# Classification of proteomic data with multiclass Logistic Partial Least Squares algorithm

Zhenqiu Liu*

Division of Biostatistics,
Greenebaum Cancer Center,
University of Maryland Medicine,
22 South Greene Street, Baltimore, 21201 MD, USA
E-mail: zliu@umm.edu
*Corresponding author

Dechang Chen

Department of Preventive Medicine and Biometrics,
Uniformed Services,
University of the Health Sciences,
4301 Jones Bridge Road, Bethesda, 20814 MD, USA
E-mail: dchen@usuhs.mil

Jianjun Paul Tian

Mathematics Department,
College of William and Mary,
Williamsburg, 23187, VA, USA
E-mail: jtian@wm.edu

**Abstract:** Early detection of cancer is crucial for successful treatments. In this paper, we propose a multiclass Logistic Partial Least Squares (LPLS) algorithm for classification of normal vs. cancer using Mass Spectrometry (MS). LPLS combines the multiclass logistic regression with Partial Least Squares (PLS) algorithm. Wavelet decomposition is also proposed for pre-processing of original data. Wavelet decomposition and the proposed LPLS are applied to real life cancer data. Experimental comparisons show that LPLS with wavelet decomposition outperforms other methods in the analysis of MS data.

**Keywords:** Mass Spectrometry; MS; wavelet; logistic regression; Partial Least Squares; PLS; bioinformatics.

**Reference** to this paper should be made as follows: Liu, Z., Chen, D. and Tian, J.P. (2008) 'Classification of proteomic data with multiclass Logistic Partial Least Squares algorithm', *Int. J. Bioinformatics Research and Applications*, Vol. 4, No. 1, pp.1–10.

**Biographical notes:** Zhenqiu Liu, PhD, is an Assistant Professor at the Division of Biostatistics and Bioinformatics, Department of Epidemiology and Preventive Medicine and Greenebaum Cancer Center, The University of Maryland. His research interests include bioinformatics, statistical genetics and data mining. He has extensive experience in microarray, mass-spectrometry,

SNP, and DNA sequence data analysis. Currently he is concentrating on proteomics and dynamic pathway modelling in cancer research.

Dechang Chen is an Associate Professor at the Division of Epidemiology and Biostatistics, Department of Preventive Medicine and Biometrics, Uniformed Services, University of the Health Sciences, Maryland, USA. His research interests include applied statistics, bioinformatics, machine learning, ad hoc and sensor networks and differential equations.

Jianjun Paul Tian is a postdoctoral fellow at the Mathematical Biosciences Institute at the Ohio State University. His current research is focusing on mathematical and computational modelling of mechanism of disease occurrence and progression with experimental verification, and on computational statistics in study of cancer vs. health data.

## 1 Introduction

Proteins carry out and modulate the vast majority of chemical reactions which together constitute 'life'. Proteomics is an integral part of the process of understanding biological systems and uncovering disease mechanisms. Because of their high level of variability and complexity, it is an extremely challenging endeavour to conduct massive analysis of thousands of proteins. In the last decade, MS has increasingly become the method of choice for analysis of complex protein samples. MS measures two properties of ion mixtures in the gas phase under the vacuum environment: the mass-to-charge ratio ($m/z$) of ionised proteins in the mixture and the number of ions present at different $m/z$ values. The output is a chart with a series of spike peaks. The heights of peaks and the $m/z$ values of the peaks are a fingerprint of the sample. MS has not only been used intensively to identify proteins via peptide mass fingerprints, but also had promising applications in cancer classification (Petricoin et al., 2002a; Adam et al., 2002; Lilien, et al., 2003; Qu et al., 2002, 2003; Wu et al., 2003; Diamandis, 2003; Master, 2005). An important goal of cancer classification is to predict cancer on the basis of peptide/protein intensities.

