

## O-glycan profiling of serum glycan for potential renal cancer biomarkers

GAO Yi<sup>1</sup>, MA Fei<sup>2,3</sup>, ZHANG Wei<sup>4</sup>, ZHONG Fan<sup>5</sup>, TANG HaiLin<sup>4</sup>,  
XU DanFeng<sup>1\*</sup> & ZHAO Liang<sup>2\*</sup>

<sup>1</sup>Department of Urology, Changzheng Hospital of Second Military Medical University, Shanghai 200003, China;

<sup>2</sup>State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing 102206, China;

<sup>3</sup>Shanghai Key Laboratory of Functional Materials Chemistry, School of Chemistry and Molecular Engineering, East China University of Science and Technology, Shanghai 200237, China;

<sup>4</sup>Department of Automatic Control, College of Mechatronics and Automation, National University of Defense Technology, Changsha 410073, China;

<sup>5</sup>Laboratory of Systems Biology, Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China

Received April 16, 2013; accepted May 8, 2013; published online June 17, 2013

Serum was obtained from 25 male renal cell carcinoma (RCC) patients and 21 healthy males. O-glycans were released by a  $\beta$ -elimination reaction and purified by graphitized carbon cartridge solid phase extraction, then profiled by matrix-assisted laser desorption/ionisation-time of flight mass spectrometry. After noise removal and peak alignment, 1372 peaks were extracted from 200000 data points. Feature peaks were analyzed by calculation of differential sensitivity and specificity. The combination of two feature peaks was chosen as a biomarker and could clearly differentiate RCC and normal samples in our study group.

**mass spectrometry, glycan, renal cell carcinoma, biomarkers**

**Citation:** Gao Y, Ma F, Zhang W, et al. O-glycan profiling of serum glycan for potential renal cancer biomarkers. *Sci China Life Sci*, 2013, 56: 739–744, doi: 10.1007/s11427-013-4498-9

Approximately 210000 cases of renal cell carcinoma (RCC) are diagnosed annually worldwide, accounting for 102000 deaths [1]. In the European Union and United States, RCC is the ninth most common cancer [2], while it is the sixth most common cancer in China [3]. Local RCC incidence increased from 0.4 per 1000 in 2000 to 0.6 in 2010 [3], and has become an important threat to health in China. Environmental risk factors are considered as major causes for the increase in the past 10 years. The five-year survival rate of RCC is over 50% in patients without distant metastases; however five-year survival falls to less than 10% in cases of distant metastases [4]. Because there are currently no avail-

able biomarkers for RCC, one-third of patients are found to have metastasis at first diagnosis [5]. Sensitive and convenient RCC screening can greatly improve prognosis and five-year survival; therefore efforts to improve screening methods are urgently required in RCC research.

Protein glycosylation is one of the most important post-translational modifications (PTMs), and over 50% proteins are glycosylated [6]. Glycans attached to proteins have very important functions in cancer biology. Glycan changes occur in all stages of tumor progression, including proliferation, invasion, angiogenesis, metastasis and immunity [7]. Therefore, changes in glycans are a hallmark of cancer states [8]. With rapid developments in mass spectrometry, glycans can now be profiled. Moreover, they are

\*Corresponding author (email: zhaoliangbj@189.cn; xu-danfeng@hotmail.com)

easier to identify and quantify compared with proteins because of their smaller molecular weight and greater stability [9]. All these factors make glycans appropriate potential biomarkers [10] for cancer.

Many studies have investigated glycan cancer biomarkers in recent years. The majority focused on N-linked glycans [11–13], because there is no broad-spectrum glycosidase for O-glycosylation, and this hinders O-glycan biomarker studies. In fact, both O-linked glycans and N-linked glycans are important in cancer. Mucins (MUC) are used as cancer biomarkers in many cancers that are characterized by dense clusters of O-linked glycans, and *MUC1* is highly overexpressed in RCC [14,15]. Examination of O-glycan changes by serum O-glycan profiling will lead to new strategies for RCC biomarker discovery.

In this research, O-linked glycans were released by  $\beta$ -elimination and purified through graphitized carbon solid-phase extraction cartridges for mass spectra analysis [16]. Because there was no high-performance liquid chromatography separation section, the experimental time was considerably smaller. This would be beneficial in further large sample size validation.

