academicJournals

Vol. 13(20), pp. 2084-2089, 14 May, 2014
DOI: 10.5897/AJB2013.13473
Article Number: 6D2245244587
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Application of bioinformatics to optimization of serum proteome in oral leukoplakia and oral squamous cell carcinoma

Hong He[1], Gang Sun[1], Feiyun Ping[2*], Zhijian Xie[1], Huiming Wang[1*], Yining Li[1] and Kai Zhang[1]

[1]Stomatology Hospital, School of Medicine, Zhejiang University. 395 Yan an Rd, Hangzhou, 310006, China.
[2]Second Hospital Affiliated to Medicine College of Zhejiang University 88 Jie fang Rd, Hangzhou 310009, China.

Received 15 November, 2013; Accepted 2 May, 2014

Intervention to the evolution and metastasis of oral squamous cell carcinoma (OSCC) from oral leukoplakia (OLK) out of normal mucosa is currently far from ultimate. New technologies and insights are reckoned on for etiology and clinical protocol of OSCC and mucosa lesions. SELDI-TOF-MS technology, Support vector machine, Discriminate Analysis, and CM10 protein chip were applied to study the sera proteomes of 32 healthy volunteers, 6 patients with oral mucosa leukoplakia, 28 OSCC patients, and 8 patients with metastatic OSCC. Ratios of protein mass to its charge (m/z) showed in valve peak value were delivered and discriminated out of the huge amount of protein data as group markers for identifications. Protein peak values 4181 and 4651 were high in volunteers serum while low in patients with OLK, the sensitivity of which was 100.00% (32/32), specificity was 83.33%(5/6), accuracy was 97.37%(37/38). And m/z 4162, 6886 of 87.82, 92.86 and 66.67%; 4289, 5661, 6195, 4352, 5072 of 97.22, 100.00 and 87.5% were discriminates between OLK and local OSCC, between local OSCC and regionally metastatic OSCC, respectively. Conclusively, researches are encouraged to launch a proteomics assistance and guidance in modern molecular diagnostic approaches for understanding and controlling the mucosa lesions especially in conquering the malignant progress.

Key words: Oral squamous cell carcinoma (OSCC); oral leukoplakia (OLK), ZUCI-PDAS (Zhejiang University Cancer Institute ProteinChip Data Analysis System), bioinformatics technology, discriminate analysis, proteomes of optimization.

INTRODUCTION

Mucosa lesions have never ceased threatening the human's health even with the development of modern science and technology. Oral leukoplakia (OLK) is the first one among the precancerous oral mucosal lesions, with a canceration rate of 7 to 15%, and 14 to 50% in non-homogeneous leukoplakia. Although, studies have

*Corresponding authors. E-mail: honghehh@zju.edu.cn. Tel: +86-571-87217431. Fax: +86-571-87217423.

Abbreviations: Oral squamous cell carcinoma (**OSCC**), oral leukoplakia (**OLK**)

illuminated that there exist in biomarkers as cytokines, genes, and antigens in the precancerous lesions to escape the failure in intervening the marching of malignance and metastasis of tumor. The reality is that 80% of oral malignant tumors, which rank the sixth in the queue of human malignant tumors, are the oral squamous cell carcinoma (OSCC), whilst the deformation and recurrence rate after surgery treatment are high and the prognosis is out of anticipation (Chen et al., 2004; Ferlito et al., 2002). To search for any meritorious biomarkers to trace the lesions evolution, scientists found that in the mass spectrometer, samples could be respired by laser irradiation and dissociated into gasified ions, in the uniform electric field conditions these ions perform an accelerative out fly, fly through a vacuum tube with no electric field, and eventually captured by the ion detecting receiver. The square of the ion flight time is in inverse proportion to the ratio of molecule size and molecular charge, when the numbers of charge (ions evoked by Surface Enhanced Laser Desorption/Ionization SELDI (He et al., 2009; He et al., 2011; Hu et al., 2005) are usually a single charge are equal, the time of flight associates with the molecule size, the smaller the molecule is, the shorter the flight time would be, hence small molecules were first to reach the receiver. If reflected in a mass spectrum diagram, the peak position corresponds to molecule weight of the relevant protein compositions; and the spectral peak height corresponds to the number of the relevant protein compositions. Thereby proteins could be identified with the peak position and height. This is what the protein fingerprint is. This research is aimed to obtain a sequence of fingerprints of serum protein between OLK and OSCC.