While MS is increasingly used for protein profiles, significant challenges have arisen with regard to analysing the data. Specifically, MS data analysis may be divided into three steps: data pre-processing, feature selection, and classification. The critical pre-processing step includes baseline correction, peak identification and alignment, data normalisation and visualisation. The feature selection step extracts the relevant features and reduces the dimension of features greatly. The final step is the classification of disease status using the selected features. Recent publications on cancer classification with MS data have mainly focused on how to choose features for classification and which classification method is more accurate than others. In particular, test statistics, including T statistic, and Principal Component Analysis (PCA) have been used to select features (Chen et al., 2005; Levner, 2005). Classification methods such as Linear Discrimination (LD) analysis, *k*-nearest neighbour classification, decision trees (Adam et al., 2002), and support vector machines have been used to distinguish between cancer and normal samples (Qu et al., 2003; Lilien et al., 2003).

In this paper, we first discuss data pre-processing based on wavelet decomposition. We then propose a novel analysis procedure LPLS for feature selection and classification of MS data. LPLS combines the multiclass logistic regression with PLS regression

in a natural way. Experiments can be used to show that features derived from PLS usually provide more accurate predictions than those from PCA (Liu and Chen, 2004). The proposed algorithm LPLS can not only predict the class label but also provide the probability of each sample falling into a specific class. Wavelet decomposition and LPLS are assessed using two real life MS data sets.

This paper is organised as follows. We first discuss wavelet decomposition of original data and present the LPLS algorithm. We then describe computational results. And finally, we provide conclusions and remarks.

## 2 Wavelet decomposition for data pre-processing

A MS data set with n samples is a $p \times (n + 1)p$ matrix $(mz, X) = [mz, x_1, \ldots, x_n]$ where $p$ is the number of $m/z$ ratios, $mz$ is a column vector for the measured $m/z$ ratios, and $x_j$ are the corresponding intensities of the $j$th sample. Our goal is to predict the class of a sample based on its intensity profile $x$. As usual, such prediction often requires the task of data pre-processing. In the following, we propose one method based on the wavelet decomposition.

MS data have several special characteristics: the dimension of the data is large, the data points are not necessarily independent, and the measurements are usually more or less noisy. These problems motivate the use of compression techniques to describe the sequential data with a few features that capture the basic shape of the sequence. A wavelet transform is a way to decompose a signal in a chosen number of its constituent parts. Fourier analysis also has this property but wavelet analysis has some advantages when analysing signals of non-stationary nature. Wavelet provides more irregular shapes and the wavelet decomposition is a local one, so that if the information relevant to our prediction problem is constrained in a particular part or parts of the curve, as typically it is, this information will be carried in a very small number of wavelet coefficients. These properties make wavelets ideal for analysing signals with discontinuities and sharp changes while allowing temporal locating the features of the signal. Wavelets are families of functions that can accurately describe other functions in a parsimonious way. The signal is projected into the time frequency plane. The basis functions are $\Psi_{j,k}(t) = 2^{j/2}\Psi(2^j t - k)$, where $\Psi$ is the mother wavelet function. Any square integrable real function $f(t)$ can be represented in terms of bases as

$$f(t) = \sum_{j,k} c_{j,k} \Psi_{j,k}(t)$$

where $c_{j,k} = \langle \Psi_{j,k}(t), f(t) \rangle$ are the coefficients of the Discrete Wavelet Transform (DWT). There is a fast algorithm to get the coefficients with $O(n)$ time. A simple and commonly used wavelet is the Haar wavelet (Burrus et al., 1998) with the mother function

$$\Psi_{\text{Haar}}(t) = \begin{cases} 1, & \text{if } 0 < t < 0.5 \\ -1, & \text{if } 0.5 < t < 1. \\ 0, & \text{otherwise} \end{cases}$$

In this paper, we use the Haar wavelet because of its simplicity. One can use any other orthonormal wavelet basis functions and achieve similar results as we present here.

