

# Serum profiling by mass spectrometry combined with bioinformatics for the biomarkers discovery in diffuse large B-cell lymphoma

Wenhong Xu · Yue Hu · Xuexin He · Jun Li · Tao Pan · Hai Liu · Xianguo Wu · Hong He · Weiting Ge · Jiekai Yu · Qichun Wei · Shu Zheng · Suzhan Zhang · Yiding Chen

Received: 1 August 2014 / Accepted: 7 November 2014 / Published online: 20 November 2014  
© International Society of Oncology and BioMarkers (ISOBM) 2014

**Abstract** The aim of this study was to identify potential serum biomarkers of diffuse large B-cell lymphoma (DLBCL) and to detect DLBCL therapy response biomarkers. DLBCL serum proteomic analysis was performed using the CM10 ProteinChip mass spectrometry (SELDI-TOF-MS) approach combined with bioinformatics. A total of 178 samples were analyzed in this study from untreated early stage DLBCL patients (38), patients with inflammatory lymphadenopathy (13), healthy donors (35), post-treatment non-relapsed DLBCL patients (53), and relapsed DLBCL patients (39). Model 1 formed by nine protein peaks ( $m/z$ : 6443, 5913, 6198, 4098, 7775, 9293, 5946, 5977, and 4628) could be used to distinguish DLBCL patients from healthy individuals with an accuracy of 95.89 % (70/73). The diagnostic pattern constructed using the support vector machine including the nine

proteins of model 1, showed a maximum Youden's Index. Model 2 formed by three protein peaks ( $m/z$ : 3942, 6639, and 4121) could be used to distinguish DLBCL patients from those with inflammatory lymphadenopathy with an accuracy of 94.12 % (48/51). Model 3 formed by six protein peaks could distinguish patients with inflammatory lymphadenopathy from healthy individuals with an accuracy of 97.92 % (47/48). Model 4 could be used to distinguish non-relapsed DLBCL patients from relapsed DLBCL patients with an accuracy of 84.78 % (78/92). The four patterns were validated by leave-one-out cross-validation. These data demonstrate that the CM10 ProteinChip and SELDI-TOF-MS approach combined with bioinformatics can be used effectively to screen for the differential protein expression profiles of DLBCL patients and to predict the response to therapy.

W. Xu · Q. Wei

Department of Radiation Oncology, the Second Affiliated Hospital, Zhejiang University, College of Medicine, Hangzhou, Zhejiang, China

W. Xu · Y. Hu · X. He · J. Li · T. Pan · H. Liu · H. He · W. Ge · J. Yu · S. Zheng · S. Zhang · Y. Chen

Cancer Institute, the Second Affiliated Hospital, Zhejiang University, College of Medicine, Hangzhou, Zhejiang, China

X. Wu

Clinical Laboratory, the Second Affiliated Hospital, Zhejiang University, College of Medicine, Hangzhou, Zhejiang, China

H. He

Stomatology, the Affiliated Stomatology Hospital, Zhejiang University, College of Medicine, Hangzhou, Zhejiang, China

Y. Chen (✉)

Surgery Oncology, the Second Affiliated Hospital, Zhejiang University, College of Medicine, No. 88 Jiefang Road, 310009 Hangzhou, Zhejiang Province, China  
e-mail: birdswan@sina.com

**Keywords** Diffuse large B-cell lymphoma · Biomarkers · Proteomic · SELDI · Bioinformatics

## Introduction

Non-Hodgkin lymphoma is one of the most malignant tumors occurring with the seventh highest incidence and ninth mortality rate in males and the sixth highest incidence and seventh mortality rate in females in developed countries [1]. The incidence and mortality rate of non-Hodgkin lymphoma is also increasing in developing countries. Diffuse large B-cell lymphoma (DLBCL) is the most common form of lymphoma, accounting for 30–40 % of newly diagnosed non-Hodgkin lymphomas [2]; there are no satisfactory biomarkers available to screen for DLBCL. Detecting cancers at their earliest stages of development is critical for achieving improved rates for curing the disease. Among DLBCL patients, less than one half

of all diagnosed cases are cured with conventional chemotherapy; however, DLBCL is a heterogeneous disease with significant variations in clinical outcome [3]. Prognostic stratification is based on clinical risk factors used to determine the International Prognostic Index (IPI), which incorporates patient age at diagnosis, tumor stage, serum lactate dehydrogenase level, performance status, and the number of extranodal sites [4]. However, the use of the IPI for identification of high-risk patients is not completely reliable due to the intrinsic molecular heterogeneity of this disease [5]. Therefore, the identification of prognostic serum biomarkers that can be used to reliably distinguish DLBCL patients is highly important for improving patient outcomes.

