# Development and Validation of a Novel Plasma Protein Signature for Breast Cancer Diagnosis by Using Multiple Reaction Monitoring-based Mass Spectrometry

HAN-BYOEL LEE[1*], UN-BEOM KANG[2*], HYEONG-GON MOON[1,3], JIWOO LEE[3], KYUNG-MIN LEE[3], MINJU YI[3], YONG SUN PARK[2], JONG WON LEE[4], JONG-HAN YU[4], SEUNG HO CHOI[5], SANG HEON CHO[5], CHEOLJU LEE[6], WONSHIK HAN[1,3] and DONG-YOUNG NOH[1,3]

[1]*Department of Surgery, Seoul National University College of Medicine, Seoul, Republic of Korea;*
[2]*R&D Center, Biomedieng Co. Ltd., Gyeonggi-do, Republic of Korea;*
[3]*Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea;*
[4]*Department of Surgery, Asan Medical Center, Ulsan University College of Medicine, Seoul, Republic of Korea;*
[5]*Healthcare System Gangnam Center, Healthcare Research Institute, Seoul National University Hospital, Seoul, Republic of Korea;*
[6]*Life Sciences Division, Korea Institute of Science and Technology, Seoul, Republic of Korea*

**Abstract.** *Aim: We aimed to develop a plasma protein signature for breast cancer diagnosis by using multiple reaction monitoring (MRM)-based mass spectrometry. Materials and Methods: Based on our previous studies, we selected 124 proteins for MRM. Plasma samples from 80 patients with breast cancer and 80 healthy women were used to develop a plasma proteomic signature by an MRM approach. The proteomic signature was then validated in plasma samples from 100 patients with breast cancer and 100 healthy women. Results: A total of 56 proteins were optimized for MRM. In the verification cohort, 11 proteins exhibited significantly differential expression in plasma from patients with breast cancer. Three proteins (neural cell adhesion molecule L1-like protein, apolipoprotein C-1 and carbonic anhydrase-1) with highest statistical significance which gave consistent results for patients of stage I and II breast cancer were selected and a 3-protein signature was developed using binary logistic regression analysis [area under the curve (AUC)=0.851, sensitivity=80.6%]. The 3-protein signature showed similar performance in an independent validation cohort with an AUC of 0.797 and sensitivity of 77.2% for detection of stage I and II breast cancer. Conclusion: We developed a distinct plasma protein signature for breast cancer diagnosis based on an MRM-based approach, and the clinical value of the 3-protein signature was validated in an independent cohort.*

During the past several decades, the treatment outcome of patients with breast cancer has dramatically improved. Recent statistics have suggested that breast cancer-related mortality is actually beginning to decline in Western women. The main reasons for survival improvement are the wide use of adjuvant systemic chemotherapy and the public adoption of breast cancer screening (1).

Early detection of breast cancer through screening programs utilizes physical examination and breast imaging (screening mammography) (2). Systematic review of data from randomized trials has shown that screening mammography reduces breast cancer mortality for women aged 39 to 69 years (3). However, screening mammography also carries certain limitations, such as the substantial discomfort during the test and its limited accuracy, which lead to over-diagnosis, unnecessary biopsies, and failure to detect occult carcinomas (4, 5).

To overcome these issues, recent efforts have focused on developing novel genomic or proteomic methods with the purpose of increasing early breast cancer detection using blood samples (6). In particular, proteomics is an interesting means of developing novel diagnostic tools since cancer cells may secrete cancer-specific proteins and certain membrane

*These Authors contributed equally to this study.

*Correspondence to:* Dong-Young Noh, MD, Ph.D., Department of Surgery and Cancer Research Institute, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 110-744, Republic of Korea. Tel: +82 220722921, Fax: +82 27663975, e-mail: dynoh@snu.ac.kr

*Key Words:* Breast cancer, proteomics, multiple reaction monitoring, diagnosis, plasma.

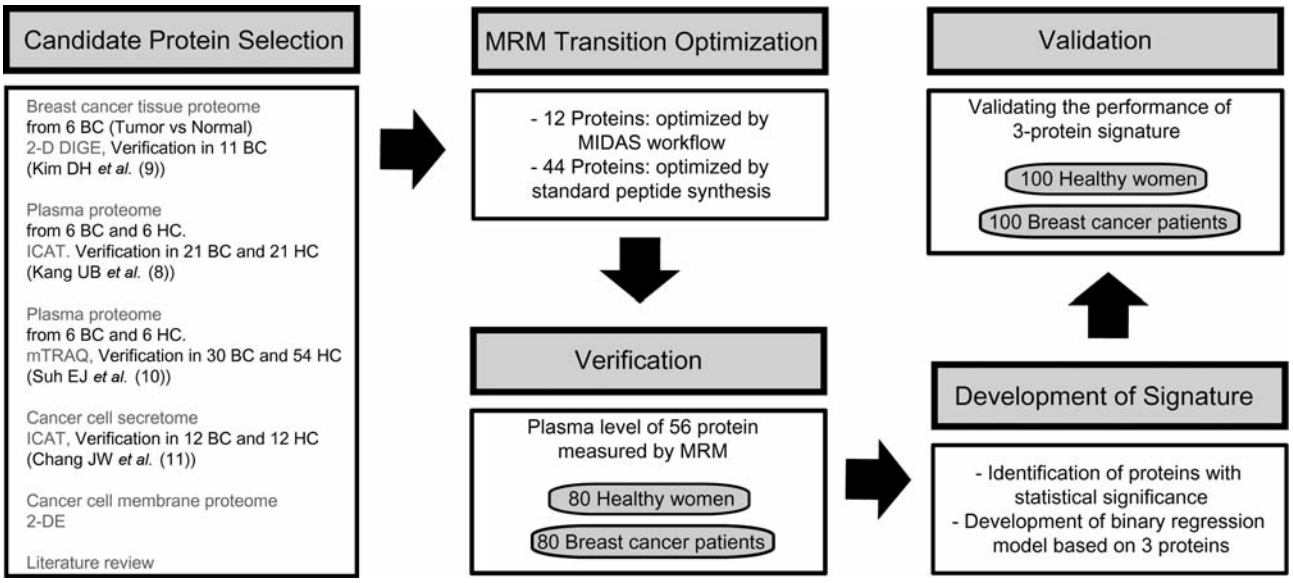


Figure 1. *The study scheme of plasma proteomic signature discovery for breast cancer. BC: Breast cancer; HC: healthy control; 2-D DIGE: 2-dimensional differential gel electrophoresis; ICAT: isotope-coded affinity tagging; mTRAQ: mass differential tags for relative and absolute quantification; 2-DE: 2-dimensional electrophoresis; MRM: multiple reaction monitoring; MIDAS: multiple reaction monitoring-initiated detection and sequencing.*

proteins can be shed from carcinomas during cell proliferation. It is well-appreciated that complex cancer biology can be better-captured by multi-marker signatures than by a single-marker approach (7). Recent advances in proteomic technologies, such as multiple reaction monitoring (MRM), have made the multi-marker approach feasible in a limited amount of human sample.

