Journal of
Proteomics & Bioinformatics

Reyzer et al., J Proteomics Bioinform 2013, S6
DOI: 10.4172/jpb.S6-004

Research Article Open Access

MALDI MS Profiles Distinguish ER-Negative Breast Cancers from Lung Adenocarcinoma

Michelle L Reyzer[1#], Jun Won Park[2#], Jamie L Allen[1], Oleg Chertov[3], Dae Yong Kim[4], Han Sung Kang[2], Geon Kook Lee[2], Richard Caprioli[1] and Hark Kyun Kim[2*]

[1]*Mass Spectrometry Research Center, Vanderbilt University, Nashville, TN 37232-8575, USA*
[2]*National Cancer Center, Goyang, Gyeonggi 410-769, Republic of Korea*
[3]*Protein Chemistry Laboratory, SAIC-Frederick, Inc., National Cancer Institute at Frederick, Frederick, MD, 21702, USA*
[4]*College of Veterinary Medicine, Seoul National University, Seoul 151-742, Republic of Korea*
[#]*These authors equally contributed to this paper*

Abstract

Background: There is an unmet need for tissue-specific biomarkers that distinguish estrogen receptor (ER)-negative breast cancers from primary lung adenocarcinomas.

Methods: To identify proteins that differ between primary breast and lung cancers, frozen resected samples collected from 10 ER-negative breast and 18 lung adenocarcinomas patients were analyzed using Matrix-Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry (MS).

Results: MALDI MS profiles were significantly different between primary breast and lung adenocarcinomas. Importantly, 4 peaks that differentially expressed between breast and lung adenocarcinomas correctly predicted the class label of a lung metastasis of breast cancer. A peak at m/z 10,093, which was significantly overexpressed in primary lung cancer, was identified as S100A6. According to immunohistochemistry study using commercially available tissue microarray slides, S100A6 expression was significantly lower in breast cancer samples than in lung adenocarcinomas samples.

Conclusion: We identified MALDI MS profiles that may distinguish primary lung adenocarcinoma from ER-negative breast adenocarcinoma. S100A6 was identified as one of the informative peaks in the MALDI MS profiles.

Keywords: MALDI; Breast cancer; Lung cancer

Introduction

The lung is a frequent site of metastasis of breast adenocarcinoma. When a solitary pulmonary nodule is detected during the follow-up of estrogen receptor (ER)-negative breast cancer patients, series of immunohistochemical studies for the lung biopsy are often needed to make a differential diagnosis of the lung lesion. Vollmer proposed a model based on immunohistocheimcal stains for thyroid transcription factor (TTF)-1, mammaglobin, p63, and ER and epidemiologic data about primary lung and metastatic breast cancers in women [1]. This model can yield nearly certain diagnoses in approximately 80% of tumors, but TTF-1-negative lung cancers may not be distinguished from ER-negative breast cancer in small biopsy samples [1]. Therefore, there is an unmet need for tissue-specific biomarkers that distinguish primary lung adenocarcinoma from metastatic ER-negative breast cancers in the lung. We undertook a study to identify tissue-specific biomarkers to help distinguish between primary lung adenocarcinoma and breast cancer metastases in lung. Since fresh frozen tissue of the breast cancer metastasis in lung is not readily available for research, we chose to compare protein expression profiles between primary ER-negative breast cancers and primary lung adenocarcinoma, to identify tissue-specific biomarkers. This approach has been widely employed to develop a genomic predictor of tissue of origin [2], based on the assumption that expression profiles are generally similar between primary and metastatic lesions [3].

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) has been demonstrated to be useful for direct molecular profiling of common solid tumors [4]. In this approach, mass spectra are obtained from discrete locations on the tissue cryosections, aided by sinapinic acid matrix. The resulting spectra are composed primarily of singly charged ions of proteins present in the tissue at the locations sampled. In our previous study of gastric cancer, protein profiles were found to accurately classify tumor from non-tumor tissue [5]. It was previously demonstrated that MALDI MS profiles can distinguish hepatocellular carcinomas from intrahepatic cholangiocarcinomas [6]. Meding et al. [7] also reported that MALDI imaging classified 6 common tumors (thyroid gland, esophagus, stomach, colon, liver, and breast) according to primary tissue origin at the 82.74% accuracy. These promising results prompted us to test the feasibility of using direct tissue MALDI-MS to classify common solid tumors according to the tissue origin of primary tumors. Here we report MALDI MS profiles that may distinguish primary lung adenocarcinoma and ER-negative breast adenocarcinoma.

