Plasma Protein Profiles Differ Between Women Diagnosed with Cervical Intraepithelial Neoplasia (CIN) 1 and 3

Chandrika J. Piyathilake[1], Denise K. Oelschlager[2], Sreelatha Meleth[3], Edward E. Partridge[4] and William E. Grizzle[2]

[1]Department of Nutrition Sciences, [2]Department of Pathology, [3]Department of Biostatistics, [4]Department of Obstetrics & Gynecology, University of Alabama at Birmingham, Birmingham, Alabama 35294.

Abstract: Early detection of precancerous cells in the cervix and their clinical management is the main purpose of cervical cancer prevention and treatment programs. Cytological findings or testing for high risk (HR)-human papillomavirus (HPV) are inadequately sensitive for use in triage of women at high risk for cervical cancer. The current study is an exploratory study to identify candidate surface-enhanced laser desorption/ionization (SELDI) time of flight (TOF) mass spectrometry (MS) protein profiles in plasma that may distinguish cervical intraepithelial neoplasia (CIN 3) from CIN 1 among women infected with HR-HPV. We evaluated the SELDI-TOF-MS plasma protein profiles of HR-HPV positive 32 women with CIN 3 (cases) and 28 women with CIN1 (controls). Case-control status was kept blinded and triplicates of each sample and quality control plasma samples were randomized and after robotic sample preparations were run on WCX2 chips. After alignment of mass/charge (m-z values), an iterative method was used to develop a classifier on a training data set that had 28 cases and 22 controls. The classifier developed was used to classify the subjects in a test data set that has six cases and six controls. The classifier separated the cases from controls in the test set with 100% sensitivity and 100% specificity suggesting the possibility of using plasma SELDI protein profiles to identify women who are likely to have CIN 3 lesions.

Keywords: cervical, neoplasia, protein profiles, SELDI.

Introduction

Pap smear is the most commonly used diagnostic test used to identify women who are at risk of developing cervical cancer. An estimated 50 million Pap smears are performed each year in the United States (Walsh, 1998) and 6% of these are diagnosed as low-grade squamous intraepithelial lesions (LSILs) or as atypical squamous cells of undetermined significance (ASCUS) (Paavonen et al. 1990). Although there is general consensus that cytologically diagnosed high-grade squamous intraepithelial lesions (HSILs) should be evaluated by colposcopy and biopsy, there has been no consensus as to the clinical management of women diagnosed with ASCUS or LSILs. Current options include immediate colposcopy and directed biopsy or follow-up with repeat cytology every 4–6 months and colposcopy with repeated abnormal cytology. The annual cost associated with these treatment options in the U.S.A. is estimated to be over one billion dollars (Schechter, 1996). Since HR-HPV is the main risk factor for cervical cancer, recent studies have investigated the use of HPV testing in the clinical management of these lesions. The ASCUS/LSIL Triage Study for cervical cancer (ALTS) conducted by the NCI demonstrated that HPV testing is sensitive in detecting underlying precancerous lesions (CIN 2 or 3) among women with a Pap test diagnosis of ASCUS (Solomon et al. 2001), but HPV testing is not useful for women with a Pap test diagnosis of LSIL (ALTS investigators, 2000). Therefore, novel markers with higher specificity for the presence of precancerous lesions among HR-HPV positive are needed and this will improve cervical cancer screening and reduce the cost associated with patient care. As discussed below, recent advancements in proteomic research have yielded techniques to aid in novel protein biomarker identification.

Because the cause of most human diseases lies in the functional dysregulation of protein interactions, advances in proteomic technologies represent exciting ways to explore the disease processes at the molecular level. The identification, quantification, classification and functional assignment of proteins

Correspondence: Chandrika J. Piyathilake, Department of Nutrition Sciences, Division of Nutritional Biochemistry and Genomics, University of Alabama at Birmingham (UAB), 1675 University Blvd, Webb 318A, Birmingham, AL 35294. Tel: 205-975-5398; Fax: 205-966-2859; Email: piyathic@uab.edu

are essential to the full understanding of these molecular events. Since mammalian systems are much more complex than can be deciphered by their genes alone, expression analysis directly at the protein level is thought to be necessary to unravel the critical changes that occur as part of pathogenesis of a disease (Srinivas et al. 2001).