Recently, the focus of cancer research has expanded from the analysis of the genetic information encoded by the human genome to the analysis of protein expression. ProteinChip is a new proteomics technique that is based on surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) [6]. The ProteinChip array surface binds proteins non-selectively, enabling high-throughput protein profiling and providing a sensitive system for detecting and analyzing multiple proteins. This technology has been applied successfully to the discovery of biomarkers of a number of cancers [7–9].

The purpose of this study was to utilize SELDI-TOF-MS and bioinformatics techniques to generate and analyze protein profiles in serum samples from patients with DLBCL and to distinguish prognostic markers for patients with DLBCL.

## Materials and methods

### Patients and samples

This study was approved by the Ethics Committee of Zhejiang University (China), and written informed consent was obtained from all patients and donors prior to inclusion. In total, 178 individuals (38 patients diagnosed as untreated early stage DLBCL, 13 patients with inflammatory lymphadenopathy, 35 healthy donors, 53 post-treatment non-relapsed DLBCL patients, and 39 relapsed DLBCL patients) who presented to the Second Affiliated Hospital, Zhejiang University, College of Medicine (China) were included in our study. The diagnoses of malignant tumors and inflammatory lymphadenopathy were confirmed by biopsy pathology. The patients with early stage DLBCL were classified as stage I according to the Ann Arbor staging system for lymphomas. All the enrolled post-treatment patients received a standard R-CHOP chemotherapy. Each group was age- and sex-matched with a corresponding control group.

### Sample preparation and SELDI-TOF-MS analysis

The Protein Biological System (PBS) II ProteinChip plus SELDI-TOF-MS and the CM10 ProteinChip were purchased from Ciphergen Biosystems (Fremont, CA, USA). Sinapinic acid (SPA) was purchased from Fluka (Buchs, Switzerland). All other reagents were purchased from Sigma (Sigma-Aldrich; St. Louis, MO, USA). Serum samples were collected from patients in the early morning before breakfast and were stored at room temperature for 1–2 h before centrifugation at 3000 rpm for 10 min. The samples were then stored at  $-80^{\circ}\text{C}$  until required.

Prior to analysis, the samples were thawed and centrifuged at 10,000 rpm for 2 min. The supernatants were removed, and 5  $\mu\text{l}$  of each sample was added to 10  $\mu\text{l}$  of 0.5 % U9 buffer (9 M urea, 0.2 % CHAPS, 0.1 % dithiothreitol) in a 96-well plate. Samples were then incubated for 30 min at  $4^{\circ}\text{C}$  on a platform shaker set at 600 rpm.

The ProteinChip array cassette was placed in a 96-well bioprocessor, and 200  $\mu\text{l}$  NaAC (50 mM, pH 4.0) was added to each well. The cassette was incubated for 5 min at  $4^{\circ}\text{C}$  and shaken at 600 rpm. The liquid was removed and retained, and the procedure was repeated. A further 185  $\mu\text{l}$  NaAC was added to each well, and the cassette was incubated for 2 min at  $4^{\circ}\text{C}$  with shaking at 600 rpm. Samples retained from different patients were added to separate wells (100  $\mu\text{l}$ /well), and the cassette was incubated for 1 h at  $4^{\circ}\text{C}$ , with shaking at 600 rpm. After removing excess liquid, the array was washed twice with 200  $\mu\text{l}$  NaAC for 5 min and shaken at 600 rpm and rinsed twice with high-performance liquid chromatography water.