In the present study, we selected a list of candidate proteins based on our previous proteomic studies using breast cancer biospecimens as well as literature review. Using an MRM-based approach, we established a plasma proteomic signature comprising of three plasma proteins for breast cancer diagnosis. The diagnostic performance of the plasma proteomic signature was further validated in an independent cohort of patients with breast cancer and healthy women.

## Materials and Methods

Characteristics of studied women and blood sampling. For verification and validation, plasma samples of patients with breast cancer were obtained before the definitive treatment of their breast cancer at Seoul National University Hospital. Patients with *in situ* carcinoma and patients with recurrent breast cancer were excluded from this study.

To obtain blood samples from healthy women, those who visited the Seoul National University Hospital Healthcare System Gangnam Center for routine health check-up were invited. During the comprehensive explanation for their health check-up, the participants were also informed of this research and the additional blood sampling of 3-5 cc for research purposes. Between September 2010 and October 2010, a total of 200 healthy women who did not show any abnormality in their breast cancer screening consented to this study and donated their blood samples. This study comprises of two analytical evaluations, the verification and the validation study. Two independent MRM experiments were carried out for verification and validation using plasma from 80 and 100 patients with breast cancer and healthy volunteer pairs for each cohort, respectively (Figure 1).

This study was approved by the Institutional Review Board of Seoul National University Hospital (IRB No. 0512-502-163), and was conducted according to the declaration of Helsinki.

*Preparation of blood samples*. Plasma samples were drawn from peripheral veins and were temporarily stored in tubes containing ethylene diaminetetra-acetic acid (EDTA) to prevent coagulation. Samples were then transferred to the laboratory and underwent centrifugation at 1,300 × *g* for 10 min at 4˚C. The supernatant plasma was filtered through a cellulose acetate filter (0.2 μm pore size) and platelet-free plasma was stored at −80˚C for further use.

For mass spectrometry (MS), plasma protein concentration was determined by the Bradford assay. The plasma protein samples (200 μg) were denatured by incubation in 50 mM Tris buffer (pH 8.0) containing 3 M urea at 37˚C for 30 min. Samples were reduced with 10 mM dithiothreitol for 1 h at 56˚C, treated with 60 mM iodoacetamide for 1 h at room temperature in the dark, and then diluted 10-fold with 50 mM ammonium bicarbonate. Digestion was performed with sequencing-grade trypsin (Promega, Madison, WI, USA) at 37˚C overnight at a protein:trypsin molar ratio of 50:1.

Tryptic digests were desalted using a C18 SPE cartridge (Waters, Milford, MA, USA) and dried *in vacuo*. The dried samples were dissolved in 0.1% formic acid. One hundred femtomoles of a beta-galactosidase (β-Gal) peptide (residues 954-962, GDFQFNISR) was added to the desalted peptide mixture as a relative internal standard peptide for the MRM runs.

*Synthetic peptide for MRM transition optimization*. Eighty-three peptides from 44 proteins that were not identified through the MIDAS workflow were synthesized for MRM verification. These 83 peptides were selected based on conditions for containing the tryptic end, and not including any modification sites within the useful range of 8-20 amino acids. To specify peptides in designated protein targets, peptides were searched with BLASTP for exact matches against the human SWISS-PROT peptides using NCBI BLAST (www.ncbi.nlm.nih.gov/blast). The peptides were synthesized by JPT SpikeTides™ services (http://www.jpt.com).

*MRM*. MRM runs were performed for the predetermined transitions using a triple quadrupole linear ion trap in the MRM mode. An aliquot (~10 μg) was then injected into a reversed-phase HALO C18 column (Advanced Materials Technology, Inc., Wilmington, DE, USA) (10 cm ×500 μm) on an Eksigent micro-UPLC system (AB Sciex, Foster City, CA, USA) at a flow rate of 30 μl/min. The column was equilibrated with 95% buffer A (0.1% formic acid in water) and 5% buffer B (0.1% formic acid in acetonitrile) prior to use. The peptides from plasma with internal standard peptide (β-Gal) were eluted with a linear gradient of 10-40% buffer B over 40 min.

Electrospray MS data were collected using the Turbo V™ Source on a 5500 Q TRAP hybrid triple quadupole/linear ion trap instrument (AB Sciex), and the peaks were integrated using quantitation procedures in the Analyst software 1.4.2 (Intelli-Quan algorithm). MRM transitions were acquired at unit resolution in both Q1 and Q3 quadrupoles to maximize specificity. The source temperature was set to 400˚C, and the source voltage was set to 4500 V. The declustering potential was set to 100 V and entrance potential was set to 10 V. The curtain gas was set to 15, and collision gas was set to medium. The collision energy (CE) for each transition was based on theoretical values that were calculated from the equation CE=0.044 × (m/z) + 8.5 for ($M^{+2}H^{+}$) ions. The scan time was maintained at 50 ms for each transition, and the pause time between transition scans was set to 5 ms.

*Data acquisition and statistical analysis*. Data from the intensity chromatograms of the transitions were extracted with the MultiQuant program (AB Sciex). Peak areas for transitions were extracted and normalized *versus* internal standard transitions (Q1/Q3 transitions at 542.3/636.3 m/z for the β-Gal peptide; see above for sequence). Each normalized peak area for the individual transitions was compared with the corresponding transition peaks of other runs to estimate the relative differences between individual plasma samples.

