Identification of candidate biomarker mass (m/z) ranges in serous ovarian adenocarcinoma using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling

Amutha Periyasamy1, Gopal Gopisetty1#, Sridevi Veluswami2, Malliga Joyimallaya Subramanium3, and Rajkumar Thangarajan1#

1Department of Molecular Oncology, 2Department of Surgical Oncology, and 3Department of Preventive Oncology, Cancer Institute (WIA), Chennai, Tamil Nadu, India

Abstract

Objective: To differentiate plasma from ovarian cancer and healthy individuals using MALDI-TOF mass spectrometry.

Materials and methods: MALDI-TOF was used to generate profiles of immuno-depleted plasma samples (89 cancers and 199 healthy individuals) that were fractionated using three types of magnetic beads (HIC8, WCX and IMAC-Cu).

Results: Differentially expressed mass ranges showing >1.5–2-fold change in expression from HIC8 (30), WCX (12) and IMAC-Cu (6) fractions were identified. Cross validation and recognition capability scores for the models indicated discrimination between the classes.

Conclusions: Spectral profiles can differentiate plasma samples of ovarian cancer patients from healthy individuals.

Keywords

HIC8, IMAC-Cu, MALDI-TOF, ovarian cancer, plasma, WCX

Introduction

Ovarian cancer is the seventh most common cancer in women and is associated with a high mortality rate worldwide (Ferlay et al., 2013). According to Madras Metropolitan Tumor Registry (MMTR) report (2006–2008) ovarian cancer is the third leading cause of cancer death in women in south India compared to other types of cancer. The crude incidence rate and age adjusted rate were 6.9 and 7.4 per 100,000 persons, respectively (Shanta & Swaminathan, 2008). The most common types of ovarian cancer include epithelial, germ cell and sex-cord stromal. Among the various types of ovarian cancer, the majority of cases (90%) diagnosed are epithelial in origin of which serous adenocarcinoma is the most common type. About 75–80% of the epithelial origin cancers are diagnosed at late stage (III/IV) and the prognosis is dismal. Despite effective treatment strategies such as cyto-reductive surgery and combination of chemotherapy, late diagnosis of the disease contributes to the high mortality and also high rate of disease relapse which in due course decreases the 5 years survival rate to ~27% (Jelovac & Armstrong, 2011; Kaku et al., 2003; SEER). Since ovarian cancer does not present clear physical symptoms consequently, surveillance and early diagnosis of the disease using biomarkers is of vital importance in increasing the survival rates for ovarian cancer.

Proteomic technologies have been previously employed in the identification of new blood based tumour markers for various cancers including ovarian cancer. However, the use of whole blood/plasma/serum for biomarker discovery has met with limited success. Plasma for instance is a complex mixture contains more than 100 thousand proteins constituted in a wide dynamic concentration range which limits direct analysis. This can be partly overcome by employing various pre-fractionation methods to reduce the complexity of samples which are later analyzed by various proteomic technologies such as 2D gel electrophoresis, surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-ToF MS), matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF MS) and electrospray ionization (Anderson & Anderson, 2002). Here, we have applied immune-depletion combined with magnetic beads based pre-fractionation strategy to enrich the low abundant proteins from plasma. The proteins/peptides profile was acquired using MALDI-TOF mass spectrometer in linear mode. Statistical software tool ClinProt was applied to compare the mass spectrum profile of ovarian cancer against healthy controls to identify the discriminating peaks in the training sample set. Data from our analysis indicate presence of biomarker mass ranges which can differentiate ovarian cancer samples from healthy controls. The classification models generated using the training set spectral profiles, were...
further validated with an independent test sample set. Based on the fold change in expression, receiver operating curve (ROC) values, and classification models generated we determined candidate biomarker masses for ovarian cancer in our patient samples.

Materials and methods

Study group and sample collection

A total of 89 patients with epithelial ovarian cancer (age 22–70, mean age 51 years) newly diagnosed and a total of 199 healthy volunteers (age 28–58, mean age 39 years) who had undergone general breast examination and cervical smear test and had no reported abnormalities were enrolled in this study. Written informed consent was collected from all participants prior to sample collection. All samples were collected before the day of chemotherapy. The sample collection was approved by the Institutional ethical committee overseeing the study.