The surface-enhanced laser desorption/ionization (SELDI) time of flight (TOF) mass spectrometry (MS) is one such technique with considerable potential for the detection and quantitation of protein in a wide variety of human tissue samples (Wright et al. 1999) and appear to be an important diagnostic tool for a whole range of diseases. Preliminary applications of proteomic techniques are published in the area of cervical cancer using two-dimensional gel analysis (Bae et al. 2005) and SELDI (Wong et al. 2004). These studies demonstrate the ability of proteomics approach to distinguish cervical cancer from its normal counterpart. If this approach is also workable in the analysis of non invasive samples such as plasma, urine, exfoliated cervical cells or cervical mucous, it might potentially be used in the diagnosis of underlying precancerous cervical lesions in high-risk women. To our knowledge, there are no published reports demonstrating the usefulness of SELDI protein profiles to identify HR-HPV positive subjects who are likely to have underlying true precancerous lesions of the cervix.

The current study is an exploratory study to identify candidate SELDI-TOF MS protein profiles in plasma that may aid in distinguishing CIN 3 from CIN 1 among women positive for HR-HPV.

Materials and Methods

Samples

The plasma samples used were from a previously conducted study of cervical neoplasia (nutrition ancillary study of the ASCUS/LSIL Triage Study for cervical cancer (ALTS) where HPV status (as detected by the Hybrid Capture 2 assay-Digene Diagnostics) and cervical diagnoses were known (Piyathilake et al. 2004). Plasma samples from sixty HR-HPV positive women diagnosed with CIN 1 (n = 28) and CIN 3 (n = 32) were randomly selected from the prior study and were used in this study. A randomized template that was generated

by the statistician and that included triplicate samples from each case (CIN 3) and control (CIN 1) plasma sample was developed to ensure samples were run randomly on the system. Case-control status was kept blinded and the protein profile data generated as described below was sent directly to the statistician for analysis.

Methods

The plasma samples were prepared in triplicate and diluted 1:10 in binding buffer (100mM Ammonium Acetate pH 4.5, 0.1% Triton X-100) using the Biomek 2000 Robotics System (Beckman Coulter). The plasma samples (100 µl) were added to a pre-washed bioprocessor containing the WCX2, eight spot protein chips (Ciphergen Biosystem, Inc). The samples were incubated on a shaker for 2 hours at room temperature. The chips were washed twice with binding buffer and twice with HPLC grade H_20. The bioprocessor was disassembled and the chips were washed several times with HPLC grade H_20. Each spot was allowed to air dry and 0.5 µl of sinapinic acid (energy absorbing molecule) was added twice to each spot. The chips were read on the Ciphergen Protein Chip Reader Series PBS II using the following parameters: Molecular Weight range–1000 to 15,000, High Mass–20,000, Intensity–180, and Sensitivity–7.

Statistical Analysis

Pre-processing of data and creating training and test sets

The mean spectrum for an individual was obtained by averaging across the three replicate spectra for that individual. The m-z values for the CIN 3 group were slightly different from the CIN 1 group, but these differences were within the expected margin of error (0.2%) of the instrument. The average m-z values of both groups at each point were used as the estimated m-z value. The error of +/– 0.2% was used to develop a new variable index, which was the upper limit of the m-z interval within this error range. A total of 530 index variables were used to represent the 10,300 m-z values. Peaks were defined as the maximum intensity value within each set of m-z values represented by the new variable index. Each group had 530 peaks. The data sets were then randomly divided into

training and test sets. Six patients from each group were randomly selected to create the test set. Therefore, the training set consisted of 28 CIN 3 and 22 CIN 1.

Principal component analysis (PCA)

PCA plot which places the components in terms of the proportion of variance in the data that is attributable to the particular combination was used to visually demonstrate the differences in the groups before building a classifier in order to assess the possibility of obtaining a good classifier. The PCA components are orthogonal to each other. Therefore, the first principal component accounts for the largest amount of variability and the second component accounts for the largest amount of variability unaccounted for by the first. The orthogonality of the components allows their representation as axes in a two-dimensional plot. If the variability explained by the components is the variability associated with the differences in the two groups (CIN 3 and CIN 1), and not the random noise, the principal component plot will visually demonstrate the differences between the groups.

Building a classifier

An iterative process was implemented to develop a classifier. At each iteration, a random sample of 18 CIN 1 and 22 CIN 3 was selected from the training data set. The peaks that were statistically significantly different based on a Wilcoxon Rank Sum test were then subjected to a stepwise discriminant analysis. This procedure produced a list of index values at which the intensities were most likely to separate the two groups. This list of possible discriminators was stored. The entire process was repeated 300 times in this preliminary analysis. Index values that occurred in at least 270 of the 300 (90%) lists were selected to put into the classifier. This error rate of the classifier was estimated by using the classifier to classify the groups in the test data set.