For statistical analysis, Student's *t*-test was used to compare the concentration of proteins between patients with cancer and healthy women, and binary logistic regression analysis was used to create the 3-protein signature for breast cancer diagnosis. All statistical analyses were performed using IBM SPSS Statistic software version 19 (IBM, Armonk, NY, USA).

## Results

*Selection of plasma protein candidates for MRM-based breast cancer diagnostics*. Based on our previous studies on breast cancer-specific biomarkers (8-11), we were able to identify a total of 124 proteins that were significantly differentially expressed between cancer and normal tissue (Supplementary Table I available at http://lbcb.snu.ac.kr/g4/bbs/board.php?bo_table=pds&wr_id=2). Briefly, our previous studies were based on five experimental settings; i) isotope-coded affinity tag (ICAT) labeling and tandem MS using plasma from patients with breast cancer and healthy women (8); ii) 2-dimensional differential gel electrophoresis (2-D DIGE) experiment comparing breast cancer tissue and non-cancerous tissues (9); iii) mass differential tags for relative and absolute quantification (mTRAQ)-based stable isotope-labeling MS using plasma from patients with breast cancer and healthy women (10), iv) 2-D electrophoresis of breast cancer cell line secretome (11), and ICAT analysis of cancer cell line membrane proteins. We also performed a literature review of plasma proteins reported to be specific to patients with breast cancer and identified 29 proteins (12-14). Among the 124 candidate proteins, the candidates for MRM were narrowed down to 56 proteins based on the results of the MIDAS workflow (12 proteins) and the successful identification after peptide synthesis (44 proteins) (Figure 1) (Table I).

*Identification of three significant plasma proteins and development of a diagnostic model*. The plasma concentration of the selected 56 proteins was measured by MRM for 80 healthy women and 80 patients with breast cancer (verification cohort). Among the 56 plasma proteins, 11 proteins were statistically significantly differentially expressed in blood from patients with breast cancer when compared to that of healthy women (Figure 2A). The plasma levels of these proteins were also compared between healthy women and patients with stage I or II breast cancer since these tumors are often asymptomatic, and hence are best candidates for breast cancer screening. Among the 11 proteins with significant differential expression, four lost their statistical significance in stage I and II breast cancer (gelsolin, plasma protease C1 inhibitor, apolipoprotein D, and glutathione peroxidase 3) (Figure 2B).

The three proteins (neural cell adhesion molecule L1-like protein, apolipoprotein C-1, carbonic anhydrase-1) with highest statistical significance for both normal *vs*. all stage and normal *vs*. stage I/II were selected for further binary regression analysis. Based on the B weight from the binary regression analysis, the following model was created for the three proteins to predict breast cancer: model=0.604× [carbonic anhydrase-1] + 7.575×[neural cell adhesion molecule L1-like protein] − 0.523×[apolipoprotein C-1]. In the verification

Table I. *The list of 56 proteins used for multiple reaction monitoring-based mass spectrometry.*