In all cases, 5 ml of blood sample was collected in a vacutainer tube (Becton, Dickinson and Company, Franklin Lakes, NJ) and allowed to settle down the RBCs for 30 min at room temperature. The supernatant plasma was transferred to another centrifuge tube and then centrifuged at 2500 g at 4°C until analysis.

Quantitation and SDS–PAGE analysis of immuno-depleted fraction

Protein concentration in immuno-depleted plasma fraction was analysed by bicinchoninic acid (BCA) method using Pierce BCA Protein assay kit (Thermo Scientific, Waltham, MA). An aliquot of fraction was resolved using 4–12% SDS–PAGE and proteins were visualized by coomassie brilliant blue (G250) staining.

ClinProt magnetic beads pre-fractionation

The immuno-depleted plasma fraction was fractionated by chemically coated magnetic beads (MBs) (ClinProt; Bruker Daltonics, Bremen, Germany) with defined surfaces hydrophobic interaction C8 (HIC8), immobilized metal ion-Cu (IMAC-Cu) and weak cation exchange (WCX) according to the manufacturer’s protocol.

HIC8 MBs were shaken for multiple times to get a homogeneous suspension. Ten microlitres (1 μg/μl) of immune-depleted sample was mixed with 20 μl of binding buffer in 0.2 ml thin wall PCR-tube. Five microlitres of MBs was added to the mixture and incubated for 1 min. Later the tube was placed in a magnetic separator for 20 s to separate the unbound proteins in supernatant. Unbound fraction was discarded and MBs were washed thrice with 100 μl washing buffer. Finally, the bound proteins were eluted in 10 μl of elution solution (50% acetonitrile in water).

IMAC-Cu MBs were vortexed thoroughly for 1 min to get a homogenous suspension. Of these, 5 μl MBs mixed with 50 μl binding buffer in a thin wall PCR tube and placed on a magnetic separator and supernatant was removed. This step was repeated twice and MBs were resuspended in 40 μl of binding buffer. Twenty microlitres (1 μg/μl) of immune-depleted sample was added to the pre-treated MBs and incubated for 5 min. After that the tube was placed in a magnetic separator for 20 s to separate the unbound proteins in supernatant. Unbound fraction was discarded and MBs were washed thrice with 100 μl washing buffer. Finally, MBs were resuspended in 10 μl of elution solution. After 5 min, tube was placed on a magnetic separator and the supernatant was transferred into a fresh tube.

WCX MBs were resuspended thoroughly to get a homogenous suspension. Ten microlitres MBs mixed with 40 μl of binding buffer. Twenty microlitres (1 μg/μl) of immune-depleted sample was added to the mixture and incubated for 5 min. After that the tube was placed in a magnetic separator for 20 s to separate the unbound proteins into the supernatant. Unbound fraction was discarded and MBs were washed thrice with 100 μl washing buffer. Finally, MBs were resuspended in 5 μl of proprietary elution solution. After 5 min, tube was placed on a magnetic separator and bound peptides were
collected in supernatant and transferred to fresh tube containing 5 µl stabilization buffer supplied with the kit.

**MALDI-ToF MS analysis**

The profile spectrum of protein fractions was obtained using MALDI-TOF MS (UltraFlex; Bruker Daltonics). The calibration standard for the analysis comprised mixture of peptides and proteins in the mass range of 1–17 kDa. The spectrum was acquired in linear positive mode in the mass range of 1–20 kDa. To acquire MS spectrum, 1 µl of the elutes was mixed with 1 µl HCCA matrix (50% acetonitrile in 0.1% TFA water) and 1 µl of this mixture was spotted on to a MALDI target plate in four repeats. Final mass spectra were obtained by averaging 5000 laser spots collected at 10 different positions.