Materials and Methods

Processing clinical material and MALDI MS data acquisition

Tissues were obtained, with informed consent and institutional review board approval, from patients undergoing surgery at National Cancer Center in Korea and had been frozen in liquid nitrogen until the analysis. Samples were prepared for MALDI analysis as described previously [4]. Briefly, thin (12 μm) sections were obtained from the

***Corresponding author:** Hark K Kim, Chief, Biomolecular Function Research Branch, National Cancer Center, 323 Ilsanro, Ilsandong, Goyang, Gyeonggi 410-769, Republic of Korea, E-mail: hkim@ncc.re.kr

Received April 01, 2013; **Accepted** May 02, 2013; **Published** May 06, 2013

Citation: Reyzer ML, Park JW, Allen JL, Chertov O, Kim DY, et al. (2013) MALDI MS Profiles Distinguish ER-Negative Breast Cancers from Lung Adenocarcinoma. J Proteomics Bioinform S6: 004. doi:10.4172/jpb.S6-004

frozen tissues with a cryostat (Leica CM 3050S, Leica Microsystems Inc., Bannockburn, IL). For each tissue, two serial sections were obtained. One section was stained with hematoxylin and eosin (H&E) for visual examination. An optical image was acquired with a digital microscope slide scanner (Mirax, Budapest, Hungary). The other section was thaw-mounted onto a gold-coated stainless steel MALDI plate and washed with graded ethanol solutions (70, 90, 95% ethanol for 30 sec each) for subsequent mass spectral analysis. The optical image of the H&E-stained serial section was evaluated by pathologists, who digitally placed 200 μm diameter circles on discrete regions of interest on the tissue section. These circles were intended to cover areas of the tissue enriched with at least 75% of a particular cell type, i.e., normal epithelial cells or tumor cells. The two images (H&E section and ethanol-fixed section) were then overlayed in Photoshop (Adobe Systems Inc, San Jose, CA) in order to align features of the two serial sections. Distinct *x,y*-coordinates were obtained from each spot and imported into a robotic device for automated matrix deposition.

Matrix was deposited via an acoustic reagent multispotter (Portrait 630, LabCyte, Sunnyvale, CA) that uses focused acoustic energy to eject matrix droplets onto a target.10 Sinapinic acid (20 mg/ml, 50:50 acetonitrile:water with 0.1% trifluoroacetic acid) was used as the MALDI matrix. A total volume of ~9 nl matrix was deposited at each discrete location on tissue, resulting in dried crystal spots of ~200 μm diameter.

Mass spectra were acquired using an Autoflex Speed (Bruker Daltonics, Billerica, MA) time-of-flight (TOF) mass spectrometer equipped with a SmartBeam laser (Nd:YAG, 355 nm) and run using a linear-mode acquisition method optimized for 2-20 kDa. Data were acquired in an automated fashion from each matrix spot, with a total of 384 laser shots acquired via random walk over the entire spot for each mass spectrum.

Data processing and statistical analysis

ClinProTools (version 2.2, Bruker Daltonics) was used for baseline subtraction, spectral recalibration, and spectral peak area calculation. A resolution of 300 was applied to the peak detection method. The Top Hat baseline with 10% minimal baseline width was used for baseline subtraction. Data reduction was performed at a factor of 4. Spectra were recalibrated with a maximal peak shift of 2,000 ppm between reference and peak masses. The value of the '% Match to Calibrant Peaks' parameter was set to 20%. Spectra that were not recalibratable were excluded. All data with signal-to-noise ratios>5 were acquired, and peak area was used for peak calculation with zero level integration. An average peak list was set up for each tissue sample by picking peaks on the calculated total average spectrum for each tissue sample to create one average spectrum per patient.