Prior to SELDI-TOF-MS analysis, 1  $\mu\text{l}$  of a saturated solution of SPA energy-absorbing molecules was applied onto each chip array. The array surface was allowed to air-dry, and the application was repeated. The CM10 ProteinChip is a weak cation exchange array with a carboxylate surface that binds cationic proteins; therefore, the negatively charged carboxylate groups interact with the positive charges exposed on target proteins. The ProteinChip was placed on the PBS II mass spectrometer reader, the automatic baseline correction was applied, and peaks between 2000 and 20,000 Da were detected automatically. Peaks with mass-to-charge ( $m/z$ ) ratios  $<2000$  Da were mainly due to ion noise from the matrix and were excluded. Ciphergen ProteinChip Software v. 3.2 was used to analyze the data.

### Bioinformatics and statistical analysis

Data analysis was performed using the Zhejiang University Cancer Institute-ProteinChip Data Analysis System (ZUCI-PDAS) on a MATLAB Web Server v. 1.2.4 (MathWorks Inc., Natick, MA, USA). The undecimated discrete wavelet

**Table 1** The cross-validation blind test results of the test set in model 1 (cases)

Groups	Normal	DLBCL	Total
Normal	33	1	34
DLBCL	2	37	39
Total	35	38	73

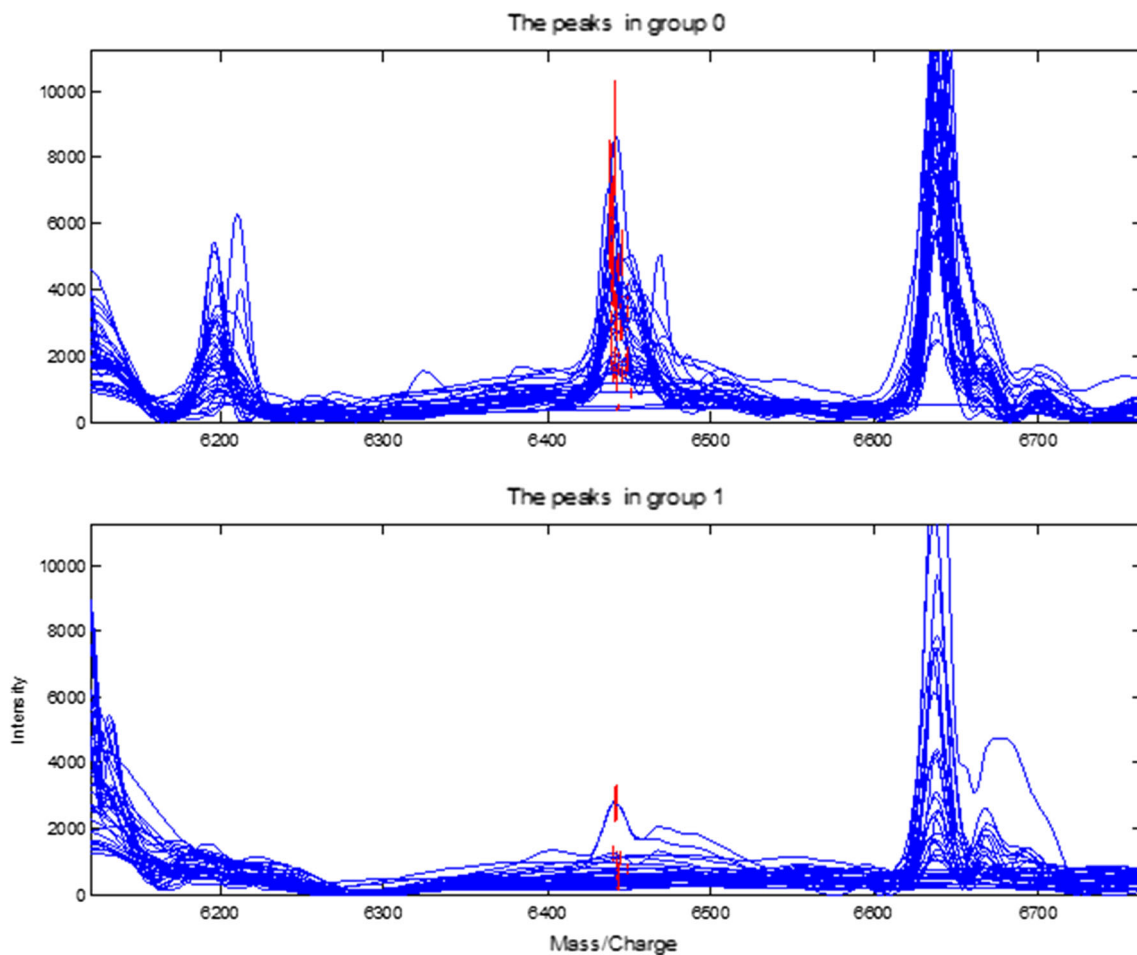
Sensitivity of 94.29 % (33/35), specificity of 97.37 % (37/38), and accuracy of 95.89 % (70/73)