## 1 Materials and methods

### 1.1 Human serum sample preparation

Serum samples from healthy male individuals ( $n=21$ ) and male patients with histologically confirmed RCC ( $n=25$ ) were obtained from the Shanghai Municipal Center for Disease Control & Prevention (Shanghai CDC) and Shanghai Changzheng Hospital, respectively. All samples were processed as follows: blood samples were first incubated in anticoagulant-free tubes at 37°C for 1 h, and then centrifuged at 2400 $\times$ g for 10 min. The supernatant was collected and stored at -80°C as serum samples for subsequent experiments.

### 1.2 Release of O-glycans

A total of 10  $\mu$ L serum was mixed with 10  $\mu$ L 2 mol L<sup>-1</sup> NaBH<sub>4</sub> (China National Pharmaceutical Group Co., Shanghai, China) in 0.1 mol L<sup>-1</sup> NaOH (Sinopharm Chemical Reagent Co., Ltd). This mixture was incubated at 42°C for 18 h, and then 1 mol L<sup>-1</sup> HCl (China National Pharmaceutical Group) was slowly added until the pH was between 3 and 5 to quench the reaction. The solution was centrifuged at 2400 $\times$ g for 5 min at 4°C, and the supernatant was diluted to 200  $\mu$ L and subjected to solid-phase extraction for oligosaccharide purification and enrichment.

### 1.3 Solid-phase extraction (SPE) of O-glycans

Solid-phase extraction was performed using graphitized carbon cartridges (Alltech Associates, Inc., Chicago, Illinois,

USA). Cartridges were preconditioned with 1 mL ACN/H<sub>2</sub>O (80%/20%, v/v) mixture containing 0.05% chromatographically pure TFA (Sigma-Aldrich Co. Ltd., Gillingham, UK) at a flow rate of 0.5 mL min<sup>-1</sup>, followed by 2 mL H<sub>2</sub>O at the flow rate of 0.5 mL min<sup>-1</sup>. In the loading phase, 200  $\mu$ L  $\beta$ -elimination product was added and the cartridge was slowly rotated approximately 3–6 times without external pressure. In the desalting phase, 5 mL H<sub>2</sub>O was worked through the column at a flow rate of 0.5 mL min<sup>-1</sup>. Then 1 mL 10% ACN (China National Pharmaceutical Group) was eluted at a flow rate of 0.5 mL min<sup>-1</sup>. The eluate was collected and concentrated to 10  $\mu$ L for mass spectrometry (MS) analysis.

### 1.4 Mass spectrometry

Mass spectra with an external matrix-assisted laser desorption/ionization (MALDI) source were carried out on Bruker Ultraflex III TOF/TOF (Bruker Daltonics, Billerica, Massachusetts, USA) at positive ion mode. Purified glycan samples were combined in a 1:1 ratio with 2,5-dihydroxybenzoic acid (DHB) matrix solution. The matrix solution was prepared by 10 mg DHB in 1 mL 50% ACN/H<sub>2</sub>O (v/v). Mass spectra were controlled by FlexControl 3.3 software (Bruker Daltonics) using the following parameters: 25 kV acceleration voltage, 21.85 kV pulse voltage, 30 ns ion extraction delay, and 1.4 kV detector gain. Results were analyzed using FlexAnalysis 3.3 software (Bruker Daltonics Ins). The data were collected and imaged using FlexImaging 3.0 software (Bruker Daltonics).

### 1.5 Raw mass data processing

Detection of signals from raw data consisted of two parts: peak detection from single mass spectrum and peak alignment among multiple spectra. Peak detection from single mass spectrum was performed using the following steps: peak finding, chemical noise removal, normalization and deisotoping. Undecimated discrete wavelet transform (UDWT) was applied to remove the high frequency random noise from the MALDI-TOF data. The baseline was corrected by linear interpolation and chemical noise was removed by a robust regression method.

After removal of the chemical noise, peak heights were used to estimate the abundance for every glycan peak. Median normalization was applied to eliminate the system error. Finally, the isotope clusters were identified by detecting whether the distance between two peaks was within (1 $\pm$ 0.2) Da, and the highest peak among an isotope cluster was selected.

If the distance between two peaks from different spectra was within -0.25 to 0.25 Da, the two peaks were viewed as a MS signal from the same glycan in different spectra. All peaks were aligned using the aforementioned principle.