Average-normalized data was subjected to statistical analysis using BRB-ArrayTools (NCI, version 4.1) [8]. A principal component analysis (PCA) plot was generated using multi-dimensional scaling analysis of BRB-ArrayTools, which graphically represents correlation coefficients among samples without forcing the samples into specific clusters. The three primary principal components were used as the axes for the 3-dimensional scaling representation. Class comparison and class prediction analyses were also performed using BRB-ArrayTools [8]. The class comparison analysis computed a Student *t*-test for each peak, and listed peaks differentially expressed among the classes at selected statistical significance level. Then, 100 random permutations of the class labels were performed. For each random permutation, all of the *t*-tests were re-computed for each peak. The class comparison tool computed the proportion of the random permutations that gave as many peaks significant at the selected level of significance as were found by comparing the true class labels. Protein profiles of the classes were considered different if this probability (designated as *Permutation P value*) was calculated to be less than 0.05.

To evaluate whether classes have different protein profiles or not, class prediction analyses were performed using all samples as a training set. The cross-validation was performed using 0.632+bootstrap cross validation method of BRB-ArrayTools [9]. The cross-validated misclassification rate was computed for all classifier functions (compound covariate predictor (CCP), diagonal linear discriminant analysis (LDA), 1- and 3-nearest neighbors (NN), nearest centroid (NC), and support vector machine (SVM)) in the training set. Then, class labels were randomly shuffled and the cross-validated misclassification rate was computed for each random dataset. Permutation *P* value, which is defined as the proportion of random datasets that give as small misclassification rate as is obtained with real class labels, was then calculated. MALDI MS profiles of the classes were considered different if this permutation *P* value was <0.05. A statistical classification model was developed to distinguish primary breast from lung cancers using half the samples and validated with the other half. Informative peaks identified in the training set were then used to predict the class label of a proof-of-concept test sample.

Protein identification

A surgically removed lung adenocarcinoma specimen was used for peak identification. Protein was extracted using phosphate-buffered saline and hypotonic saline. These extracts were further fractionated by high-performance liquid chromatography (HPLC). Aliquots of HPLC fractions were analyzed by MALDI time-of -flight (TOF) mass spectrometry (MS) using sinapinic acid as matrix on a Voyager-De Pro TOF MS instrument (PerSeptive, Framingham, MA). Lyophilized aliquots of selected fractions were further fractionated by reverse phase HPLC. LC-MS/MS analyses of the HPLC fractions (treated with trypsin) were performed, and theoretical masses of intact proteins identified in particular fractions were compared with MALDI-TOF MS data of these fractions.

Immunohistochemistry

S100A6 immunohistochemistry was performed using anti-S100A6 (A5115, DAKO, Carpinteria, CA) and commercial tissue microarray (TMA) slides (BB6, VA2, MB4, and 201(VI), Superbiochips Laboratories, Seoul, Korea; A206V and A202VII, Isu Abxis, Seoul, Korea). Immunostaining results could be obtained from 16 breast and 12 lung adenocarcinomas with Superbiochips. Also, results could be obtained from 24 breast and 18 lung adenocarcinomas with Isu Abxis. Positive staining for S100A6 was defined as staining unequivocally deeper than background. The intensity of immunoreactivity was graded on a scale from 0 to 4+. Chi-square test was used to compare S100A6 immunostaining grades.

Results

Data acquisition

Ten breast adenocarcinoma and 18 lung adenocarcinoma were used as a training set, and a breast cancer lung metastasis was used as a proof-of-concept test sample (Table 1S). Mass spectra were acquired on individual spots for each tissue section, and these spectra were averaged together after pre-processing to create one average spectrum per patient in order to minimize intra-sample variability. The average

spectra are composed of 4 to 21 individual measurements for breast cancer samples (with a median value of 6). The average spectra for primary breast and lung cancers are shown in figure 1A. Post-spectral processing identified 63 features across the entire mass range for all of the samples studied.