| Accession Number | Gene | Description | Transition | | Sequence | CE | Source |
|---|---|---|---|---|---|---|---|
| | | | Q1 | Q3 | | | |
| P02654 | *APOC1* | Apolipoprotein C-I | 526.8 | 776.4 | EFGNTLEDK | 31.68 | MIDAS |
| P02671 | *FGA* | Fibrinogen Alpha Fragment | 553.8 | 879.5 | VQHIQLLQK | 32.87 | MIDAS |
| P05090 | *APOD* | Apolipoprotein D | 615.8 | 890.5 | NILTSNNIDVK | 35.60 | MIDAS |
| P05155 | *SERPING1* | Plasma protease C1 inhibitor | 609.7 | 771.4 | GVTSVSQIFHSPDLAIR | 34.98 | MIDAS |
| P06396 | *GSN* | Gelsolin | 441.7 | 710.4 | TGAQELLR | 27.94 | MIDAS |
| P08519 | *LPA* | Apolipoprotein(a) | 521.8 | 721.4 | GTYSTTVTGR | 31.46 | MIDAS |
| P0C0L5 | *C4B-1* | Complement component C4B | 782.4 | 836.5 | TTNIQGINLLFSSR | 42.93 | MIDAS |
| P19652 | *ORM2* | Alpha-1-acid glycoprotein 2 | 617.9 | 869.5 | EHVAHLLFLR | 35.69 | MIDAS |
| P43251 | *BTD* | Biotinidase | 654.4 | 751.4 | LSSGLVTAALYGR | 37.29 | MIDAS |
| P55786 | *NPEPPS* | Puromycin-sensitive aminopeptidase | 565.3 | 870.5 | AGIISTVEVLK | 33.38 | MIDAS |
| Q03591 | *CFHR1* | Complement factor H-related protein 1 | 665.8 | 552.3 | INHGILYDEEK | 37.80 | MIDAS |
| P05109 | *S100A8* | Protein S100-A8 | 636.9 | 774.4 | ALNSIIDVYHK | 36.52 | IDA |
| P08571 | *CD14* | Monocyte differentiation antigen CD14 | 456.8 | 641.3 | ATVNPSAPR | 28.60 | IDA |
| O00533 | *CHL1* | Neural cell adhesion molecule L1-like protein | 642.8 | 836.4 | GDLYFANVEEK | 36.78 | IDA |
| P06276 | *BCHE* | Cholinesterase | 614.3 | 820.4 | IFFPGVSEFGK | 35.53 | IDA |
| Q9UGM5 | *FETUB* | Fetuin-B | 456.8 | 700.4 | LVVLPFPK | 28.60 | IDA |
| P06702 | *S100A9* | Protein S100-A9 | 904.0 | 1236.6 | NIETIINTFHQYSVK | 48.28 | IDA |
| P02751 | *FN1* | Fibronectin | 555.8 | 821.4 | STTPDITGYR | 32.95 | IDA |
| P22352 | *GPX3* | Glutathione peroxidase 3 | 374.7 | 546.3 | TTVSNVK | 24.99 | IDA |
| P00915 | *CA1* | Carbonic anhydrase 1 | 485.8 | 758.4 | VLDALQAIK | 29.88 | IDA |
| O15439 | *ABCC4* | Multidrug resistance-associated protein 4 | 538.3 | 733.4 | AEAAALTETAK | 32.19 | IDA |
| O95045 | *UPP2* | Uridine phosphorylase 2 | 723.9 | 943.6 | FEQVILDNIVTR | 40.35 | IDA |
| P00746 | *CFD* | Complement factor D | 895.6 | 331.2 | VQVLLGAHSLSQPEPSK | 47.91 | IDA |
| P02741 | *CRP* | C-Reactive protein | 564.8 | 696.4 | ESDTSYVSLK | 33.35 | IDA |
| P06126 | *CD1A* | T-Cell surface glycoprotein CD1a | 523.8 | 786.5 | FILGLLDAGK | 31.55 | IDA |
| P07996 | *THBS1* | Thrombospondin-1 | 651.9 | 960.5 | AGTLDLSLTVQGK | 37.18 | IDA |
| P09086 | *POU2F2* | POU domain, class 2, transcription factor 2 | 751.4 | 939.5 | LYGNDFSQTTISR | 41.56 | IDA |
| P09172 | *DBH* | Dopamine beta-hydroxylase | 559.8 | 791.5 | TPEGLTLLFK | 33.13 | IDA |
| P18031 | *PTPN1* | Tyrosine-protein phosphatase non-receptor type 1 | 430.7 | 648.4 | DPSSVDIK | 27.45 | IDA |
| P31944 | *CASP14* | Caspase-14 | 782.4 | 922.5 | DPTAEQFQEELEK | 42.92 | IDA |
| P42338 | *PIK3CB* | Phosphatidylinositol-4,5-*bis*phosphate 3-kinase catalytic subunit beta isoform | 523.3 | 875.4 | GLHEFDSLK | 31.52 | IDA |
| P50991 | *CCT4* | T-Complex protein 1 subunit delta | 709.4 | 932.5 | GDVTITNDGATILK | 39.71 | IDA |
| P51884 | *LUM* | Lumican; Keratan sulfate proteoglycan lumican | 469.2 | 523.3 | EDAVSAAFK | 29.15 | IDA |
| P61769 | *B2MG* | Beta-2-microglobulin | 575.1 | 460.7 | VEHSDLSFSK | 33.80 | IDA |
| Q12986 | *NFX1* | Transcriptional repressor NF-X1 | 755.4 | 501.3 | FNTDAAEFIPQEK | 41.74 | IDA |
| Q13608 | *PEX6* | Peroxisome assembly factor 2 | 611.3 | 816.4 | YLEGSIAPEDK | 35.40 | IDA |
| Q38SD2 | *LRRK1* | Leucine-rich repeat serine/ threonine-protein kinase 1 | 758.7 | 530.3 | LTELPALFLHSFK | 41.88 | IDA |
| Q68E01 | *INTS3* | Integrator complex subunit 3 | 717.9 | 830.4 | VLAHLAPLFDNPK | 40.09 | IDA |
| Q6LA40 | *UMODL1* | Uromodulin-like 1 | 625.3 | 835.5 | TNAQVFEVTIK | 36.02 | IDA |
| Q6P052 | *MAN1A1* | MAN1A1 protein | 840.2 | 754.9 | GLPPVDFVPPIGVESR | 45.47 | IDA |
| Q6RI45 | *BRWD3* | Bromodomain and WD repeat-containing protein 3 | 786.9 | 931.5 | LINEGDVPHLPVNR | 43.13 | IDA |
| Q6TDP4 | *KLHL17* | Kelch-like protein 17 | 659.9 | 815.4 | YVLQHFVDVAK | 37.53 | IDA |
| Q6ZRS5 | | cDNA FLJ46139 fis | 492.3 | 653.4 | ADGSLHLDR | 30.16 | IDA |
| Q7Z7G2 | *CPLX4* | Complexin-4 | 504.8 | 809.4 | AQATFTEIK | 30.71 | IDA |
| Q8TBF2 | *C1orf93* | Uncharacterized protein C1orf93 | 591.3 | 874.5 | HAVTGEAVELR | 34.52 | IDA |
| Q92954 | *PRG4* | Proteoglycan 4 | 559.3 | 689.4 | TTPETTTAAPK | 33.11 | IDA |
| Q96AC1 | *FERMT2* | Fermitin family homolog 2 | 647.3 | 967.5 | YYSFFDLNPK | 36.98 | IDA |
| Q96DD0 | *LRRC39* | Leucine-rich repeat-containing protein 39 | 610.9 | 624.3 | LQELILSYNK | 35.38 | IDA |
| Q96EZ7 | *PPM1J* | Protein phosphatase 1J | 694.5 | 725.3 | VLSAYEPNDHSR | 39.06 | IDA |
| Q96MN2 | *NLRP4* | NACHT, LRR and PYD domains-containing protein 4 | 724.9 | 892.5 | DQVTISEIYQPR | 40.39 | IDA |
| Q9H2G9 | *BLZF1* | Golgin-45 | 802.4 | 972.5 | GEFLGQSEGVIEPNK | 43.81 | IDA |
| Q9P2F8 | *SIPA1L2* | Signal-induced proliferation-associated 1-like protein 2 | 651.3 | 816.4 | LDEQGLSFQHK | 37.16 | IDA |
| Q9UFB7 | *ZBTB47* | Zinc finger and BTB domain-containing protein 47 | 521.7 | 658.3 | GTPEPEEAGR | 31.46 | IDA |
| Q9UPU9 | *SAMD4A* | Protein Smaug homolog 1 | 512.6 | 894.4 | LGLLGTSGFVSSNQR | 31.06 | IDA |
| Q9Y623 | *MYH4* | Myosin-4 | 738.4 | 907.5 | DPLNETVVGLYQK | 40.99 | IDA |

CE: Collision energy; MIDAS: multiple reaction monitoring-initiated detection and sequencing; IDA: information-dependent acquisition.