MATERIALS AND METHODS

Serum sample collection and processing

Under informed consent, sera samples of patients and healthy volunteers from the Affiliated Stomatology Hospital and 2nd Hospital of Zhejiang University during Feb 2010 to March 2014 were obtained before any treatment was implemented, and collected in the early morning before breakfast, then immediately separated and stored at -80°C until use, with diagnoses confirmed by post surgical pathology. There were 19 males and 9 females in the 28 cases of local OSCC (L. OSCC) group with a median age as 58.6 years (35 to 88 years range). Four males and 2 females with a median age as 53 years (44 to 59 years range) constituted the OLK group. Seven males and 1 female with a median age as 56.4 years (37 to 71 years range) constituted the regional metastasis group (R. OSCC). 22 males and 10 female with a median age as 50.9 years (34 to 71 years range) constituted the volunteers group (N.).

Proteins bind chip and spectrometer detection

After thawing and 2 min of centrifugation (10,000 r/min), 5 µl serum sample was added into 10 µl 0.5% U9 (9 mol/L urea, 0.2% CHAPS (3[(3-cholamidopropyl) dimethylammonio]-1-propane sulfonate), 0.1% DTT (DL-dithiothreitol) in a 96-well plate and incubated for 30 min at 4°C with 600 r/min vigorous shaking. The ProteinChip array cassette was put into a 96-well bioprocessor and 200 µl NaAc (50 mmol/L, pH 4.0) was put into each well, also incubated for 5 min at 4°C with 600 r/min vigorous shaking. Supernatant was collected and the procedure is repeated once. Then, 185 µl NaAc was added into each well in the 96-well plate (600 r/min, 2 min) and 100 µl samples of different patients disposed as above were separately added into different well of the ProteinChip array cassette (600 r/min, 1 h). After the content from each well was removed, each well was washed with 200 µl NaAc (pH = 4.0, 600 r/min, 5 min). This procedure was repeated two more times. Each spot was washed with 200 µl HPLC water, which was removed immediately. This was repeated once. After natural air drying, 1 µl SPA (sinapic acid) was applied to each spot. After natural drying for another 5 min, another 1 µl SPA was applied. Naturally dried again.

American Ciphergen SELDI Protein Biology System II plus (PBS II plus) and ProteinChip Software (Version 3.0, Ciphergen Biosystems) were used to read the chips and analyze the data. Setting parameters: laser intensity 165, 65 laser shots per sample, detector sensitivity 7, automatically detected peaks from 2,000 to 30,000 *m*/*z*. Mass accuracy was calibrated to less than 0.1% using the All-in-1 peptide molecular mass standard chip (Ciphergen Biosystems). The peaks were normalized and noises were filtrated (first Signal to Noise ratio >2.5). Peak clusters were completed using second-pass peak selection (Signal to Noise ratio>4, within 0.3% mass window) and estimated peaks were added. Discrepant mass peaks of different groups were identified by support vector machines (SVM) applying radially based kernel function with Gamma value as 0.6, Penalty score function C as 19, each m/z peak was proved with Wilconxon rank test ($p < 0.05$).

Database analysis and statistical validation

Bioinformatics studies were integrated in the ZUCI-PDAS (Zhejiang University Cancer Institute ProteinChip Data Analysis System) available at www.zlzx.net. Samples of group models from different stage were developed and validated by SVM, discriminate analysis and time-sequence analysis. These statistical analysis tools were implemented by Matlab-nn Tools software. Training was conducted to converge on the training data and to minimize the biases. Discriminate analysis and SVM models introduced random perturbations in multiple runs to test the consistency of the top 10 ranked peaks, measured by the *P* value of *m*/*z* peaks of computed ranks from multiple runs. Stage models were built using the selected peaks. Moreover, leave-one-out cross-validation approach was applied to estimate the accuracy of the classifier to determine the misclassification rate. For each step of the cross-validation, one sample was left out. The possibility of obtaining a small cross-validated misclassification rate by chance was obtained by repeating the entire cross-validation procedure using *n* random permutations of the class labels for the clinical criteria being evaluated. The ultimate candidate biomarkers of the highest Youden's index are selected out during group validation. The models established are based on these selected biomarkers should be further validated independently. In such studies, validation datasets should be preferably derived from sources different from that of the original training dataset. This is one way to ensure that the performance of the selected biomarkers is not influenced by systematic biases between different groups. Time-sequence analysis was used to distinguish different stage groups.