A PCA plot graphically demonstrates that primary breast cancers and primary lung cancers are separately clustered in an unsupervised analysis (Figure 1B). When a class comparison analysis was performed using BRB-ArrayTools, the proportion of the random permutations that gave as many significant peaks at a feature selection of $P<0.01$ as were found by comparing the true class labels (breast cancer vs lung cancer) was less than 0.023, suggesting that breast and lung cancer samples are significantly different in their protein profiles. Class prediction analysis was performed using all of 28 samples as a training set. The permutation P value for 0.632+bootstrap cross-validation misclassification rate was less than 0.05 for the most classifiers (including support vector machines, nearest centroid, 1-nearest neighbor, and 3-nearest neighbor), suggesting that breast and lung cancer samples are significantly different in their protein profiles.

Next, class prediction analysis was performed after dividing the entire set of samples into two groups based upon the chronological order of the enrollment. The first half of the samples was used as a training set to develop predictors for the tissue origin. When peaks differentially expressed between breast and lung cancers in the first half at feature selection $P<0.01$ were applied to samples in the latter half, they predicted the class labels with 71-86% accuracy. Although there was only one breast cancer metastasis sample available, we sought to test the classification power of our MALDI MS profile in this sample as a proof of concept. When 4 peaks that were differentially expressed between 10 breast and 18 lung primary adenocarcinomas at feature selection $P<0.01$ were used to predict the tissue origin of a proof-of-concept test sample, it was correctly predicted as a breast primary by all classifiers. These results warrant further larger-scaled studies to validate the predictive power of the MALDI MS profiles. Table 1 lists these 4 peaks that were differentially expressed between primary breast and lung adenocarcinomas (Table 1).

Identification of discriminatory proteins and immunohistochemistry

Discriminatory protein identification was performed using a lung adenocarcinoma tissue sample as described in Materials and Methods. A peak at m/z 10,093, which was the largest peak that was significantly overexpressed in the lung cancer, was identified as S100A6 (Figure 2A). The other three peaks could not be identified.

S100A6, the only one peak identified in the MALDI MS profile, was then tested for classification power for the tissue origin using a different platform. S100A6 immunostaining was significantly lower in breast cancer samples than in lung adenocarcinoma samples, according to 2 independent tissue microarray studies. Median S100A6 immunostain grades were 1+ and 3+ for breast and lung adenocarcinomas, respectively, with the Superbiochip tissue microarray analysis (P=0.001). Median S100A6 immunostain grades were 1+ and 3+ for breast and lung adenocarcinoma, respectively, with the Isu Abxis tissue microarray analysis (P=0.023). When all samples were dichotomized according to S100A6 immunostaining grade as S100A6-high (2+ or above) or as S100A6-low (<2+), the overall sensitivity and specificity of the S100A6 for predicting breast cancer were 78% and 83% respectively [positive and negative predictive values, 86 and 74%, respectively]. Representative immunohistochemical study results are shown in figure 2B.

Discussion

This work demonstrates that a protein profile obtained from frozen resected samples via MALDI MS can differentiate primary breast cancers from lung adenocarcinoma samples. While MALDI MS profiling studies have been performed for primary breast [10,11] and lung cancers [4], respectively, there have been no MALDI MS studies that identified differentially expressed peaks between breast and lung