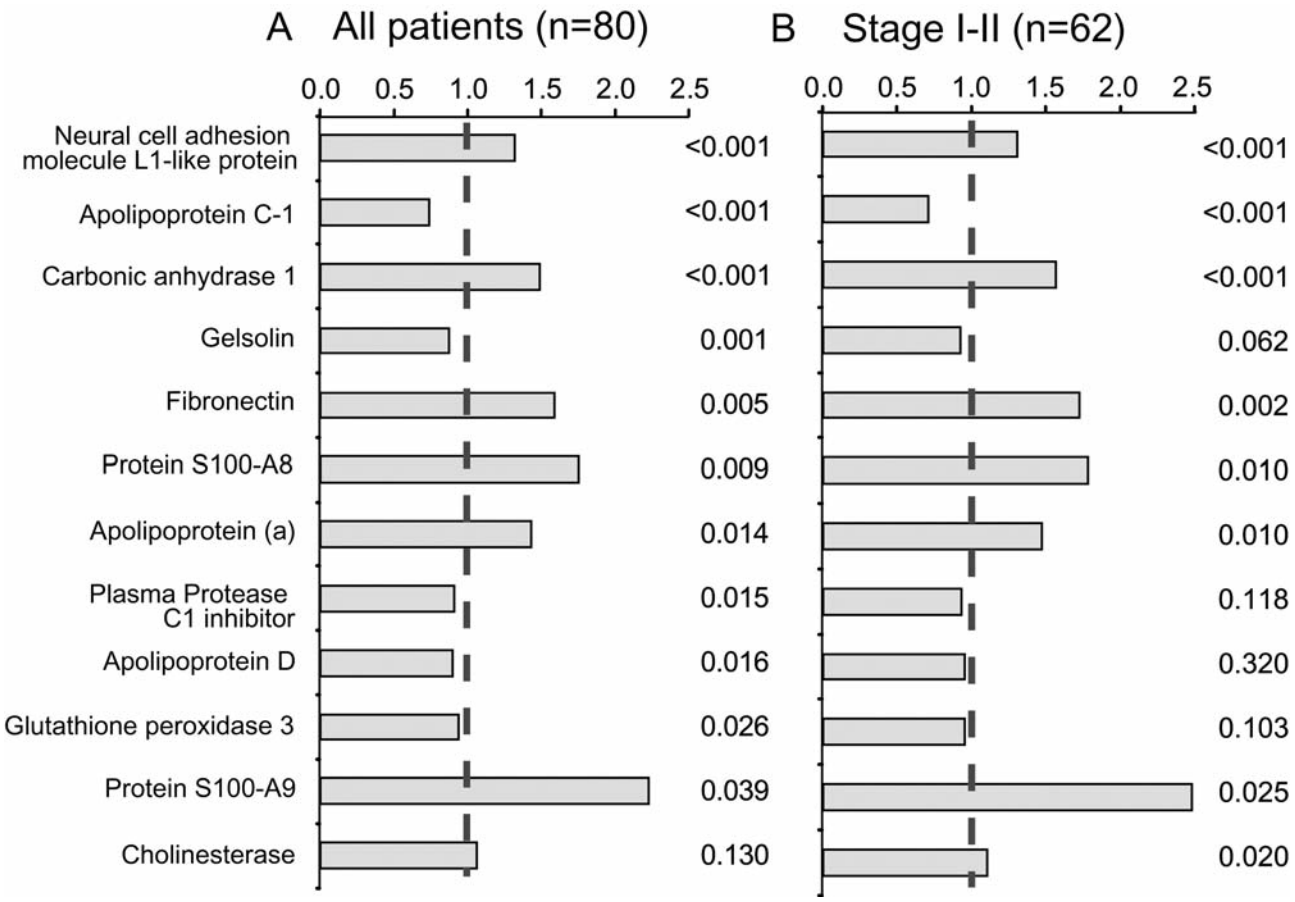


Figure 2. *Relative plasma concentration of significant proteins in the verification cohort. Statistically significant differential expression of 11 proteins in all patients with breast cancer (A) and in patients with stage I and II breast cancer (B) are shown in comparison to those in healthy women.*

cohort, the sensitivity, specificity and area under the curve (AUC) of the diagnostic model was 78.75%, 78.75% and 0.831, respectively (Figure 3A). The performance of the diagnostic model was further improved for patients with stage I and II disease, with an AUC of 0.851 (Figure 3B).

*Performance of the 3-protein diagnostic model in the validation cohort*. The performance of the 3-protein diagnostic model was validated in an independent cohort of 100 patients with breast cancer and 100 healthy women. The patterns of individual plasma concentration of the three proteins were similar in both verification and validation cohorts (Figure 4). The overall accuracy of the 3-protein model was slightly lower in the validation cohort when compared to that of the verification cohort (Figure 3C and D). However, the performance of the 3-protein model was still accurate in the validation cohort, with sensitivity, specificity and AUC of 68.7%, 69.4% and 0.746, respectively. The performance of the model was higher for stage I and stage II disease, as it was in the verification cohort (sensitivity=77.2%, specificity=63.3%, and AUC=0.797).

## Discussion

In this study, we developed a plasma protein model that may discriminate healthy women from those with breast cancer, based on our previous proteomic studies and from literature review. Our 3-protein model was capable of detecting breast cancer when tested in an independent cohort of 100 healthy women and 100 women with breast cancer, and its performance was higher in patients with stage I and stage II breast cancer, who may benefit from breast cancer screening by detecting tumors in the asymptomatic period (sojourn time) (15, 16).

We have adopted the strategy of using the recently established MS-based MRM method which allows efficient quantification of plasma protein concentration without the use of antibodies (16, 17). By using this MRM approach, we