## 2 Results and discussion

### 2.1 MALDI-TOF data processing

Because matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) offers a simple means for screening complex mixtures [17], it plays a crucial role in the characterization of glycosylation [18]. Peak detection is a pivotal first step in biomarker discovery. It is not only an important extraction step but also indispensable for subsequent glycan identification and quantification, and can significantly influence the results of downstream data analysis steps. Hence processing of MALDI-TOF MS raw data is a central issue for glycan biomarker analysis.

The MALDI-TOF MS raw data consists of the possible peak signal, random noise and chemical noise. Random noise is normally high frequency noise caused by the instrument. Chemical noise is low frequency noise, resulting from the response to MALDI matrix and chemical reagents. Chemical noise peaks are separated by a little more than 1 Da, and sinusoidal in shape. Removal of chemical noise and random noise significantly improves the signal-to-noise ratio (SNR).

Figure 1 shows the TOF data. Chemical noise signals have the same pattern. Peak intensity exhibited some variation locally, and decreased with increasing  $m/z$ . Spectrum features were similar to Q-TOF data [19] and pTOF data [20].

Peak detection from single mass spectra includes peak identification, chemical noise removal, normalization and deisotoping. Figure 2 gives a concrete example of peak identification and chemical noise removal. In this glycomic

experiment, a large quantity of MS data were generated. Each mass spectrum consisted of more 200000 data points. It was very difficult to perform biomarker analysis directly on the raw MS data. Following an existing analytic approach [21], high frequency random noise was first removed by the undecimated discrete wavelet transform (UDWT) method. Then, the baseline was corrected using linear interpolation. Finally, the local maximum of  $N$  neighboring points was viewed as the peak. It is worth noting that the peaks obtained using this method were not only the glycan peaks, but also the chemical noise peaks.

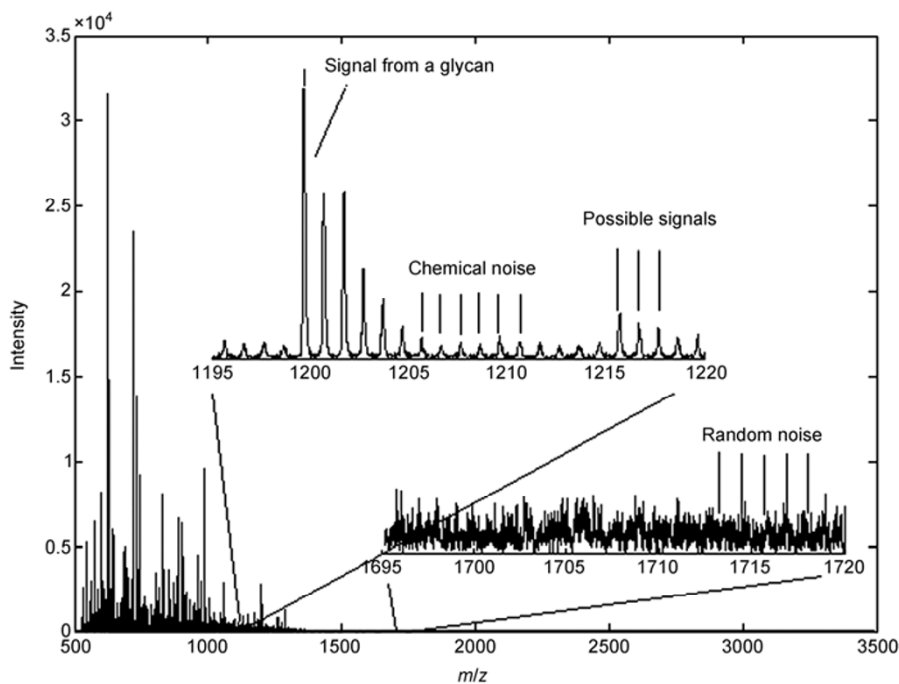
The robust regression method [19] was applied to detect glycan peaks from chemical noise. This method views the glycan peaks as outliers of chemical noise distributions. Because the robust regression method includes the effect of outliers, a linear robust regression model is used to estimate the intensity distribution and the location distribution of chemical noise peaks. First, using all peaks intensities, a linear robust regression model,