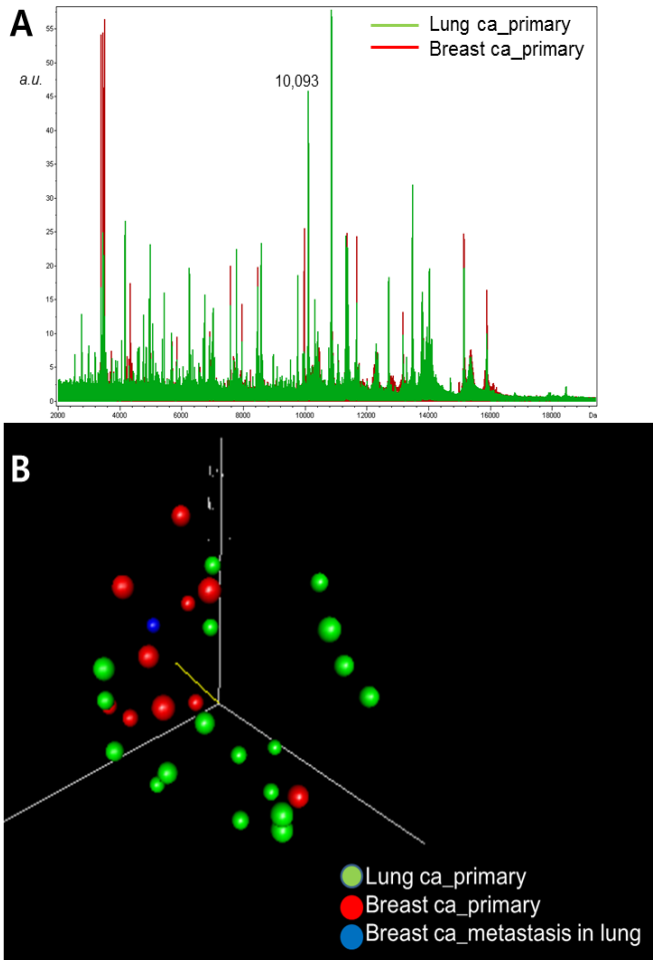

Figure 1: (A) Overlay of average mass spectra for protein obtained from primary breast (*shown in red*) and lung adenocarcinomas (*shown in green*). (B) A principal component analysis plot for 10 primary breast cancer samples (*shown in red*) and 18 primary lung adenocarcinoma samples (*shown in green*), which graphically represents Euclidean distances among samples. Each sphere represents a single sample, and samples whose protein expression profiles are very similar are shown clustered together. A breast cancer metastasis to the lung (*shown in blue*) is located close to primary breast cancers.

m/z	*P*	Peak area		Ratio[1]	Assignment
		Breast	Lung		
2,752	0.008	16	26	0.6	
6,277	0.004	29	48	0.6	
8,566	0.004	102	189	0.5	
10,093	0.008	175	353	0.5	S100A6

Ratio[1], the peak area ratio of breast to lung cancer

Table 1: Peaks differentially expressed between primary breast and lung adenocarcinomas at feature selection $P<0.01$.

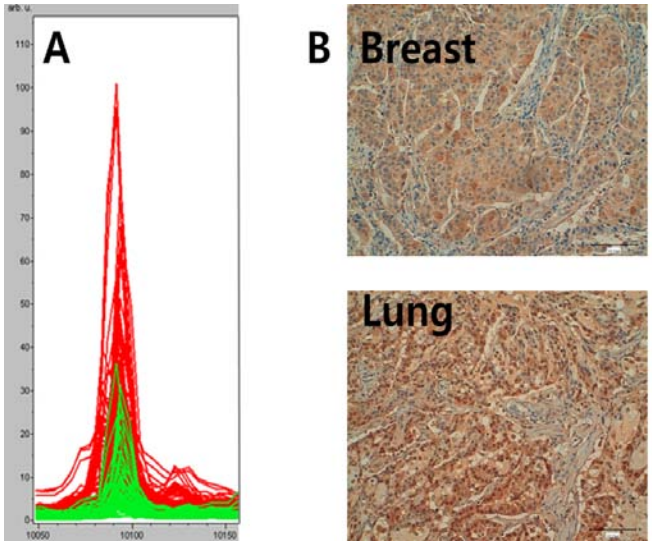

Figure 2: (A) Intensity profile for an identified discriminatory protein, S100A6, among 10 primary breast cancers (*shown in green*) and 18 primary lung adenocarcinomas (*shown in red*). (B) Representative S100A6 immunostaining for primary breast and lung adenocarcinoma samples in commercial tissue microarray slides.

adenocarcinomas, two common fatal diseases. The distinction between a primary lung adenocarcinoma and a metastatic breast cancer is critical for determining a treatment plan, but it is not uncommon to find a poorly differentiated, solitary lung lesion existing in breast cancer patients. We demonstrated the feasibility of using direct tissue MALDI MS to classify common solid tumors according to the tissue origin of primary tumors, suggesting a possible clinical utility.