**Mass spectrometric data analysis**

The following settings were used for spectra preparation:
- Resolution 800 resolution, baseline subtraction Top Hat Baseline, 10% minimal baseline width, mass range 1000–20000 m/z, support spectra grouping, enable similarity selection, 1000 ppm maximal peak shift, recalibration of spectra, 30% match to calibrant peaks and excluded not recalibratable spectra and null spectra. Settings for peak calculation: Peak calculation using total average spectrum, signal to noise ratio (S/N) 5 on average spectra, peak calculation using intensities, mutation rate 0.20 and cross-over rate 0.50.

ClinProt Tools 2.2 (CPT) (Bruker Daltonics) software was applied for data analysis. CPT offers three basic workflow such as peak statistics calculation, model generation and external validation to identify the differentially expressed proteins between two classes. For peak statistic calculation spectra of two different classes (normal and ovarian tumour) were loaded in CPT. The data were subjected to standard data preparation workflow using default parameter settings. Baseline subtraction on spectra (top Hat), normalization of spectra to their total ion current, recalibration of spectra, average spectra calculation from recalibrated spectra and average peak list calculation with S/N >5 was performed. The p values from t-test, analysis of variance, Wilcoxon test and Anderson–Darling test were assessed. p Value <0.05 in the entire three tests was considered as significant. Finally, CPT generated a peak statistic table along with statistical difference between two classes was studied for potential biomarker masses. Further, fold change in expression of the markers were calculated from the average intensity of the peaks in the mass spectrum. Significant differential peaks showing greater than or equal to 1.5-fold change in expression was noted. For classification model generation algorithms present in CPT namely genetic algorithm (GA), supervised neural network (SNN) and quick classifier (QC) were used. From the generated model the cross validation (CV) and recognition capability (RC) of the training set were assessed to define the accuracy of class prediction. The external validation was performed with an independent test set in CPT.

**Results**

A total of 199 healthy subjects and 89 ovarian cancer of the serous adenocarcinoma type were analysed. In this study we used two step enrichment protocol to increase the levels of low expression proteins in plasma. The plasma sample was immuno-depleted off 14 abundant proteins using affinity chromatography and the resultant fraction was quantitated. Comparison between immuno-depleted fraction and whole plasma using SDS–PAGE analysis indicated enrichment of proteins (lane 2) not observed in undepleted plasma (lane 4) (Figure 1).

The immuno-depleted plasma proteins were further fractionated using magnetic beads coated with defined surfaces. Three different magnetic beads surface modifications were used (hydrophobic interaction, cation exchange, metal affinity). Previous studies have shown that magnetic bead based fractionation decreased pre-analytical variability in sample preparation and possessed high reproducibility (Baumann et al., 2005).

**Plasma protein profiles**

The profile spectra of plasma samples were imported into the ClinProt software. The training set (n = 100) comprised 50 confirmed ovarian serous adenocarcinoma (ADC) cases and 50 normal samples which were designated as class I and class II, respectively. After the initial data preparation the peak statistic table was generated. For each fraction, classifier models were generated using SNN, GA and QC. The default parameter settings were used of the model generation, for instance genetic algorithm, number of selected peaks for generation of model was set to not more than 5 and number of generations not more than 50. In this k-nearest neighbour was set to 3 in order to achieve significant differences in peaks of diagnostic potential in the groups analysed. In the case of SNN the number of peaks set did not exceed 25 peaks. We selected all three affinity purification methods to
generate spectrums to widen the scope of EOC biomarker identification.

**Identification of biomarker mass ranges in HIC8 fractionated sample**

Mass spectrometric profile of HIC8 affinity purified samples showed significant differences between EOC and normal samples along with minimal overlap of spectral groups (Figure 2). A total of 112 discriminating peaks were identified of which 45 peaks (30 peaks unregulated and 15 peaks down regulated in tumour) (Supplementary Table 1) had a p value less than 0.01 and greater than or equal to 1.5-fold change in expression. The candidate biomarker marker mass ranges showing greater than two-fold overexpression in tumours along with ROC values are shown in Table 1. CV and RC parameters calculated for the three models for the HIC8 fraction are listed in Table 4. Our analysis indicates high CV and RC values with GA model showing the highest CV of 94.36% and RC of 97.22% among the models. Validation was performed with an independent test set comprising of 39 ovarian serous ADC and 149 normal samples. The sensitivity and specificity for the three models generated are listed in Table 5. The results showed that all the models had higher sensitivity compared to specificity with GA model showing the highest sensitivity of 91.9% and SNN model showing highest specificity of 78.1%.