RESULTS

After the tremendous comparison data from the mass spectrums, 112, 108, 120 discrepant protein mass peaks between sera of 32 healthy volunteers and 6 OLK patients,

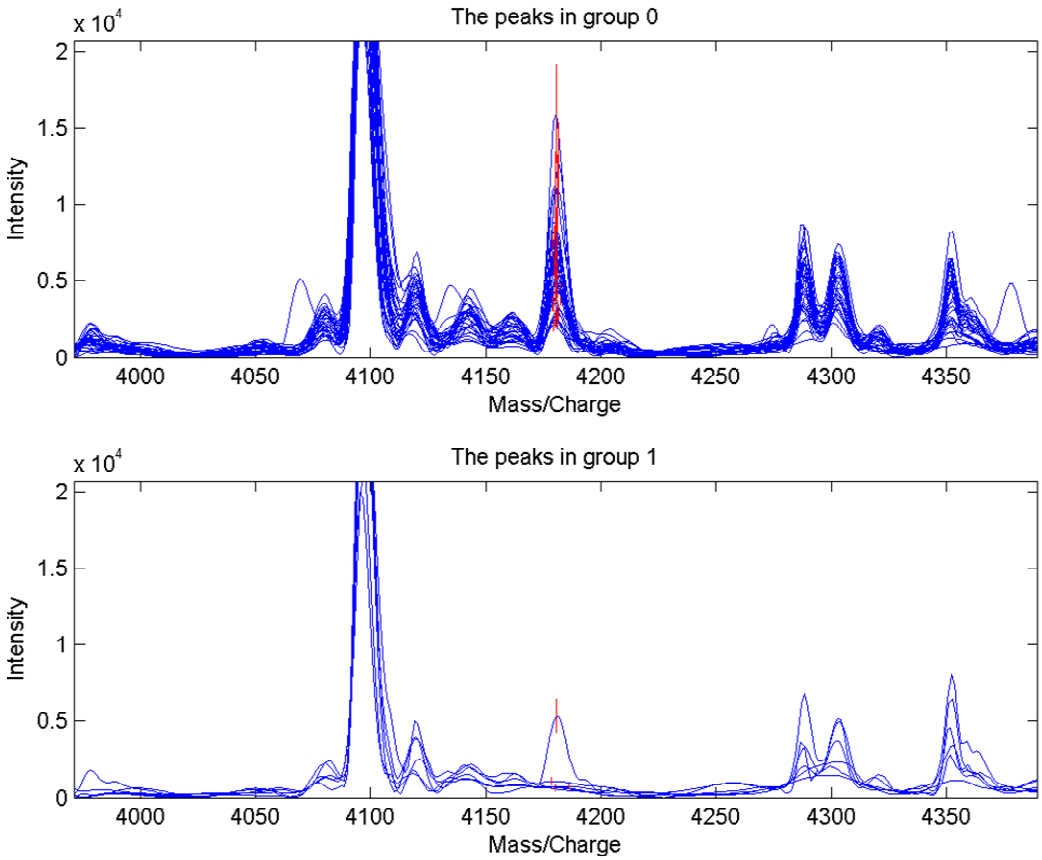

Figure 1a. Serum protein mass/charge wave peak *4181* expressed significantly high in the healthy volunteers group (group 0), while remarkably low in the group of patients with OLK (group 1), *P*=0.0001307483.

28 OSCC patients, 28 locally sited OSCC patients and 8 regionally metastatic OSCC were respectively exhibited as significant difference, among which ten of the lowest *p* value of the integrated proteomes constituted respectively of 2, 2, 5 protein peaks (Figure 1a to c, Tables 1 to 4) were eventually selected out as the highest Youden's index via SVM Discriminate Analysis and as common key sense upon corresponding and simultaneous changes. They were 4181 mutually together with 4651; 4162 together with 6886; 6195, 5661, 4289~~ and 4352 together with 5072. These ultimately obtained individual protein peaks and integrated discrepant serum proteomic selections with sensitivity, specificity, and accuracy and with some of the peak figures were presented as shown in Figures 1a to c.