transform (UDWT) method was applied to denoise the signals using the Rice Wavelet Toolbox v. 2.4. Baseline correction was achieved by aligning the spectra with a monotone local minimum curve and mass calibration was carried out by adjusting the intensity scale to three peaks that were present in all the spectra. Detection and quantification of the peaks was performed using an algorithm that evaluated the maximum height of every denoised, baseline-corrected, and calibrated mass spectrum. The minimum height of each peak was set at 10. The peaks were filtered to maintain a signal-to-noise ratio (SNR) >3, where the SNR was defined as the height of

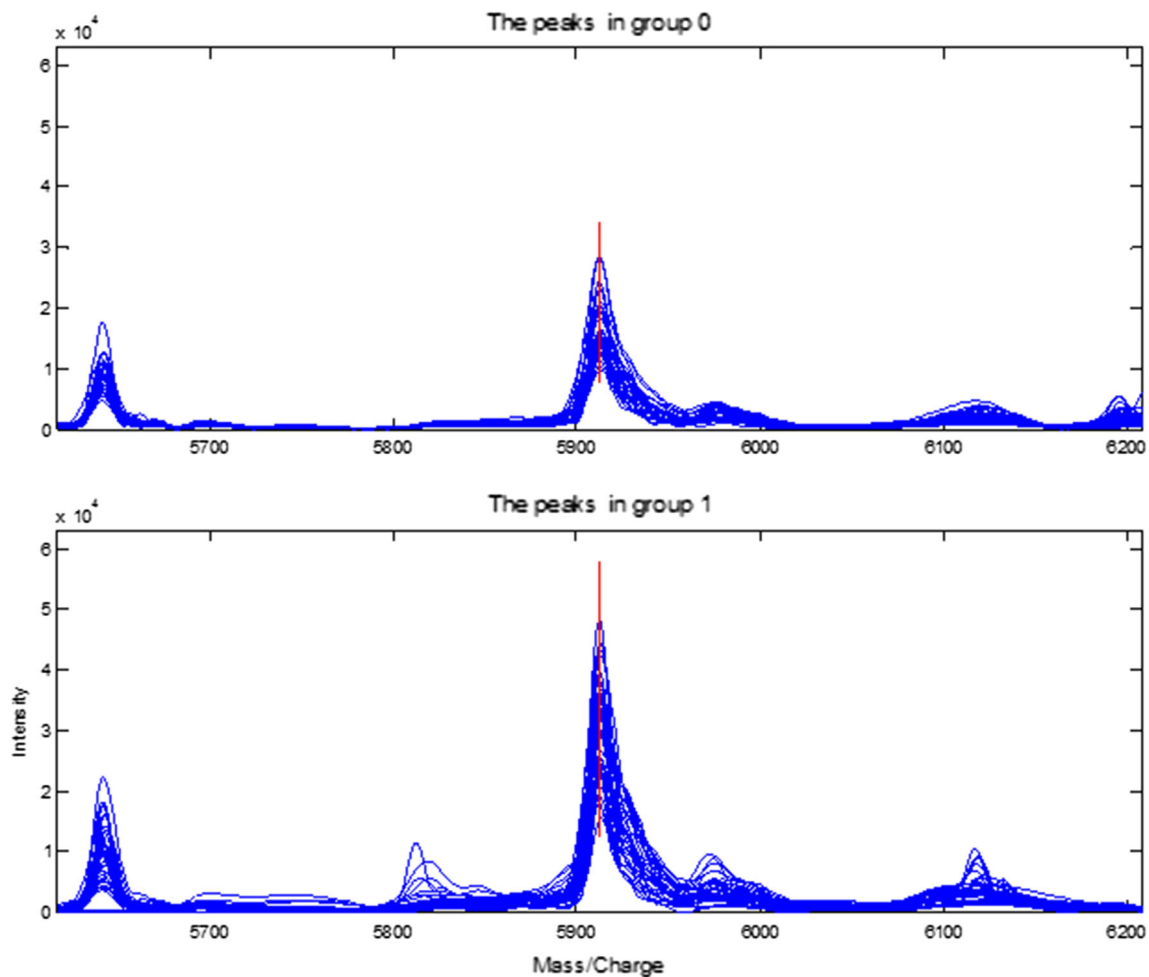
the peak above the baseline to the wavelet-defined noise. To match peaks across the spectra, peaks with a relative mass difference  $\leq 0.3$  % were pooled and defined as a cluster. The value of any peak that was absent from a cluster was set to the maximum height of that cluster. Normalization was performed only within peak clusters.