**Identification of biomarker mass ranges in WCX fractionated sample**

Mass spectrometric profile of WCX affinity purified samples showed differences between EOC and healthy samples, however, some degree of overlapping of spectra between the two groups was also present (Figure 3). A total of 103 discriminating peaks were identified of which 17 peaks (12 peaks up regulated and 5 peaks down regulated in tumours) had p value <0.01 and greater than or equal to 1.5-fold change in expression (Supplementary Table 2) and the candidate biomarker peaks showing greater than 2-fold overexpression in tumours along with ROC values are shown in Table 2. Results from model generation for the WCX fraction are listed in Table 4. Our analysis indicate high CV and RC values with GA model showing the highest CV of 89.33% and RC of 94.71% among the models. Validation was performed with an independent sample set comprising of 39 serous ovarian adenocarcinoma tumour and 149 healthy individual samples. The sensitivity and specificity for the three models generated are listed in Table 5. The results showed that among the models QC had the highest sensitivity 91% and SNN (78.8%) and QC (74.3%) models showed higher specificity.

| Peak index | Mass (m/z) | Fold change | ROC value (AUC) |
|------------|------------|-------------|-----------------|
| 6          | 1066.08    | 2.92        | 0.831           |
| 45         | 4467.05    | 2.79        | 0.829           |
| 97         | 11693.29   | 2.54        | 0.854           |
| 98         | 11735.91   | 2.45        | 0.854           |
| 80         | 8601.58    | 2.44        | 0.722           |
| 13         | 1277.19    | 2.39        | 0.795           |
| 7          | 1082.24    | 2.32        | 0.775           |
| 8          | 1087.8     | 2.26        | 0.792           |
| 14         | 1293.36    | 2.10        | 0.741           |
| 109        | 17105.23   | 2.08        | 0.863           |
Identification of biomarker mass ranges in IMAC-Cu fractionated sample

IMAC-Cu affinity purified samples also showed differences between EOC and healthy samples in 2D mass spectrum profile although the separation between the two groups of spectra was not prominent and presented with considerable overlapping (Figure 4). A total of 82 discriminating peaks were identified of which 30 peaks had p value less than 0.01 and greater than or equal to 1.5-fold change in expression (Supplementary Table 3) and peaks showing greater than 1.5 overexpression in tumour samples are listed in Table 3, peaks with >2-fold overexpression were not observed in this fraction. CV and RC parameters calculated for the three models are listed in Table 4. Comparison of their values showed highest CV for the SNN model (78.28%) and highest RC for the GA model (91.36%). Validation was performed with an independent set, for the want adequate sample for analysis the number of tumours and normals was restricted to 24 and 124 samples, respectively. The sensitivity and specificity for the three models generated are listed in Table 5. The results showed that all the models had higher sensitivity compared to specificity and QC had the highest sensitivity 85.7% and SNN model (79.8%) showed higher specificity.

Discussion

Proteomics based biomarker identification using MALDI-TOF mass spectrometry can be used to compare protein/peptide profiles of large number of samples within a short time and investigate for evidence of differentiating features between the two classes of samples. This technology has been successfully applied to study complex mixtures like plasma, urine or synovial fluid (Bosso et al., 2008; Kojima et al., 2008; Pan et al., 2012). Mass spectrometry combined with CPT system was applied in biomarker identification for various cancer such as breast, oral, head and neck cancers and ovarian cancer (Cheng et al., 2005; Fan et al., 2012; Freed et al., 2008; Qiu et al., 2009). Unlike previous studies we focused our analysis on serous adenocarcinoma of the ovary which is the most common type of EOC, this was done to
minimize sample variability and ensure a degree of homogeneity. Several studies had reported high abundant proteins like fibrinogen, complement protein and apolipoprotein as biomarker and could have potentially excluded the information about low abundant proteins (Chang et al., 2006; Cheng et al., 2005; Kojima et al., 2008; Ornellas et al., 2012). It is plausible that most of the tumour specific markers exist in plasma at low level concentrations, however identifying them is beset with limitations posed by dynamic range of the protein concentrations. One approach that has been employed is to remove the high abundant proteins which hinder the detection of low abundant proteins by using antibody based multiple affinity removal system. In this study we have used MARS column to deplete 14 high abundant proteins from plasma samples prior to subsequent fractionation using magnetic beads with defined surfaces. This was performed with an aim to increase the chance of identification of proteins/peptides present in low concentration.

