB were found in a higher concentration among the male (m) plasma samples, whereas the pregnancy zone protein (PZP) median plasma concentration was higher in the female (f) samples (Figure 4). Our findings were thus in accordance with published results.

## Discussion

While protein profiling using mass spectrometry high throughput methods, such as data dependent or data independent acquisition, are well established techniques in biomarker discovery studies, the determination of absolute concentrations in large protein sets remains a challenge in the proteomics field [3]. To our knowledge, quantitation methods of up to 200 high and medium abundant plasma proteins in a single run are only established on MRM basis, and currently limited to few specialized laboratories [11,12]. Instead, ELISA based methods are routinely used, though offering only single protein quantification per assay. The aim of the present study was to test the applicability of a MSE method for the absolute quantification of proteins in plasma previously described by Silva et al. [6], whereas some improvements were made. Through a state-of-the-art nano-LC Q-TOF UDMS$^E$ proteomics approach, we first quantified a total number of 242 non-redundant plasma

proteins in a single run analysis. Later, we explored the protein set in more detail and showed a high correlation with absolute values reviewed in the literature.

Silva et al. [6] have established an absolute quantification method for proteins by using a single standard spike in for a single point calibration and demonstrated to be suitable for LC-MSE applications on complex protein samples [6]. However, there are some differences worth noting when comparing the method of Silva et al. [6] and ours. First of all, the spike in we employed is a synthetic standard peptide mixture from *E.coli,* whereas Silva et al. [6] used a purified yeast enolase digest. Secondly, to overcome a limitation highlighted by Silva et al. [6] regarding the accurate selection of the same top 3 peptides in every sample, we determined the three most stable *E.coli* peptides among several runs. Thus, only these three peptides were included in the peptide database. This ensured a non-varying selection of peptides for an accurate absolute quantification across all runs.

First, we evaluated the peptides identified on the whole as well as the corresponding protein set. As recently shown in our previous research, the UDMS$^{E}$ method enables the identification of a high number of unique peptides ensuring the quantification of a large set of non-redundant proteins with good instrument and technical variation [8]. The analysis of the plasma samples of this study identified 242 quantifiable plasma proteins. It was carried out without any steps susceptible to variability, such as pre-fractionation or depletion of highly abundant proteins. The total number of proteins was found to be lower in comparison to the 311 EDTA plasma proteins quantified relatively after depletion of six highly abundant plasma proteins within our previous study [8]. This difference is caused by the ion suppression effect of the high abundant proteins corresponding peptides in the mass spectrometry detector. Thus, the total peptide and protein coverage is slightly smaller. This effect is also reflected in the overall variation of the proteins. In this study, protein variation was found to be higher (CV = 0.36) than the variation identified for the depleted plasma samples (CV = 0.28) in our previous study [8]. Nevertheless, the wide dynamic range of plasma proteins itself is also covered by the wide dynamic quantification range of 105 of our method, being also in agreement with Percy et al. [11] and Domanski et al. [12]. However, the overall findings seem to be in accordance with the current research status on plasma proteins.

Next, we examined the overall characteristics of the large set of the 242 quantified plasma proteins. Our first attempt was to determine the general protein coverage. Hence, the majority of the proteins were represented by classical proteins, predominantly secreted from the liver. The set mostly included proteins which are frequently tested in clinical laboratories and which are involved in important biological processes, such as immune response or the coagulation cascade. Among those proteins, we disclosed a very stable set of 63 proteins, which included medium to highly abundant proteins. Looking into more detail, important protein families, such as apolipoproteins, complement components, coagulation factors, and carrier molecules were among the set. These proteins are involved in the metabolic syndrome, coagulation disorders, the immune system, and the acute phase response. Therefore, their quantification could support screening, diagnosis and monitoring of the health status as well as disease related alterations. Thus, our method proved to be applicable for diverse protein panels.

It is tempting to verify the methodical concept of our study. We chose to compare the plasma concentration of our quantified proteins with reported values. First, we compared our results to the absolute amount of 153 high and medium abundant plasma proteins reviewed by Hortin et al. [10]. Among these, we found 93 proteins in common with our protein set and the absolute concentrations showed high correlation. One possible explanation for the fact that the correlation was not stronger is the source diversity of the reviewed proteins by Hortin et al. [10]. Protein concentrations were reported after being obtained by different proteomic methods, namely 2D electrophoresis and different mass spectrometry approaches. Also, concentrations were reported from diverse blood specimens, such as plasma and serum. Moreover, some of the concentrations could only be reported as an estimated value. Therefore, we consider the results of our comparison to the data reviewed by Hortin et al. [10] as well-founded.

Furthermore, Percy et al. [11] quantified 142 proteins implicated in non-communicable diseases using a multiplexed MRM approach. Among this set, we found 88 proteins in common with our protein set and the absolute plasma concentrations were found to highly correlate. The study of Percy et al. was based on 18 samples taken from subjects of very different ages and with different ethnicities, as well as an unequal male to female ratio. Moreover, the subjects were afflicted with cardiovascular diseases. By contrast, biological variance was kept at the lowest level possible within our study by ensuring low age difference, equal sex ratio and healthy status. It is well known that biological variance is an important factor that leads to differences in the plasma concentrations of proteins. This bias could therefore explain why the comparison of the common proteins did not exhibit strong correlation.