DISCUSSION

The expression profiling of serum proteins from human with OLK or OSCC and healthy volunteers was investigated by a combination of proteomic techniques in this study. Different to Panicker's research (Kordy et al., 2012), which is currently lack of diseases links such as to carcinoma with metastasis or not, and in some cases to pre-invasive disease, and although our study is temporally lack of protein identification, our present research can be used to further some understanding of the serum proteomic behaving from individual to universal of oral mucosal lesion for tumorgenesis etiology as well as tumor evolution, and provide a snapshot of the proteome for diagnosis, prognosis and prophylaxis of oral mucosal disease. Validated protein biomarkers could be useful in early detection of disease and any possibility to be getting malignant or metastatic, monitoring disease progression or monitoring response to treatment. Upon neoplasia arises, it is likely to study proteins produced by the local lesion as well as by the host reaction in serum with response to the lesion. Advances in proteomic technologies have greatly accelerated the field of protein biomarker discovery (Nemet et al., 2005; Panicker et al., 2009; 2010). High-throughput technology Surface Enhanced Laser Desorption and Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS) has achieved tremendous harvests from different cancerous field to infectious diseases or drugs for researches (Sun et al.,

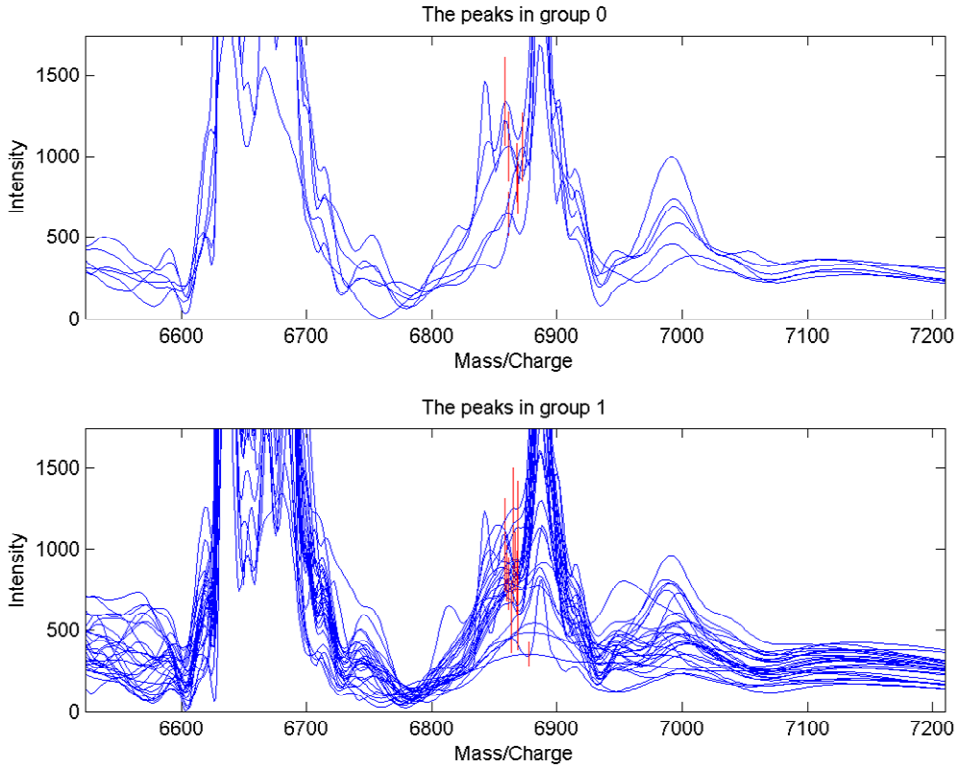

Figure 1b. Serum protein mass/charge wave peak 6886 expressed statistically different between the group of patients with OLK (group 0) and the group of patients with OSCC (group 1), *P*=0.0015641816, together with m/z 4162 protein, established the distinguished proteome with a highest Youden's index.