In this study, S100A6 is identified as a biomarker that distinguishes primary lung adenocarcinoma and ER-negative breast adenocarcinoma. S100A6 expression is increased during the progression of lung adenocarcinoma [12], and decreased in breast cancers compared with normal mammary epithelium [13]. S100A6 functions as a Ca^{2+} sensor, and is involved in cell cycle regulation, intracellular calcium homeostasis and signaling, ion transport, exocytosis of insulin from pancreatic cells, cytoskeletal rearrangement, and ubiquitinated proteolytic degradation [14]. The functional implication of its relative abundance in lung adenocarcinoma is uncertain. While S100A6 was initially identified as a MALDI MS peak underexpressed in our ER-negative breast cancers compared with primary lung adenocarcinoma, its immunoreactivity was also found to be lower in ER-positive breast cancers than in lung cancers. Importantly, S100A6 expression was lower not only in primary breast cancers, but also in a lung metastasis, than in primary lung adenocarcinoma. Notably, however, our study is limited for the small size of validation set, which is due to the fact that metastatic breast cancers are seldom resected in clinical practice.

This article was originally published in a special issue, **Microarray Proteomics** handled by Editor(s). Dr. Qiangwei Xia, University of Wisconsin-Madison, USA

Acknowledgement

The work was supported by grants from Converging Research Center Program (2012K001506), through the Ministry of Education, Science and Technology of Korea and Proteogenomic Research Program through the National Research Foundation of Korea funded by the Korean Ministry of Education, Science and Technology.

References

1. Vollmer RT (2009) Primary Lung Cancer vs Metastatic Breast Cancer: A Probabilistic Approach. Am J Clin Pathol 132: 391-395
2. Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, et al. (2008) MicroRNAs accurately identify cancer tissue origin. Nat Biotechnol 26: 462-469.
3. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. (2000) Molecular portraits of human breast tumours. Nature 406: 747-752.
4. Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, et al. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet 362: 433-439.
5. Kim HK, Reyzer ML, Choi IJ, Kim CG, Kim HS, et al. (2010) Gastric cancer-specific protein profile identified using endoscopic biopsy samples via MALDI mass spectrometry. J Proteome Res 9: 4123-4130.
6. Jeon YE, Lee SC, Paik SS, Lee KG, Jin SY, et al. (2011) Histology-directed matrix-assisted laser desorption/ionization analysis reveals tissue origin and p53 status of primary liver cancers. Pathol Int 61: 449-455.
7. Meding S, Nitsche U, Balluff B, Elsner M, Rauser S, et al. (2012) Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging. J Proteome Res 11: 1996-2003.
8. Simon R, Lam A, Li MC, Ngan M, Menenzes S, et al. (2007) Analysis of gene expression data using BRB-ArrayTools. Cancer Inform 3: 11-17.
9. Sahiner B, Chan HP, Hadjiiski L (2008) Classifier performance prediction for computer-aided diagnosis using a limited dataset. Med Phys 35: 1559-1570.
10. Cornett DS, Mobley JA, Dias EC, Andersson M, Arteaga CL, et al. (2006) A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. Mol Cell Proteomics 5: 1975-1983.
11. Rauser S, Marquardt C, Balluff B, Deininger SO, Albers C, et al. (2010) Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. J Proteome Res 9: 1854-1863.
12. Ishii A, Suzuki M, Satomi K, Kobayashi H, Sakashita S, et al. (2009) Increased cytoplasmic S100A6 expression is associated with pulmonary adenocarcinoma progression. Pathol Int 59: 623-630.
13. Sanders ME, Dias EC, Xu BJ, Mobley JA, Billheimer D, et al. (2008) Differentiating proteomic biomarkers in breast cancer by laser capture microdissection and MALDI MS. J Proteome Res 7: 1500-1507.
14. Breen EC, Tang K (2003) Calcyclin (S100A6) regulates pulmonary fibroblast proliferation, morphology, and cytoskeletal organization in vitro. J Cell Biochem 88: 848-854.