Moreover, hemoglobin subunits and pregnancy zone protein concentrations were evaluated. HBA and HBB plasma levels in the male samples were higher than in the female samples, as also reviewed by Murphy et al. [13]. Pregnancy zone plasma levels were also in agreement with the reviewed values [14], being higher among females than males. Altogether, our method is in accordance with the current research status on plasma proteins which are known to be a reliable matrix of potential biomarkers for the clinical use [15].

Nevertheless, our results need to be verified in larger plasma sample sets. Furthermore, technical variations in mass spectrometric analyses are substantially higher than those of ELISAs developed for clinical diagnosis, and have therefore to be considered in the interpretation of the results. Plasma protein concentration databases could be developed further and could be used for the screening and monitoring of the health status.

## Conclusions

Our study describes an absolute plasma protein quantification method using a state-of-the-art data independent acquisition nano-LC Q-TOF UDMS$^E$ proteomics approach. A high number of proteins in blood plasma could be analyzed in a single run, covering a wide dynamic range and showing high stability at protein level. The method has proven to be a reliable tool for the proteomics community, offering a simple and fast method for the absolute quantification of blood proteins in a single run analysis. Without requiring additional pre-fractionation steps, which are time consuming and costly, and being entirely in accordance with other methods currently employed for biomarker research, this method is worth further validation over a larger number of samples and application for biomarker verification in the clinical practice.

## Acknowledgements

We would like to acknowledge the European Social Fund, Human Resources Development Operational Programme 2007-2013 (Project no. POSDRU/159/1.5/136893), the Deutscher Akademischer Austauschdienst (German Academic Exchange Service) (Programme ID 57130104, Personal number: 91558112), the ERASMUS + Traineeship (Contract no. 06/24/08/2016) and the Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca Romania (Grant no. 7690/57/2016 and 5200/50/2017) for the research grants assigned to Maria Ilies. We thank Thomas Thiele (Institute for Transfusion Medicine, University Medicine Greifswald) for performing blood collection.

**References**

1. Riondino S, Ferroni P, Spila A, Alessandroni J, D'Alessandro R, Formica V, et al. Ensuring sample quality for biomarker discovery studies - use of ICT tools to trace biosample life-cycle. Cancer Genomics Proteomics. 2015;12:291-299.
2. Huang Z, Ma L, Huang C, Li Q, Nice EC. Proteomic profiling of human plasma for cancer biomarker discovery. Proteomics. 2017 Mar;17(6). doi: 10.1002/pmic.201600240. Epub 2016 Oct 17.
3. Sabbagh B, Mindt S, Neumaier M, Findeisen P. Clinical applications of MS-based protein quantification. Proteomics Clin Appl. 2016;10:323-345.
4. Zhao X, Qureshi F, Eastman PS, Manning WC, Alexander C, Robinson WH, et al. Pre-analytical effects of blood sampling and handling in quantitative immunoassays for rheumatoid arthritis. J Immunol Methods. 2012;378:72-80.
5. Liu Y, Buil A, Collins BC, Gillet LC, Blum LC, Cheng LY, et al. Quantitative variability of 342 plasma proteins in a human twin population. Mol Syst Biol. 2015 Feb 4;11(1):786.
6. Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics. 2006;5:144-156.
7. Ilies M, Iuga CA, Loghin F, Dhople VM, Thiele T, Völker U, et al. Data on the impact of the blood sample collection methods on blood protein profiling studies. Data Brief. 2017;14:313-319.
8. Ilies M, Iuga CA, Loghin F, Dhople VM, Thiele T, Völker U, et al. Impact of blood sample collection methods on blood protein profiling studies. Clin Chim Acta. 2017;471:128-134.
9. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72:248-254.
10. Hortin GL, Sviridov D, Anderson NL. High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. Clin Chem. 2008;54:1608-1616.
11. Percy AJ, Chambers AG, Yang J, Hardie DB, Borchers CH. Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility. Biochim Biophys Acta. 2014;1844:917-926.
12. Domanski D, Percy AJ, Yang J, Chambers AG, Hill JS, Freue GV, et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics. 2012;12:1222-1243.
13. Murphy WG. The sex difference in haemoglobin levels in adults - mechanisms, causes, and consequences. Blood Rev. 2014;28:41-47.
14. Folkersen J, Teisner B, Grunnet N, Grudzinskas JG, Westergaard JG, Hindersson P. Circulating levels of pregnancy zone protein: normal range and the influence of age and gender. Clin Chim Acta. 1981;110:139-145.
15. Sajic T, Liu Y, Aebersold R. Using data-independent, high-resolution mass spectrometry in protein biomarker research: perspectives and clinical applications. Proteomics Clin Appl. 2015;9:307-321.