Results

The PCA was based on the entire set of 530 index values. The first two components explained 37 percent of the variability in the data. As shown in Figure 1, the PCA plot between CIN 3 and CIN using these first two components suggested that it is possible to obtain a good separation between the groups. There were five regions in the spectra that occurred in 90% of the lists produced by the stepwise discriminant procedure. These consisted of the region between m_z 2261–2282b (index 28, 29), m_z 2371–2392 (index, 38), and m_z 2426–2437 (index 43, 44). As shown in Figure 2, a subset of these five m_z regions i.e. m_z 2426–2437 groups provided 100% separation between the groups.
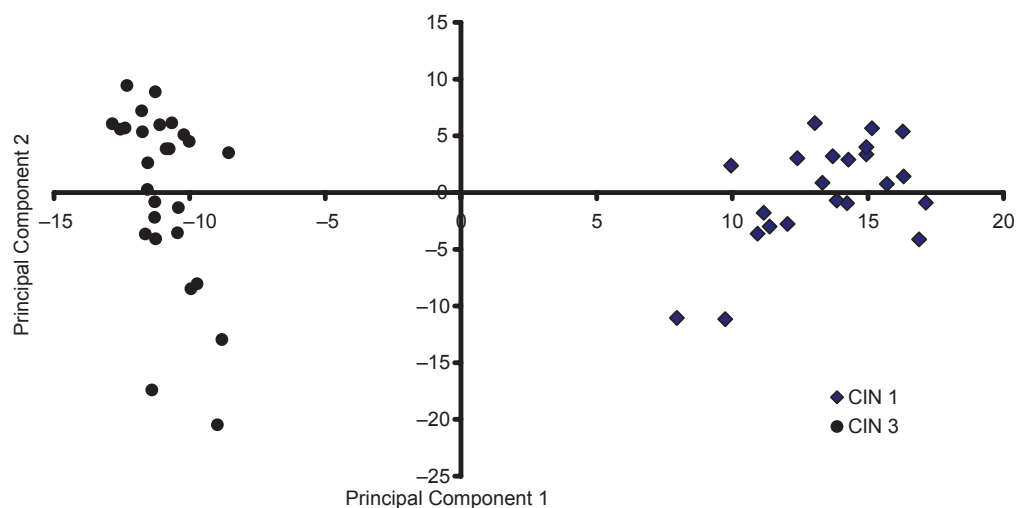

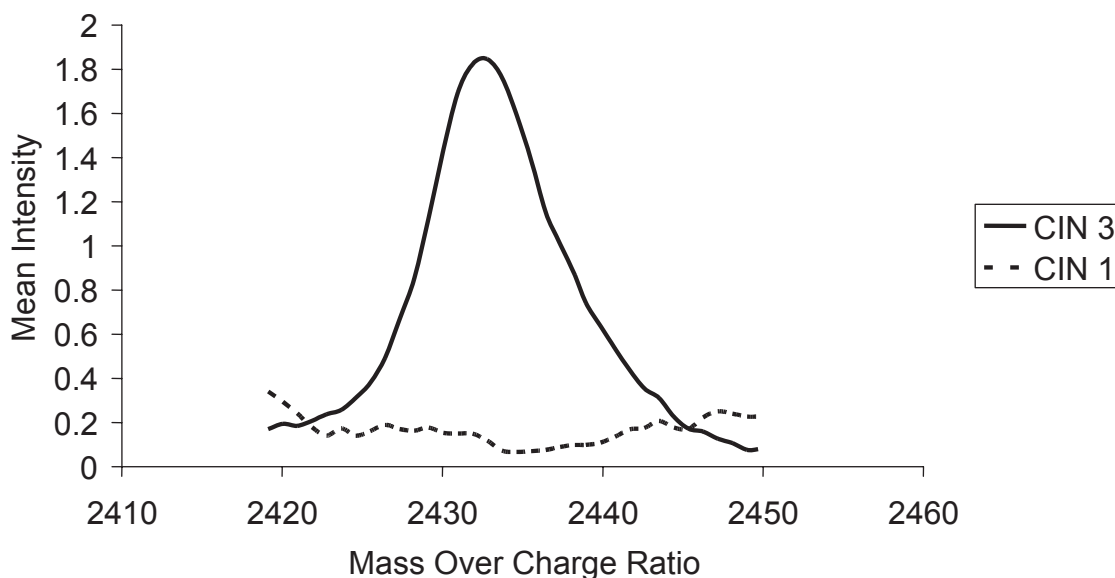**Figure 1.** PCA plot between CIN 3 and CIN 1.

Figure 2. The region of the spectra of CIN 3 and CIN 1 groups that provided 100% separation between the groups.