Mass spectrometry has been widely used to analyze the peptides/proteins in their intact form without further fragmentation. In MALDI-TOF MS the proteins are co-crystallized with energy absorbing matrix on a plate. Pulsed laser energy is used to desorb/ionize the sample, time-of-flight affected by means of mass and charge of the ionized particle and subsequent detections of ions results in spectrum representing the masses ($m/z$) of proteins/peptides present in the sample (Lu et al., 2007). Since the complexity of the sample can directly affect the quality and reproducibility of the spectra, the selective enrichment of specific proteins according to their chemical and physical properties can improve the spectra quality significantly (Wong et al., 2010).

Table 4. CV and RC of various models generated using the test set of samples.

| Fraction | Algorithm | CV % | RC % |
|----------|-----------|------|------|
| HIC8     | SNN       | 94.36| 97.22|
|          | QC        | 84.97| 88.17|
|          | GA        | 92.27| 90.72|
| WCX      | SNN       | 86.06| 91.69|
|          | QC        | 79.44| 85.31|
|          | GA        | 89.33| 94.71|
| IMAC-Cu  | SNN       | 78.28| 78.16|
|          | QC        | 71.96| 78.84|
|          | GA        | 77.72| 91.36|

SNN, supervised neural network; QC, quick classifier; GA, genetic algorithm.

Table 5. Sensitivity and specificity of various models generated using the validation set of samples.

| Fraction | Algorithm | Sensitivity % | Specificity % |
|----------|-----------|---------------|---------------|
| HIC8     | SNN       | 85.1          | 78.1          |
|          | QC        | 86.5          | 68.1          |
|          | GA        | 91.9          | 61.3          |
| WCX      | SNN       | 77.8          | 78.8          |
|          | QC        | 91            | 69.6          |
|          | GA        | 68.1          | 74.3          |
| IMAC-Cu  | SNN       | 79.1          | 79.8          |
|          | QC        | 85.7          | 59.4          |
|          | GA        | 79.1          | 62.6          |

SNN, supervised neural network; QC, quick classifier; GA, genetic algorithm.

Figure 4. Distribution of IMAC-Cu fraction spectra from serous ovarian adenocarcinoma (crosses) and healthy control samples (small circles).
Here, we have employed the use of magnetic beads based fractionation system and MALDI mass spectrometry combined with immune-depletion to increase our chances of detection of low abundant biomarker proteins which could help in differentiation of samples. In this study, we have analyzed three different fractions in order to pick up the differential proteins from immuno-depleted plasma sample and have identified 297 Peaks differentially expressed between cancer and healthy individuals. Our analysis indicates that, of all the three fractions HIC8 yielded the highest number of differential peaks followed by WCX and IMAC-Cu. We narrowed down the mass ranges by setting the fold change in expression as criteria. The utility of the biomarker mass ranges was established by the fact that they were able to discriminate between the two classes as indicated from the high (>90%) CV and RC scores generated by the classification models which in turn are generated based on combination of these peaks. The classification models were further validated by an independent test set and some of the models had shown good discrimination capability with specificity showing >80% and sensitivity >75% indicating a degree of robustness in the expression of these peaks.

Based on the present study we identified valuable diagnostic biomarker peak mass ranges for EOC and potential diagnostic models were established which effectively differentiates EOC from healthy controls with high sensitivity and specificity. There are, however, additional studies required before taking this forward for clinical application. First, while spectra can by itself serve as a potential diagnostic marker, for widespread availability, it would be better to know the peptides contributing to the spectra. This will then help identify the proteins, which then offers the option of developing a simple assay for these proteins. In addition, large scale validation will be necessary to confirm the above findings before taking it to the clinic.