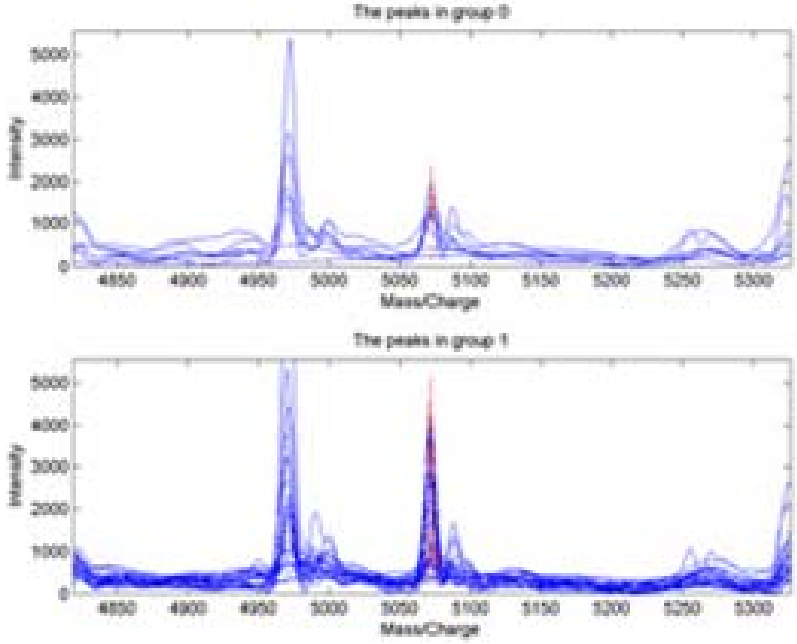

Figure 1c. Serum protein mass/charge wave peak 5072 expressed significantly low in the group of patients with metastasis OSCC (group 0), while remarkably high in group of patients with local OSCC tumor (group 1), although *P*=0.1529314178, it was one of the distinguished proteome with a highest Youden's index between the two compared groups.

Table 1. Discrepant protein peaks between OLK patients and healthy volunteers (x±SD).

m/z Peak	*p* value	Healthy volunteer (group 0)	OLK patient (group 1)
4651	0.0000086626	1905.30±548.37	686.69±386.12
4181	0.0001307483	5901.24±2548.82	1333.55±1042.86

Table 2. Discrepant protein peaks between OLK patients and local OSCC patients (x±SD).

m/z Peak	*p* value	OLK patient (0)	Local OSCC patient (1)
6886	0.0015641816	3551.45±1833.51	1818.32±921.62
4162	0.0365123759	1101.78±373.05	1644.09±579.45

Table 3. Discrepant protein peaks between local OSCC patients and regional metastatic OSCC patients (x±SD).

m/z Peak	*p* value	R. metastatic OSCC (1)	Local OSCC patient (0)
6195	0.0015387984	893.32±382.21	578.99±165.48
5661	0.0061131375	771.95±255.40	1167.44±355.59
4288	0.0289773483	1865.87±772.37	2891.56±1196.07
4352	0.0540207442	3663.07±946.03	2708.56±1249.09
5072	0.1529314178	1010.51±384.44	1518.43±952.59

Table 4. The sensitivities, specificities and accuracies of integrated discrepant serum proteomes in OLK, local (L) and regional (R) metastatic OSCC patients, and healthy volunteers.

Sample (0 vs 1)	m/z Peaks proteome	Sensitivity (%)	Specificity (%)	Accuracy (%)
N. vs OLK	*4181*，*4651*	97.37	83.33	100.00
OLK vs L. OSCC	4162, 6886	87.82	92.86	66.67
L. vs R. OSCC	4289, 5661, *6195*, *4352*, 5072	97.22	100.00	87.5

Note: Italic values of protein peak represent high expressions in group 0; regular scripts represent high expression in group 1.

2009; Weinberger et al., 2000; Xu et al., 2006; Yu et al., 2004; Yu et al., 2005; Zhou et al., 2013). Of the current proteomic tools, no single method can resolve an entire proteome. Combination of several methods like Zhejiang University Cancer Institute ProteinChip Data Analysis System (ZUCI-PDAS) with SELDI-TOF-MS (He et al., 2009; 2011; Hu et al., 2005) and SVM established the Bioinformatics. One of the merits of this study protocol is the Discriminate Analysis derived from ZUCI-PDAS and SVM has launched the other merit of it to link to the development of diseases for the biological meaning of the different preferential expressions (He et al., 2009; 2011; Hu et al., 2005).