Discussion

An important application of SELDI is the simultaneous analysis of multiple proteins to establish "fingerprint" profiles that discriminate disease from non-disease. This is an important approach since no single biomarker or protein by itself will improve the early detection/diagnosis of diseases including cancer or pre-cancer. Protein based approaches are utilized to study the natural history and treatment of several cancers including ovarian (Lin et al. 2006), breast (Wulfkuhle et al. 2001), colon/colorectal (Lawrie et al. 2001; Ward et al. 2006; Engwegen et al. 2006) prostate (Waghray et al. 2001), head and neck (Roesch-Ely et al. 2006), endometrial (Zhu et al. 2006), nasopharynx (Ho et al. 2006), bladder (Langbein et al. 2006), hepatocellular (Ward et al. 2006) and gastric cancer (Liang et al. 2006). Protein profiling of body fluids such as nipple aspirate fluid (NAF) has shown to be rapid, reproducible, and capable of identifying protein signatures that appear to differentiate NAF samples from breast cancer patients and healthy controls, including those with an abnormal mammogram who were later proven to be biopsy normal (Paweletz et al. 2001). The use of SELDI protein profiles to detect putative breast cancer markers in peripheral samples such as saliva has also been reported in recent studies (Streckfus et al. 2006). A few studies have reported on the ability of serum protein profiles to separate benign hyperplasia from cancer and non-cancer tissues (Qu et al. 2002). Although in virology, much more limited use has been made of proteomics, use of several proteomic approaches for the analysis of HIV-cellular receptor interactions, the molecular mechanisms of transport of herpes simplex virus within neurons, and the structure of the tegument of herpes simplex virus is beginning to be appreciated (Bernhard et al. 2005).

Our preliminary study demonstrated that plasma protein profiles are significantly different between women diagnosed with CIN 1 and CIN 3. Biologically, it is plausible to see difference in protein profiles between HPV positive women who develop CIN 3 and HPV positive women who don't develop CIN 3 because several biological pathways related to HPV (example: HPV integration which may result in development of CIN 3 lesions) could be different between HPV positive women who develop CIN 3 and who don't. Cervical epithelial cells exposed to HPV which undergo transformation to pre-neoplastic lesions such as CIN 3 are very likely to differentially express signaling molecules responsible for host responses following HPV infection that resulted in CIN 3 and these molecules are likely to be present in the microenvironment of cervical cells and in the circulation. Identification of such molecules will have enor-

mous clinical applications. Triage of HPV positive women by protein biomarkers could yield enormous cost savings and other important benefits include less invasive sample collection and inconvenience regarding referral.

These results, however, have to be considered with caution. SELDI-TOF-MS analysis is a very sensitive method for determining differences in protein spectra of complex samples including plasma and serum. Because of this sensitivity, the method is very susceptible to the effects of various biases. Since cases of CIN 3 and of CIN 1 were selected randomly from participants from the ALTS Nutrition Ancillary Study, we know of no biases which separate the samples of CIN 3 from those of CIN 1. Similarly, because the triplicate aliquots of the samples were run randomly, there should be no bias in the analysis. Before final conclusions can be drawn, however, it is necessary to further evaluate this classification model by using an independent replication set, which will allow more precise conclusions about the accuracy of this classification. These pilot results also warrant an expanded study using SELDI-TOF-MS analysis to evaluate samples of plasma and other samples (cervical cells and cervical mucous) from patients with a wide range of cervical lesions to determine if these can be classified successfully. A study is currently underway at the University of Alabama (U54 CA118948) to generate this data.

References

ALTS investigators. 2000. Human Papillomavirus Testing for Triage of Women With Cytologic Evidence of Low-Grade Squamous Intraepithelial Lesions: Baseline Data From a Randomized Trial. The ASCUS/LSIL Study (ALTS) Group. *JNCI*, 92(5):397–402.

Bae, S.M., Lee, C.H. and Cho, Y.L. et al. 2005. Two-dimensional gel analysis of protein expression profile in squamous cervical cancer patients. *Gynecol. Oncol.*, 99(1):26–35.

Bernard, O.K., Diefenbach, R.D. and Cunningham, L. 2005. New insights into viral structure and virus cell interaction through proteomics. *Expert Review of Proteomics*, 2:577–588.

Engwegen, J.Y., Helgason, H.H. and Cats, A. et al. 2006. Identification of serum proteins discriminating colorectal cancer patients and healthy controls using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. *World J. Gastroenterol.*, 12(10):1536–44.

Ho, D.W., Yang, Z.F. and Wong, B.Y. et al. 2006. Surface enhanced laser desorption/ionization time-of-flight mass spectrometry serum protein profiling to identify nasopharyngeal carcinoma. *Cancer*, 107(1): 99–107.