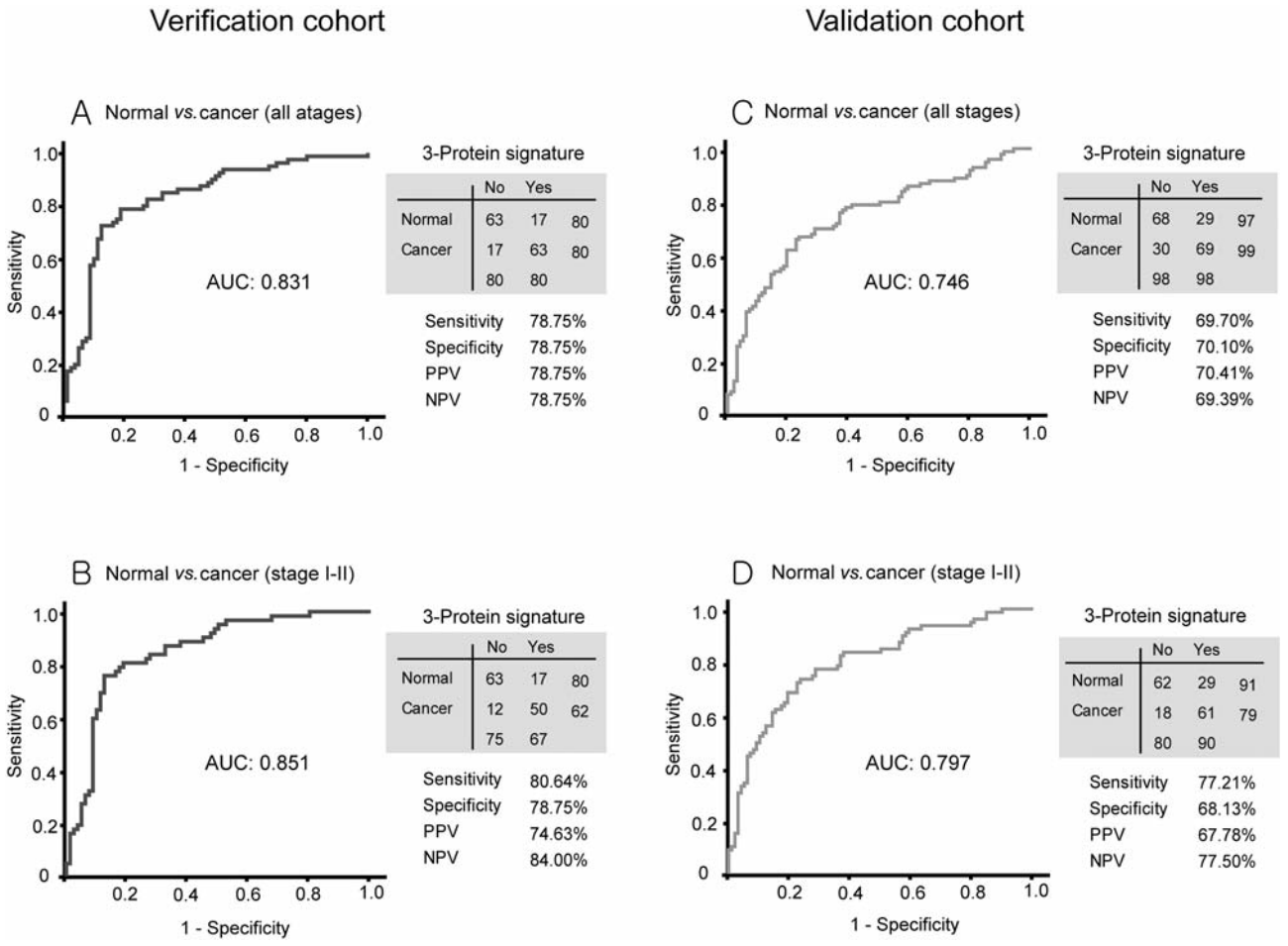


Figure 3. *Diagnostic accuracy of the 3-protein signature in the verification (A, B) and validation (C, D) cohorts. The performance of the diagnostic model was further improved for stage I and II disease [area under the curve (AUC) of 0.851 vs. 0.831) (B). The overall accuracy of the 3-protein model was slightly lower in the validation cohort when compared to that of the verification cohort (C, D). PPV: Positive predictive value; NPV: negative predictive value.*

were able to quantify the plasma concentration of various candidate proteins in a relatively large number of cases in a standardized fashion. Lessons from the recent mRNA expression signature development in cancer prognostication is that use of a multimarker panel can be far more efficient than that of the traditional single-marker approach (18). MS-based MRM seems to be a most suitable way of testing clinical relevance of multi-protein panels discovered in proteomic cancer research (19).

Many researchers have focused on developing blood markers of breast cancer detection based on the proteomic approach. Recent trends of proteomic discovery efforts can be divided into two main strategies based on the experimental materials for discovery: blood from human patients and healthy controls, and biospecimens from mouse models of breast cancer. The approach of using mouse models is promising in plasma protein discovery since it can create a relatively homogenous model for proteome identification by minimizing individual heterogeneity associated with human samples. Pitteri *et al.* reported a series of proteins up-regulated in a tumor-bearing mouse and a significant proportion of them were secretory proteins expressed in breast cancer cell lines (20). Whiteaker *et al.*, using a combination of antibody-based approaches and MS MRM based on mouse models of breast cancer, have also suggested fibulin-2 and osteopontin as potential plasma markers for breast cancer diagnosis (21). While this cell line mouse model approach can provide a biologically sound rationale, it is still unknown whether such specific designed models of mouse tumorigenesis can recapitulate the complex biology of tumor–host interaction in human plasma.