$$\mu_I(m/z) = a + b(m/z),$$

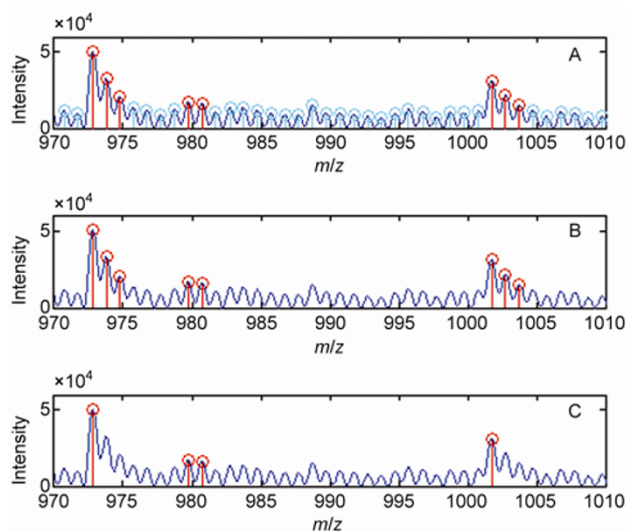
was formed to estimate the relationship between the chemical noise intensity  $\mu_I$  and  $m/z$ . The parameters  $a$  and  $b$  were estimated by the robust regression method. Peaks with intensities exceeding  $0.2\mu_I$  were retained. In contrast, chemical noise peaks are separated by approximately 1 Da, and this frequency is nearly constant over the entire  $m/z$  range. The glycan peaks may not share this pattern. Second, a robust M-type regression model,

$$\mu_M(m/z) = a + b(m/z),$$

was fitted to estimate the relationship between the location



**Figure 1** A raw spectrum, including the true or possible peak signal, chemical noise and random noise.



**Figure 2** An example of the peak finding, chemical noise removal and deisotope process. A, The spectrum after peak finding. The blue line shows the spectrum after UDWT denoising and baseline removal, the cyan line shows the chemical noise peaks, and the red line shows the true peaks. B, Peaks after chemical noise removal (red line). C, Peaks after deisotope (red line).

offset  $\mu_M$  and  $m/z$ . The parameters  $a$  and  $b$  were estimated by using the peaks with intensities greater than  $0.2 \mu_l$  based on the robust regression method. The standard deviation  $\sigma_M(m/z)$  of robust regression can be used to detect glycan peaks. The peaks with large  $\sigma_M(m/z)$  deviate from the location distribution of chemical noise peaks. Peaks with studentized residuals that exceed 2.5 are glycan peaks. Last, a linear robust regression model,

$$\mu_l(m/z) = a + b(m/z),$$

was applied to estimate the relationship between the chemical noise intensity  $\mu_l$  and  $m/z$  for the glycan peaks. The parameters  $a$  and  $b$  could also be estimated by the robust regression method. The standard deviation  $\sigma_l(m/z)$  of robust regression is used to detect glycan peaks. Peaks with large  $\sigma_l(m/z)$  deviate from the intensity distribution of chemical noise peaks. If the studentized residual of a peak exceeds 2.35, the peak is classed as a glycan peak. Using the robust regression method, the glycan peaks were detected and the default parameters for this method were selected.

Figure 2 shows the peak detection results, in which the red line shows the true peaks and the cyan line shows the chemical noise peaks. It is clear that the mass spectrum quality was greatly improved and peak complexity was greatly reduced after noise removal. Finally, 1372 peaks were extracted from 200000 data points.

## 2.2 Analysis of serum glycans

In total, 25 serum samples of serum from renal cell carcinoma patients (group A) and 21 serum samples from healthy individuals were taken (group B). Peak quantitative

character (peak intensity) and peak qualitative character (peak exists only in one group) were evaluated as differentiation markers. Qualitative character was more effective and was chosen as the differentiation character in this study (data not shown).

If one glycan peak appearing in group A alone can determine a sample as RCC, the sensitivity is  $N_A/25$  and the specificity is  $1-N_B/21$ . Similarly, if one glycan peak appearing in group B alone can determine a sample as normal, the sensitivity is  $N_B/21$  and the specificity is  $1-N_A/25$ .

Only peaks in group A or group B with peak sensitivity greater than 50% were considered. Thus the peak coverage threshold in group A and B should be above 55.2% (13/25) and 52.38% (11/21), respectively. When combining the peaks eligible for two conditions, 110 peaks were screened out from a total of 1372 peaks.