The non-linear SVM classifier was based on the shareware program OSU\_SVM v. 3.00 Toolbox of Junshui Ma and Yi Zhao. A radial-based function kernel, a parameter gamma of 0.6, and a cost of the constraint violation of 19 were used to distinguish different groups of data. The leave-one-out cross-validation (LOOCV) approach was applied to estimate the accuracy of the classifier. This approach assigns one sample as the test set, and the remaining samples are used as the training set. The process is repeated until each sample has been selected as a test set.

The Wilcoxon  $t$  test was used to estimate the potential of each peak for distinguishing different groups of data. The 10 peaks with the smallest  $P$  values were selected for further analysis. Each of the 1023 combinations of the 10 peaks was analyzed by the SVM using the LOOCV method. The



**Fig. 1** Protein mass peak 6443 expressed significantly high in normal individuals (group 0), but remarkably low in DLBCL (group 1) ( $p=8.1241e-011$ )



**Fig. 2** Protein mass peak 5913 expressed significantly high in DLBCL (group 1), but low in normal individuals (group 0) ( $p=1.1708e-010$ )

combinations that showed the highest accuracy in distinguishing different groups of data were selected as potential biomarkers. The SVM model with the highest Youden's Index was selected for the detection of DLBCL.

The reproducibility of the SELDI-TOF-MS spectra was verified using a quality control sample. The mass accuracy was  $<0.1\%$  according to the all-in-one peptide molecular mass standard (Ciphergen Biosystems).

## Results

Differential proteomics pattern of DLBCL versus normal controls (model 1)

SVM analysis using the LOOCV method revealed nine protein peaks ( $m/z$ : 6443, 5913, 6198, 4098, 7775, 9293, 5946, 5977, and 4628) that could be used to distinguish DLBCL

**Table 2** The nine selected discrepant protein mass peaks comparison of normal individuals with DLBCL patients of model 2; mean and standard deviation (SD)

Protein peaks ( $m/z$ )	Normal (mean $\pm$ SD)	DLBCL (mean $\pm$ SD)	$P$ value ( $t$ test)
6443	3293.0304 $\pm$ 2006.4098	637.7786 $\pm$ 578.1431	8.1241e-011
5913	16,191.4875 $\pm$ 4628.7639	30,643.9404 $\pm$ 8650.54	1.1708e-010
6198	8819.1506 $\pm$ 3385.9687	3230.239 $\pm$ 3811.1509	5.0573e-008
4098	2076.5592 $\pm$ 1290.3524	813.6088 $\pm$ 278.8543	6.5425e-007
7775	18,191.9167 $\pm$ 4016.9514	25,743.4027 $\pm$ 7234.7128	1.695e-006
9293	1594.5423 $\pm$ 976.1154	678.7395 $\pm$ 679.8919	4.2444e-006
5946	1632.9509 $\pm$ 651.5244	2654.1315 $\pm$ 960.4347	6.1332e-006
5977	3497.2862 $\pm$ 2631.4232	994.7109 $\pm$ 1532.0075	2.6512e-005
4628	2938.2603 $\pm$ 880.1372	1830.2684 $\pm$ 1211.3899	8.656e-005

from healthy individuals with an accuracy of 95.89 % (70/73) (Table 1, Figs. 1 and 2). The diagnostic pattern constructed using SVM analysis, including the nine proteins of model 1, showed the maximum Youden's Index (YI). Among these peaks, those at 6443, 6198, 4098, 9293, 5977, and 4628 *m/z* were found at high intensity in normal individuals, but at low intensity in DLBCL patients. In contrast, the peaks at 5913, 7775, and 5946 *m/z* were detected at significantly higher levels in DLBCL patients compared with those in normal individuals (Table 2). Representative peaks are shown in Fig. 1. The sensitivity and specificity of detection were 94.29 % (33/35) and 97.37 % (37/38), respectively (Table 1).