**Conclusion**

From this preliminary study, we found that MALDI-TOF based MS analysis of immuno-depleted plasma samples can potentially reveal novel biomarkers for ovarian cancer. However, increased sample size will be required to strengthen the models generated along with validation on a larger test set.

**Acknowledgements**

We would like to acknowledge Bruker Daltonics, India for their technical support in the proteomics facility.

**Declaration of interest**

The authors report no declarations of interest. This research was supported by Department of Science and Technology, Government of India [No.SR/S9/Z-08/2010 dated 25.06.2010]. We also thank DST-INSPIRE for supporting the JRF.

**References**

Anderson NL, Anderson NG. (2002). The human plasma proteome. Mol Cell Proteom 1:845–67.
Baumann S, Ceglarek U, Fiedler GM, et al. (2005). Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time of flight mass spectrometry. Clin Chem 51: 973–80.
Bozzo N, Chinello C, Picozzi SCM, et al. (2008). Human urine biomarkers of renal cell carcinoma evaluated by ClinProt. Proteom Clin Appl 2:1036–46.
Chang JT, Chen L, Wei S, et al. (2006). Increase diagnostic efficacy by combined use of fingerprint markers in mass spectrometry – plasma peptidomes from nasopharyngeal cancer patients for example. Clin Biochem 39:1144–51.
Cheng A, Chen L, Chien K, et al. (2005). Oral cancer plasma tumor marker identified with beads-based affinity-fractionated proteomic technology. Clin Chem 51:2236–44.
Fan N, Gao C, Zhao G, et al. (2012). Serum peptidome patterns of breast cancer based on magnetic beads separation and mass spectrometry analysis. Diagnos Pathol 7:45. doi: 10.1186/1746-1596-7-45.
Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: IARC CancerBase No. 11. Lyon: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr [last accessed Dec 2013].
Freed GL, Cazares LH, Fichandler CE, et al. (2008). Differential capture of serum proteins for expression profiling and biomarker discovery in pre- and posttreatment head and neck cancer samples. The Laryngoscope 118:61–8.
Jelovac D, Armstrong DK. (2011). Recent progress in the diagnosis and treatment of ovarian Cancer. CA Cancer J Clin 61:183–203.
Kaku T, Ogawa S, Kawano Y. (2003). Histological classification of ovarian cancer. Med Electron Microsc 36:9–17.
Kojima K, Asmellash S, Klug CA, et al. (2008). Applying proteomic-based biomarker tools for the accurate diagnosis of pancreatic cancer. J Gastrointest Surg 12:1683–90.
Lu M, Faull KF, Whitelegge JP, et al. (2007). Proteomics and mass spectrometry for cancer biomarker discovery. Biomarker Insights 2: 347–60.
Ornellas P, Ornellas AA, Chinello C, et al. (2012). Downregulation of C3 and C4A/B complement factor fragments in plasma from patients with squamous cell carcinoma of the penis. Int Braz J Urol 38:739–49.
Pan X, Huang L, Chen J, et al. (2012). Analysis of synovial fluid in knee joint of osteoarthritis: 5 proteome patterns of joint inflammation based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Inte Orthop (SICOT) 36:57–64.
Qiu F, Liu HY, Dong ZN, et al. (2009). Searching for potential ovarian cancer biomarkers with matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry. Am J Biomed Sci 1: 80–90.
SEER stat fact sheets. Ovary seer.cancer.gov/statfacts/html/ovary.html. [last accessed Sep 2013].
Shanta V, Swaminathan R. (2008). Population based cancer registry, Chennai, Cancer Institute (WIA), Adyar, individual registry write-up 2006–2008.
Wong MYM, Yu KOY, Poon TCW, et al. (2010). A magnetic beads-based serum proteomic fingerprinting method for parallel analytical analysis and micropreparative purification. Electrophoresis 31: 1721–30.

**Supplementary material available online**

Supplementary Tables 1–3