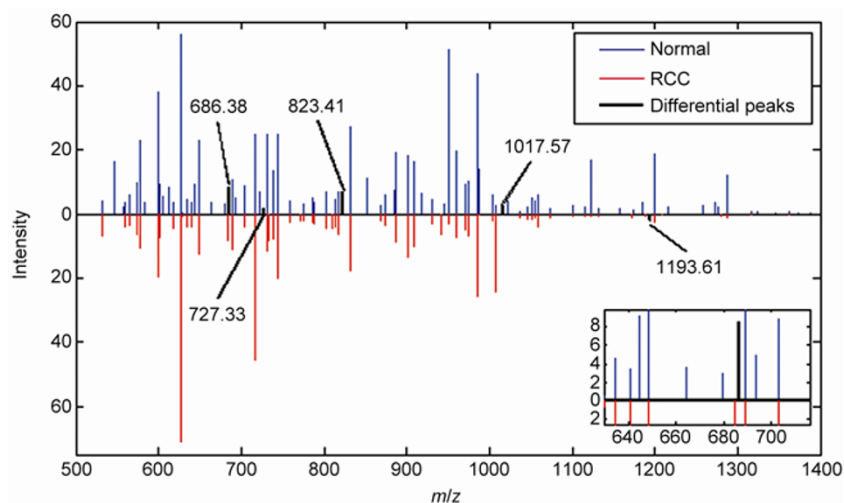
Peak sensitivity and specificity were calculated for 110 peaks selected in the first screening. Specificity was considered as a more important parameter because it can control the false positive ratio. The threshold of specificity was set at 80%. There were five peaks whose specificity was above 80%; four peaks were indicators for the normal state and one peak was an indicator for RCC state, with P72 as the peak with the best differentiation performance. Detailed information of the sensitivity and specificity of these five peaks is shown in Table 1.

Differentiation performance was evaluated by visualization. We randomly took a pair of RCC and normal O-glycan profiles as examples. In Figure 3, the marked peaks are the obtained peaks in the above analysis, and these five peaks could clearly differentiate between RCC and the normal state. Also, the average feature peak intensity was calculated for both RCC and the normal group. Average peak intensity is represented by different shade of colors in Figure 4. The five marked peaks clearly showed major differences between RCC and the normal group.

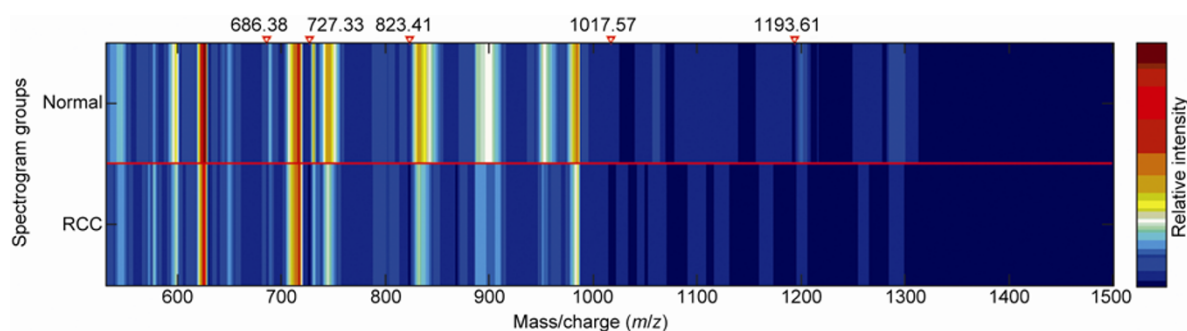
Unfortunately, sensitivity was not very high (Table 1) when the five different peaks were individually considered as biomarkers. However, when we combined two or more peaks as an indicator of disease state, the sensitivity was significantly improved. The combination of P72 and P47 was taken as a potential biomarker, and thus in this combined model the appearance of either P21 or P47 could identify the normal state. The sensitivity and specificity of this combination was calculated.