#### Differential proteomics pattern of DLBCL versus inflammatory lymphadenopathy patients (model 2)

Through LOOCV, support vector machine (SVM) analysis was used to screen out three protein peaks (*m/z*: 3942, 6639, and 4121) to build model 2, which could be used to define the peptide patterns for distinguishing DLBCL from patients with inflammatory lymphadenopathy. Combinations with the highest accuracy were chosen as potential biomarkers, and the SVM model with the highest Youden's Index (YI) was used as the diagnostic model (Table 3). The accuracy of the blind prediction was 94.12 % (48/51). The sensitivity and specificity of the detection were 84.62 % (11/13) and 97.37 % (37/38), respectively (Table 4).

#### Differential proteomics pattern of inflammatory lymphadenopathy patients versus normal individuals (model 3)

Model 3 formed by six protein peaks could be used to distinguish patients with inflammatory lymphadenopathy from healthy individuals with an accuracy of 97.92 % (47/48) (Table 5), as evaluated by LOOCV. The descriptive statistics of these six peaks are shown in Table 6.

**Table 3** The three selected protein mass peaks comparison of inflammatory lymphadenopathy patients with DLBCL patients of model 2; mean and standard deviation (SD)

Protein peaks ( <i>m/z</i> )	Inflammatory lymphadenopathy (mean±SD)	DLBCL (mean±SD)	<i>P</i> value ( <i>t</i> test)
3942	968.455±1093.5794	2599.6268±2762.2066	0.0046321
6639	777.7421±997.6501	2285.6967±2796.1499	0.013735
4121	3263.3611±1662.5648	2655.5581±1678.8008	0.19467

**Table 4** The cross-validation blind test results of the test set in model 2 (cases)

Groups	Inflammatory lymphadenopathy	DLBCL	Total
Inflammatory lymphadenopathy	11 (84.62 %)	1 (15.38 %)	12
DLBCL	2 (2.63 %)	37 (97.37 %)	39
Total	13	38	51

Sensitivity of 84.62 % (11/13), specificity of 97.37 % (37/38), and accuracy of 94.12 % (48/51)

#### Differential proteomics pattern of non-relapse DLBCL patients versus relapse DLBCL patients (model 4)

Model 4 formed by three protein peaks could be used to distinguish non-relapsed DLBCL patients from relapsed DLBCL patients with an accuracy of 84.78 % (78/92) (Table 7), as evaluated by LOOCV. The descriptive statistics of these three peaks are shown in Table 8.

## Discussion

ProteinChip technology has enabled the application of high-throughput proteomics to the discovery of potential biomarkers of a variety of diseases. The advantages of this technique include the simultaneous analysis of small samples on a single chip, a relatively short experimental duration, and the ability to separate and identify peptides expressed on the chip surface based on differences between their intensities, without a requirement for purification [10].

Support vector machines (SVMs) are a machine learning approach originally proposed and developed by Vladimir Vapnik in 1995. This approach is based on the structural risk minimization principle (SRM), which can be used to solve multiple problems in pattern recognition involving small sample sizes, nonlinearity, or high dimensional data. The central principle is based on generalization using a classifier that is required to be applicable not only on training samples but also

**Table 5** The cross-validation blind test results of the test set in model 3 (cases)

Groups	Normal	Inflammatory lymphadenopathy	Total
Normal	34	0	34
Inflammatory lymphadenopathy	1	13	14
Total	35	13	48

Sensitivity of 97.14 % (34/35), specificity of 100 % (13/13), and accuracy of 97.92 % (47/48)

**Table 6** The six selected discrepant protein mass peaks comparison of normal individuals with inflammatory lymphadenopathy patients of model 3; mean and standard deviation (SD)