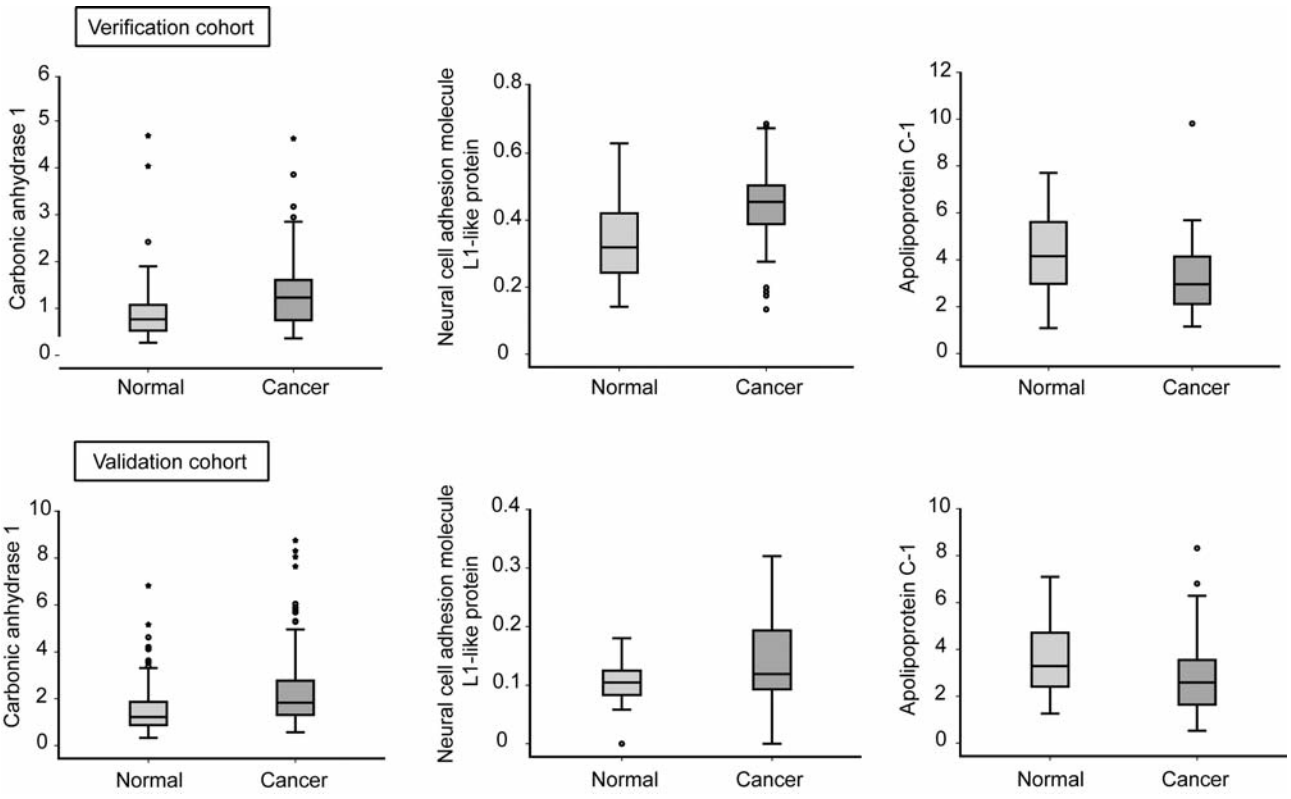


Figure 4. *Relative plasma concentration of the final three proteins in the verification and validation cohorts. The patterns of individual plasma concentration of the three proteins were similar between the verification and validation cohorts. Line in the box: median; box: 1st to 3rd quartile; whiskers: maximum and minimum value excluding outliers; dots: outlier.*

For the direct exploration of human blood samples, the surface-enhanced laser desorption/ionization–time-of-flight (SELDI-TOF) technique has often been used (6, 22-24). Other proteomic techniques such as matrix-assisted laser desorption/ionization–time-of-flight (MALDI-TOF) have also been applied for breast cancer diagnosis and have suggested various lists of differentially expressed proteins in blood of patients with breast cancer (7, 25). However, these studies of protein biomarker discovery for breast cancer diagnosis often suffer from small sample size and a lack of independent validation. In our study, we screened various proteins from our previously reported proteomic studies in 80 patients with breast cancer and 80 healthy controls. Neural cell adhesion molecule L1-like protein, apolipoprotein C-1 and carbonic anhydrase-1, with the highest statistical significance, were chosen for further validation in an independent cohort of patients with breast cancer and healthy controls. In addition to the similar performance in breast cancer diagnosis, the consistent expression patterns of these three proteins in both the verification and validation cohort suggests the potential role of these proteins for blood diagnosis of breast cancer.

The three proteins identified in our study as potential plasma markers of breast cancer diagnosis are all involved in the carcinogenesis of various human carcinoma types (26-28). Of note, apolipoprotein C-1 has been identified as a differentially expressed protein in various models of cancer proteomics. Fan *et al*. reported the decreased serum concentration of apolipoprotein C-1 in patients with breast cancer through their SELDI-TOF analysis of serum from 124 patients with breast cancer and 158 controls (13). Decreased expression of apolipoprotein C-1 in blood of patients with cancer was also observed for gastric cancer and non-small cell lung cancer (29, 30). However, other studies have suggested the increased expression of apolipoprotein C-1 in the serum of patients with prostate and pancreatic cancer, suggesting the need for further validation and potential organ-specific differences (31, 32).

A limitation of the present study is that our plasma samples were pre-collected retrospective specimens. The follow-up period of the healthy controls is also less than 2 years. A recent study has suggested that proteomic profiles may change even 3 years before the onset of clinically symptomatic breast cancer (24).

To summarize, we have developed a potential breast cancer diagnostic proteomic profile by screening our previous candidate cancer-specific proteins. The 3-protein signature (a regression model based on the relative expression of neural cell adhesion molecule L1-like protein, apolipoprotein C-1, and carbonic anhydrase-1) was validated in an independent cohort with acceptable accuracy. The clinical value of the newly developed 3-protein signature should be tested in a prospective study in comparison to standard mammographic screening.

## Conflicts of Interest

The Authors declare that they have no conflict of interest in regard to this study.

## Acknowledgements

This work was supported by grant from the Korean Healthcare Technology R&D project through the Korean Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI13C2148 and HI14C1277)

## References

1 Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD and Feuer EJ: Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med *353*: 1784-1792, 2005.

2 DeSantis C, Siegel R, Bandi P and Jemal A: Breast cancer statistics, 2011. CA Cancer J Clin *61*: 409-418, 2011.

3 Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK and Humphrey L: Screening for breast cancer: an update for the U.S. Preventive Services Task Force. Ann Intern Med *151*: 727-737, W237-742, 2009.

4 Partridge AH and Winer EP: On mammography – more agreement than disagreement. N Engl J Med *361*: 2499-2501, 2009.

5 Gotzsche PC and Nielsen M: Screening for breast cancer with mammography. Cochrane Database Syst Rev: CD001877, 2011.

6 Jotwani AC and Gralow JR: Early detection of breast cancer: New biomarker tests on the horizon? Mol Diagn Ther *13*: 349-357, 2009.

7 Pietrowska M, Marczak L, Polanska J, Behrendt K, Nowicka E, Walaszczyk A, Chmura A, Deja R, Stobiecki M, Polanski A, Tarnawski R and Widlak P: Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early-stage breast cancer. J Transl Med *7*: 60, 2009.

8 Kang UB, Ahn Y, Lee JW, Kim YH, Kim J, Yu MH, Noh DY and Lee C: Differential profiling of breast cancer plasma proteome by isotope-coded affinity tagging method reveals biotinidase as a breast cancer biomarker. BMC Cancer *10*: 114, 2010.

9 Kim DH, Bae J, Lee JW, Kim SY, Kim YH, Bae JY, Yi JK, Yu MH, Noh DY and Lee C: Proteomic analysis of breast cancer tissue reveals up-regulation of actin-remodeling proteins and its relevance to cancer invasiveness. Proteomics Clin Appl *3*: 30-40, 2009.