Langbein, S., Lehmann, J. and Harder, A. et al. 2006. Protein profiling of bladder cancer using the 2D-PAGE and Seldi-TOF-MS technique. *Technol. Cancer Res., Treat*, 5(1):67–72.

Lawrie, L.C., Curran, S. and McLeod, H.L. et al. 2001. Application of laser capture microdissection and proteomics in colon cancer. *Mol. Pathol.*, 54(4):253–8.

Liang, Y., Fang, M. and Li, J. et al. 2006. Serum Proteomic patterns for gastric lesions as revealed by SELDI mass spectrometry. *Exp. Mol. Pathol.*, 81:176–180.

Lin, Y.W., Lin, C.Y. and Lai, H.C. et al. 2006. Plasma Proteomic pattern as biomarkers for ovarian cancer. *Int. J. Gynecol. Cancer*, 16 Suppl 1:139–46.

Walsh, J.M. 1998. Cervical cancer: developments in screening and evaluation of the abnormal Pap smear. *The Western Journal Of Medicine*, 169:304–310.

Paavaonen, J., Koutsky, L.A. and Kiviat, N.B. 1990. Cervical neoplasia and other STD-related genital and ana neoplasias. In: Holmes, K.K., Mardh, P.A., Sparling, P.F., Wiesner, P.J., editors. *Sexually transmitted disease.* 2nd ed. New York (NY): McGraw-Hill; p. 561–92.

Paweletz, C.P., Trock, B. and Pennanen, M. et al. 2001. Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: Potential for new biomarkers to aid in the diagnosis of breast cancer. *DIS Markers*, 17 (4):301–7.

Piyathilake, C.J., Henao, O.L. and Macaluso, M. et al. 2004. Folate is associated with the natural history of high-risk human papillomaviruses. *Cancer Res.*, 64(23):8788–93.

Qu, Y., Adam, B.L. and Yasui, Y. et al. 2002. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from non-cancer patients. *Clin. Chem.*, 48(10):1835–43.

Roesch-Ely, M., Nees, M. and Karsai S. et al. 2006. Proteomic analysis reveals successive aberrations in protein expression from healthy mucosa to invasive head and neck cancer. Oncogene, advance online publication 3:DOI:10.1038/sj.onc.1209770.

Schechter, C.B., 1996. Cost-effectiveness of re-screening conventionally prepared cervical smears by PAPNET testing. *Acta. Cytol.*, 40(6): 1272–82.

Solomon, D., Schiffman, M. and Tarone, R., for the ALTS Group. 2001. Comparison of Three Management Strategies for Patients with Atypical Squamous Cells of Undetermined Significance. Baseline Results from a randomized trial. *JNCI*, 92(12):293–299.

Srinivas, P.R., Srivastava, S. and Hanash, S. et al. 2001. Proteomics in early detection of cancer. *Clin. Chem.*, 47:1901–11.

Streckfus, C.F., Bigler, L.R. and Zwick, M. 2006. The use of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry to detect putative breast cancer markers in saliva: a feasibility study. *J. Oral Pathol. Med.*, 35(5):292–300.

Waghray, A., Feroze, F. and Schober, M.S. et al. 2001. Identification of androgen-regulated genes in the prostate cancer cell line LNCaP by serial analysis of gene expression and proteomic analysis. *Proteomics*, 10:1327–38.

Ward, D.G., Cheng, Y. and N'Kontchou, G. et al. 2006. Changes in the serum proteome associated with the development of hepatocellular carcinoma in hepatitis C-related cirrhosis. *Br. J. Cancer*, 94(2):87–92.

Ward, D.G., Suggett, N. and Cheng, Y. et al. 2006. Identification of serum biomarkers for colon cancer by proteomic analysis. *Br. J. Cancer*, 94 (12):1898–905.

Wong, Y.F., Cheung, T.H. and Lo, K.W. et al. 2004. Protein profiling of cervical cancer by protein-biochips: proteomic scoring to discriminate cervical cancer from normal cervix. *Cancer Lett.*, 211(2):227–34.

Wright, Jr., G.L., Cazares, L.H. and Leung, S.M. et al. 1999. Proteinchip surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures. *Prostate Cancer and Prostatic Diseases,* 2:264–276.

Wulfkuhle, J.D., McLean, K.C. and Paweletz, C.P. et al. 2001. New approaches to proteomic analysis of breast cancer. *Proteomics*, 10: 1205–15.

Zhu, L.R., Zhang, W.Y. and Yu, L. et al. 2006. Serum proteomic features for detection of endometrial cancer. *Int. J. Gynecol. Cancer*, 16(3): 1374–8.