Protein peaks ( <i>m/z</i> )	Normal (mean± <i>SD</i> )	Inflammatory lymphadenopathy (mean± <i>SD</i> )	<i>P</i> value ( <i>t</i> test)
4161	10,180.8592±3946.7984	1419.0127±1667.1331	1.4464e-006
6110	3799.2852±2317.8655	570.3338±167.0196	2.6463e-006
8698	2522.8245±761.3486	1159.9095±549.2234	9.5046e-006
5895	6403.2469±3561.6559	848.4002±2295.3917	1.1903e-005
4082	2372.3278±2466.3141	1135.6346±3342.8063	4.8616e-005
2957	4893.2962±2297.818	1732.4801±1760.3489	9.0105e-005

to previously unseen samples or test samples. Although this approach was previously recognized, the development of SVM has enabled its effective application. The intrinsic complexity of biological systems and their high dimensionality compared to small sample sets render SVM an ideal tool for the analysis of high-throughput biological data.

Reduction the number of false protein peaks created through random variation remains one of the challenges in analyzing SELDI-TOF-MS-generated data [11]. To overcome this problem, we developed a bioinformatics tool based on an algorithm that was able to identify most of the true, reproducible peaks. This tool was designed using the ZUCI-PDAS with UDWT denoising and included baseline correction, peak detection, biomarker selection, and the establishment and evaluation of differential SVM patterns based on statistical theory. The SVM can be used to solve problems such as the generalization in pattern recognition, pattern selection, and over-fitting of medium and small sample sizes; therefore, it is applicable to finite samples, rather than to sample sizes approaching infinity, to give an optimal solution with the available information. In theory, obtaining the optimal point can solve problems caused by local extrema, which are otherwise unsolvable by artificial neural network (ANN) methods. The algorithm transforms the actual condition into hyperplane feature space with non-linear transformation, and the non-linear discriminant function is created by constructing a linear discriminant function in hyperplane space. The specificity of this approach enables excellent generalization and effectively solves the dimensionality problem. The LOOCV method is used to determine the accuracy of the classifier. All these steps in combination ensure that the selection of biomarkers is not influenced by systemic biases. This

**Table 7** The cross-validation blind test results of the test set in model 4 (cases)

Groups	Non-relapse DLBCL	Relapse DLBCL	Total
Non-relapse DLBCL	45	8	53
Relapse DLBCL	6	33	39
Total	51	41	92

Sensitivity of 84.91 % (45/53), specificity of 84.62 % (33/39), and accuracy of 84.78 % (78/92)

bioinformatics tool has been used previously by our group to produce a protein fingerprint in the diagnosis of papillary thyroid carcinoma [12] and to identify candidate molecular markers in oral squamous cell carcinoma [13], gastric cancer [14], non-small cell lung cancer [15, 16], and colorectal cancer [17].

In this study, we successfully used SELDI-TOF-MS and bioinformatics techniques to analyze the sera of patients with DLBCL to build four models that can be used to discriminate between different groups of patients. Model 1 could be used to distinguish DLBCL from healthy individuals with an accuracy of 95.89 % (70/73). Model 2 could be used to distinguish DLBCL from patients with inflammatory lymphadenopathy with an accuracy of 94.12 % (48/51). Model 3 could be used to distinguish patients with inflammatory lymphadenopathy from healthy individuals with an accuracy of 97.87 % (46/47). Model 4 could be used to distinguish non-relapsed DLBCL patients from relapsed DLBCL patients with an accuracy of 84.78 % (78/92). The diagnostic pattern constructed using the SVM showed the maximum YI, and the four patterns were validated by LOOCV.

Different biomarkers are often detected by using different bioinformatics methods, such as conventional SVMs, discriminant analysis, and ANNs. The LOOCV SVM method has relatively high sensitivity and specificity; therefore, we applied this method of analysis to data obtained using the SELDI-TOF-MS and bioinformatics approaches to successfully build models of serum protein fingerprints in DLBCL. Further research is required to explore the peptides identified in our study in greater detail and to confirm our findings in a larger cohort of study samples.