10 Suh EJ, Kabir MH, Kang UB, Lee JW, Yu J, Noh DY and Lee C: Comparative profiling of plasma proteome from breast cancer patients reveals thrombospondin-1 and BRWD3 as serological biomarkers. Exp Mol Med *44*: 36-44, 2012.

11 Chang JW, Kang UB, Kim DH, Yi JK, Lee JW, Noh DY, Lee C and Yu MH: Identification of circulating endorepellin LG3 fragment: Potential use as a serological biomarker for breast cancer. Proteomics Clin Appl *2*: 23-32, 2008.

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

13 Fan Y, Wang J, Yang Y, Liu Q, Yu J, Zheng S and Li M: Detection and identification of potential biomarkers of breast cancer. J Cancer Res Clin Oncol *136*: 1243-1254, 2010.

14 Gromov P, Gromova I, Bunkenborg J, Cabezon T, Moreira JM, Timmermans-Wielenga V, Roepstorff P, Rank F and Celis JE: Up-regulated proteins in the fluid bathing the tumour cell microenvironment as potential serological markers for early detection of cancer of the breast. Mol Oncol *4*: 65-89, 2010.

15 Duffy SW, Chen HH, Tabar L and Day NE: Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. Stat Med *14*: 1531-1543, 1995.

16 Latterich M, Abramovitz M and Leyland-Jones B: Proteomics: new technologies and clinical applications. Eur J Cancer *44*: 2737-2741, 2008.

17 Mollah S, Wertz IE, Phung Q, Arnott D, Dixit VM and Lill JR: Targeted mass spectrometric strategy for global mapping of ubiquitination on proteins. Rapid Commun Mass Spectrom *21*: 3357-3364, 2007.

18 Sotiriou C and Pusztai L: Gene-expression signatures in breast cancer. N Engl J Med *360*: 790-800, 2009.

19 Turtoi A, De Pauw E and Castronovo V: Innovative proteomics for the discovery of systemically accessible cancer biomarkers suitable for imaging and targeted therapies. Am J Pathol *178*: 12-18, 2011.

20 Pitteri SJ, Faca VM, Kelly-Spratt KS, Kasarda AE, Wang H, Zhang Q, Newcomb L, Krasnoselsky A, Paczesny S, Choi G, Fitzgibbon M, McIntosh MW, Kemp CJ and Hanash SM: Plasma proteome profiling of a mouse model of breast cancer identifies a set of up-regulated proteins in common with human breast cancer cells. J Proteome Res *7*: 1481-1489, 2008.

21 Whiteaker JR, Zhang H, Zhao L, Wang P, Kelly-Spratt KS, Ivey RG, Piening BD, Feng LC, Kasarda E, Gurley KE, Eng JK, Chodosh LA, Kemp CJ, McIntosh MW and Paulovich AG: Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. J Proteome Res *6*: 3962-3975, 2007.

22 Li J, Zhang Z, Rosenzweig J, Wang YY and Chan DW: Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin Chem *48*: 1296-1304, 2002.

23 Belluco C, Petricoin EF, Mammano E, Facchiano F, Ross-Rucker S, Nitti D, Di Maggio C, Liu C, Lise M, Liotta LA and Whiteley G: Serum proteomic analysis identifies a highly sensitive and specific discriminatory pattern in stage 1 breast cancer. Ann Surg Oncol *14*: 2470-2476, 2007.

24 Opstal-van Winden AW, Krop EJ, Karedal MH, Gast MC, Lindh CH, Jeppsson MC, Jonsson BA, Grobbee DE, Peeters PH, Beijnen JH, van Gils CH and Vermeulen RC: Searching for early breast cancer biomarkers by serum protein profiling of pre-diagnostic serum; a nested case–control study. BMC Cancer *11*: 381, 2011.
25 Bohm D, Keller K, Wehrwein N, Lebrecht A, Schmidt M, Kolbl H and Grus FH: Serum proteome profiling of primary breast cancer indicates a specific biomarker profile. Oncol Rep *26*: 1051-1056, 2011.
26 Pastorekova S, Parkkila S and Zavada J: Tumor-associated carbonic anhydrases and their clinical significance. Adv Clin Chem *42*: 167-216, 2006.
27 Takano S, Yoshitomi H, Togawa A, Sogawa K, Shida T, Kimura F, Shimizu H, Tomonaga T, Nomura F and Miyazaki M: Apolipoprotein C-1 maintains cell survival by preventing from apoptosis in pancreatic cancer cells. Oncogene *27*: 2810-2822, 2008.
28 Schafer MK and Altevogt P: L1CAM malfunction in the nervous system and human carcinomas. Cell Mol Life Sci *67*: 2425-2437, 2010.
29 Yang Y, Zhao S, Fan Y, Zhao F, Liu Q, Hu W, Liu D, Fan K and Wang J: Detection and identification of potential biomarkers of non-small cell lung cancer. Technol Cancer Res Treat *8*: 455-466, 2009.
30 Cohen M, Yossef R, Erez T, Kugel A, Welt M, Karpasas MM, Bones J, Rudd PM, Taieb J, Boissin H, Harats D, Noy K, Tekoah Y, Lichtenstein RG, Rubin E and Porgador A: Serum apolipoproteins C-I and C-III are reduced in stomach cancer patients: results from MALDI-based peptidome and immuno-based clinical assays. PLoS One *6*: e14540, 2011.
31 Xue A, Scarlett CJ, Chung L, Butturini G, Scarpa A, Gandy R, Wilson SR, Baxter RC and Smith RC: Discovery of serum biomarkers for pancreatic adenocarcinoma using proteomic analysis. Br J Cancer *103*: 391-400, 2010.
32 Klee EW, Bondar OP, Goodmanson MK, Dyer RB, Erdogan S, Bergstralh EJ, Bergen HR, 3rd, Sebo TJ and Klee GG: Candidate serum biomarkers for prostate adenocarcinoma identified by mRNA differences in prostate tissue and verified with protein measurements in tissue and blood. Clin Chem *58*: 599-609, 2012.

*Received July 1, 2015*
*Revised September 1, 2015*
*Accepted September 3, 2015*