On the one hand, SVM technique is proper to solve the limitation of samples, and on the other hand, our future work still would take into consideration to enlarge the samples as well as about the necessity to identify the proteins in the optimized group peaks. Current study indicates that progressive study of such prognostic biomarkers which were based on tumor phenotype and biologic behavior, no matter proteomes of protein mass peaks or characterization of individual protein, would allow clinicians not only to diagnose a disease involving OSCC as well as precancerous lesion like OLK, but also to select the most efficient treatment modalities other than absolutely radical surgery.

ACKNOWLEDGEMENTS

This work is funded by the Projects and funds sponsored by Scientific Research Foundation for the Returned Overseas Chinese Scholars, HR Office of Zhejiang Province (J20120036), Qianjiang Talent Scheme of

Zhejiang Province (2012R10049), China Scholarship Council (2009]3004), Youth Program of Natural Science Foundation of China (30901686), Natural Science Foundation of Zhejiang Province (LY13H140006), Research Program 2013KYB247 from Hygiene Bureau, Research Program Y201432699 and Y201225136 from Education Bureau, Zhejiang Province, China.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

Chen YD, Zheng S, Yu JK (2004). Artificial neural networks analysis of surface-enhanced laser desorption/ionization mass spectra of serum protein pattern distinguishes colorectal cancer from healthy population. Clin. Cancer Res. 10(24): 8380-5.

Ferlito A, Shaha AR, Rinaldo A (2002). The incidence of lymph node micrometastases in patients with pathologically staged N0 in cancer of oral cavity and oropharynx. Oral Oncol. 38: 3–5.

He H, Sun G, Ping FY (2009). Laser-Capture Microdissection and Protein Extraction for Protein Fingerprint of OSCC and OLK. Artif. Cells Blood Substit. Immobil. Biotechnol. 37(5): 208-213.

He H, Sun G, Ping FY, Cong Y (2011). A new and preliminary three-dimensional perspective: proteomes of optimization between OSCC and OLK. Artif. Cells Blood Substit. Immobil. Biotechnol. 39(1):26-30.

Hu Y, Zhang S, Yu J, Zheng S (2005). SELDI-TOF-MS: the proteomics and bioinformatics approaches in the diagnosis of breast cancer. Breast 14(4): 250-5.

Kordy HM, Baygi MH, Moradi MH (2012). A hybrid feature subset selection algorithm for analysis of high correlation proteomic data. J. Med. Signals Sens. 2(3):161-8.

Nemet Z, Velich N, Bogdan S (2005). The prognostic role of clinical, morphological and molecular markers in oral squamous cell tumors. Neoplasma 52(2): 95-102.

Panicker G, Lee DR, Unger ER (2009). Optimization of SELDI-TOF protein profiling for analysis of cervical mucous. J. Proteomics 71: 637–46.

Panicker G, Yiming Y, Dongxia W, Elizabeth RU (2010). Characterization of the Human Cervical Mucous Proteome. Clin. Proteomics 6: 18–28.

Sun G, He H, Ping FY, Zhang FF (2009). Proteomic Diagnosis Models from Serum for Early Detection of Oral Squamous Cell Carcinoma. Artif. Cells Blood Substit. Immobil. Biotechnol. 37(3): 125-129.

Weinberger, SR, Morris TS, Pawlak M (2000). Recent trends in protein biochip technology. Pharmacogenomics 1(4):395-416.

Xu WH, Chen YD, Hu Y, Yu JK (2006). Preoperatively molecular staging with CM10 ProteinChip and SELDI-TOF-MS for colorectal cancer patients. J. Zhejiang Univ. Sci. B. 7(3): 235-240.

Yu JK, Chen YD, Zheng S (2004). An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics. World J. Gastroenterol. 10(21):3127-31.

Yu JK, Zheng S, Tang Y, Li L (2005). An integrated approach utilizing proteomics and bioinformatics to detect ovarian cancer. J. Zhejiang Univ. Sci. B. 6(4): 227-31.

Zhou ZY, Tao DD, Cao JW, Luo HS (2013). Application of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry technology for the diagnosis of colorectal adenoma. Oncol. Lett. 5(6):1935-1938.