**Table 8** The three selected protein mass peaks comparison of patients of non-relapse DLBCL with relapse DLBCL of model 4; mean and standard deviation (SD)

Protein peaks ( <i>m/z</i> )	Non-relapse DLBCL (mean± <i>SD</i> )	Relapse DLBCL (mean± <i>SD</i> )	<i>P</i> value ( <i>t</i> test)
2946	1610.4981±732.0167	2833.6911±2338.1275	0.0028264
5820	1402.2886±1266.7702	2438.6621±2019.0309	0.0061753
6135	2229.0353±1002.3424	3354.8768±1315.3723	0.015854

**Acknowledgments** This work was supported by the Youth Program of Natural Science Foundation of China (30901686); Zhejiang Provincial Natural Science Foundation of China (LQ13H160015 and LY14H160031); Research Program 2013KYB247 from Hygiene Bureau; Research Program Y201432699 and Y201225136 from the Education Bureau, Zhejiang Province, China; the National High Technology Research and Development Program of China (863 Program) (2012AA02A204); and the National Natural Science Foundation of China (Grant number: 81202097).

**Conflicts of interest** None.

## References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics. *CA Cancer J Clin.* 2013;63:11–30.
2. Jardin F. Next generation sequencing and the management of diffuse large B-cell lymphoma: from whole exome analysis to targeted therapy. *Discov Med.* 2014;18:51–65.
3. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. *Blood.* 1997; 89:3909–18.
4. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med.* 1993; 329:987–94.
5. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med.* 2002;8:68–74.
6. Xu WH, Chen YD, Hu Y, Yu JK, Wu XG, Jiang TJ, et al. Preoperatively molecular staging with CM10 ProteinChip and SELDI-TOF-MS for colorectal cancer patients. *J Zhejiang Univ Sci B.* 2006;7:235–40.
7. Cao XL, Li H, Yu XL, Liang P, Dong BW, Fan J, et al. Predicting early intrahepatic recurrence of hepatocellular carcinoma after microwave ablation using SELDI-TOF proteomic signature. *PLoS One.* 2013;8(12):e82448. doi:10.1371/journal.pone.0082448. eCollection 2013.
8. Zheng GB, Gao CF, Wang XL, Zhao G, Li DH. Study on serum proteomic features in patients with and without recurrence or metastasis after surgical resection of esophageal carcinoma. *Genet Mol Res.* 2014;13:538–45.
9. Liu C, Pan C, Wang H, Yong L. Effect of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry on identifying biomarkers of laryngeal carcinoma. *Tumour Biol.* 2011;32: 1139–45.
10. Rogers MA, Clarke P, Noble J, Munro NP, Paul A, Selby PJ, et al. Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility. *Cancer Res.* 2003;63:6971–83.
11. Somorjai RL, Dolenko B, Baumgartner R. Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics.* 2003;19:1484–91.
12. Wang JX, Yu JK, Wang L, Liu QL, Zhang J, Zheng S. Application of serum protein fingerprint in diagnosis of papillary thyroid carcinoma. *Proteomics.* 2006;6:5344–9.
13. He H, Sun G, Ping F. Laser-capture microdissection and protein extraction for protein fingerprint of OSCC and OLK. *Artif Cells Blood Substit Immobil Biotechnol.* 2009;37:208–13.
14. Qiu FM, Yu JK, Chen YD, Jin QF, Sui MH, Huang J. Mining novel biomarkers for prognosis of gastric cancer with serum proteomics. *J Exp Clin Cancer Res.* 2009;28:126. doi:10.1186/1756-9966-28-126.
15. Han M, Liu Q, Yu J, Zheng S. Identification of candidate molecular markers predicting chemotherapy resistance in non-small cell lung cancer. *Clin Chem Lab Med.* 2010;48:863–7.
16. Yang S, Nan Y, Tian Y, Zhang W, Zhou B, Bu L, et al. Study of distinct protein profiles for early diagnosis of NSCLC using LCM and SELDI-TOF-MS. *Med Oncol.* 2008;25:380–6.
17. Xu W, Chen Y, He W, Fu Z, Pan T, He H, et al. Protein fingerprint of colorectal cancer, adenomatous polyps, and normal mucosa using ProteinChip analysis on laser capture microdissected cells. *Discov Med.* 2014;17:223–31.