**Table 1** Feature peaks (specificity>80%)

Peak ID	Observed $m/z$	Marker type	Sensitivity	Specificity
P26	686.38	Normal	57.14% (12/21)	84.00% (21/25)
P33	727.33	Normal	52.38% (11/21)	80.00% (20/25)
P47	823.41	Normal	61.90% (13/21)	84.00% (21/25)
P72	1017.57	Normal	57.14% (12/21)	96.00% (24/25)
P93	1193.61	RCC	56.00% (14/25)	80.95% (17/21)



**Figure 3** MS spectrum of RCC and normal state.



**Figure 4** Average peak intensity MS spectrum heat map of RCC and normal group.

$$SE_{\text{combine}} = 1 - (1 - SE_{P72})(1 - SE_{P47})$$

$$= 1 - (1 - 57.14\%) \times (1 - 61.90\%) = 83.67\%$$

$$SP_{\text{combine}} = SP_{P72} \cdot SP_{P47} = 96.00\% \times 84.00\% = 80.64\%$$

Sensitivity was as high as 83.67%, while specificity was still 80.64% (above 80%).

### 3 Conclusion

In this research for the first time, we applied a glycomic approach to obtain O-glycan MS profiling of RCC and normal serum. A series of mass spectrum data processing was performed and spikes in MS data were analyzed. A combination of two MS peaks was chosen as a biomarker to differentiate between RCC and normal samples, and ensured optimal performance in our small sample size. In the future we will expand the sample size and also analyze the glycan structure.

*This work was supported by the National Natural Science Foundation of China (30900254). The efforts of Mr. Xu Yu in initiating and coordinating this joint research are also greatly acknowledged.*

- Gutwein P, Schramme A, Sinke N, et al. Tumoural CXCL16 expression is a novel prognostic marker of longer survival times in renal cell cancer patients. *Eur J Cancer*, 2009, 45: 478–489
- Baldewijns M M, van Vloderp I J, Schouten L J, et al. Genetics and epigenetics of renal cell cancer. *Biochim et Biophys Acta*, 2008, 1785: 133–155
- Zheng L G. Clinical value of IVU, CT and MRI in renal carcinoma. *Chin Med Equip J*, 2010, 8: 83–90
- Pantuck A J, Zisman A, Belldegrun A S. The changing natural history of renal cell carcinoma. *J Urol*, 2001, 166: 1611–1623
- Drucker B J. Renal cell carcinoma: current status and future prospects. *Cancer Treat Rev*, 2005, 31: 536–545
- Taniguchi N. Toward cancer biomarker discovery using the glycomics approach. *Proteomics*, 2008, 16: 3205–3208
- Ohtsubo K, Marth J D. Glycosylation in cellular mechanisms of health and disease. *Cell*, 2006, 126: 855–867
- Alper J. Glycobiology. Turning sweet on cancer. *Science*, 2003, 301: 159–160
- Villanueva J. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Proteome Res*, 2005, 4: 1060–1072
- Reis C A, Osorio H, Silva L, et al. Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol*, 2010, 63: 322–329
- Ishizuka A, Hashimoto Y, Naka R, et al. Accumulation of free complex-type N-glycans in MKN7 and MKN45 stomach. *Biochem J*, 2008, 413: 227–237
- Kirmiz C, Li B, An H J, et al. A serum glycomics approach to breast cancer biomarkers. *Mol Cell Proteomics*, 2007, 6: 43–55
- Kyselova Z, Mechref Y, Al Bataineh M M, et al. Alterations in the serum glycome due to metastatic prostate cancer. *J Proteome Res*,

- 2007, 6: 1822–1832
- 14 Perret A G, Clemencon A, Li G, et al. Differential expression of prognostic markers in histological subtypes of papillary renal cell carcinoma. *BJU Int*, 2008, 102: 183–187
  - 15 Aubert S, Fauquette V, Hémon B, et al. MUC1, a new hypoxia inducible factor target gene, is an actor in clear renal cell carcinoma tumor progression. *Cancer Res*, 2009, 69: 5707–5715
  - 16 Morelle W, Michalski J C. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc*, 2007, 2: 1585–1602
  - 17 Diamandis E P. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool. *Mol Cell Proteomics*, 2004, 3: 367–378
  - 18 Morelle W, Canis K, Chirat F, et al. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics*, 2006, 6: 3993–4015
  - 19 McLerran D F, Feng Z, Semmes O J, et al. Signal detection in high-resolution mass spectrometry data. *J Proteome Res*, 2008, 7: 276–285
  - 20 Zhang S, DeGraba T J, Wang H, et al. A novel peak detection approach with chemical noise removal using short-time FFT for prOTOF MS data. *Proteomics*, 2009, 9: 3833–3842
  - 21 Yang C, He Z, Yu W. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics*, 2009, 10: 4–17

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.