In wavelet decomposition the signal is separated successively into slow and fast components using a pair of finite impulse response filters. In the first stage the high pass and the low pass filters separate the signal into components above $f_s/4$ and components below $f_s/4$. The second stage receives the low frequency components as input and separates them into components above $f_s/8$ and components below $f_s/8$, and so on. The number of stages depends the slowest component that is desired. Each MS input $x_j$ can be treated as an original signal for MS data. We can normalise each MS sequence within [0, 1] by the formula

$$x = \frac{x - \min(x)}{\max(x) - \min(x)}$$

and then apply DWT. The obtained coefficients can be used as the features (Qu et al., 2003).

## 3 LPLS algorithm for feature selection and classification

Note that the size of feature matrix selected with DWT method is still around $p \times n$, but many entries are zeros or near zeros. Traditionally, some heuristic rules can be applied to reduce the feature dimension. For instance, we can either keep the first few coefficients or the largest coefficients of DWT as the features. However, both methods are just based on the signal itself and do not consider the associated class information. Our proposed LPLS algorithm combines the feature selection and classification together and choose features not only based on the data but also the class labels. LPLS is a combination of the multiclass logistic regression and PLS algorithm.

### *3.1 Classification based on multiclass logistic regression*

Logistic regression is one of the popular techniques for classification. Multiclass (multinomial) logistic regression is a generalisation of logistic regression. For a $m$-class problem, we represent the class labels of sample $z$ using a '1-of-$m$' encoding vector $y = [y^{(1)}, y^{(2)}, \ldots, y^{(m)}]$, such that $y^{(j)} = 1$ if $z$ belongs to class $j$ and $y^{(j)} = 0$ otherwise. Given the training data $D = \{(z_1, y_1), \ldots, (z_n, y_n)\}$, then under the multiclass logistic regression, we have the following conditional probability:

$$P(y^{(j)} = 1 \mid z, B) = \frac{\exp(\beta^{(j)T} z)}{\sum_{j=1}^{m} \exp(\beta^{(j)T} z)}$$

where matrix $B = [\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(m)}]$ represents the parameters and $T$ represents the transpose operation. Here each column $\beta^{(1)} = [\beta_{i1}, \ldots, \beta_{ik}]'$ is a parameter vector corresponding to one class. Classification is usually done in terms of the magnitudes of these conditional probabilities. For the binary case $m = 2$, this is known as a traditional logistic regression model. Learning the parameters $B$ is done through minimising the log likelihood:

$$L(B \mid D) = \sum_{i=1}^{n} \left[ \log\left( \sum_{j=1}^{m} \exp(\beta^{(j)T} z_i) \right) - \sum_{j=1}^{m} y_{ij} \beta^{(j)T} z_i) \right]. \quad (1)$$

Since the probabilities must sum to one

$$\sum_{j=1}^{m} p(y^j \,|z, B) = 1,$$

one of the parameter vectors $\beta^{(j)}$ needs not to be estimated. Therefore, we may set $\beta^{(j)} = 0$ without loss of generality. Define

$$Z_i = \sum_{j=1}^{m} \exp(\beta^{(j)T} z_i) \quad \text{and} \quad P_{ij} = \exp(\beta^{(j)T} z_i) / Z_i.$$

Then

$$\frac{\partial P_{ij}}{\partial \beta_{kl}} = z_{il} P_{ij} [\delta_{j=k} (1 - P_{ij}) - \delta_{j \neq k} P_{ik}]$$

$$\frac{\partial \log P(y^{(j)} = 1 | z, B)}{\partial \beta_{jp}} = \sum_{i=1}^{n} z_{ip} P_{ij} - \sum_{i|i=j} y_{ip} z_{ip}$$

$$\frac{\partial^2 \log P(y^{(j)} = 1 | z, B)}{\partial \beta_{jp} \partial \beta_{kl}} = \sum_{i=1}^{n} z_{ip} z_{il} P_{ij} [\delta_{j=k} (1 - P_{ij}) - \delta_{j \neq k} P_{ik}].$$

Given the derivatives, either the gradient decent or Newton's method can be utilised to find the maximum log likelihood estimator of B. Detailed algorithms of different implementations for multiclass (multinomial) logistic regression are given in Tipping (2001), Efron et al. (2004) and Krishnapuram et al. (2005).

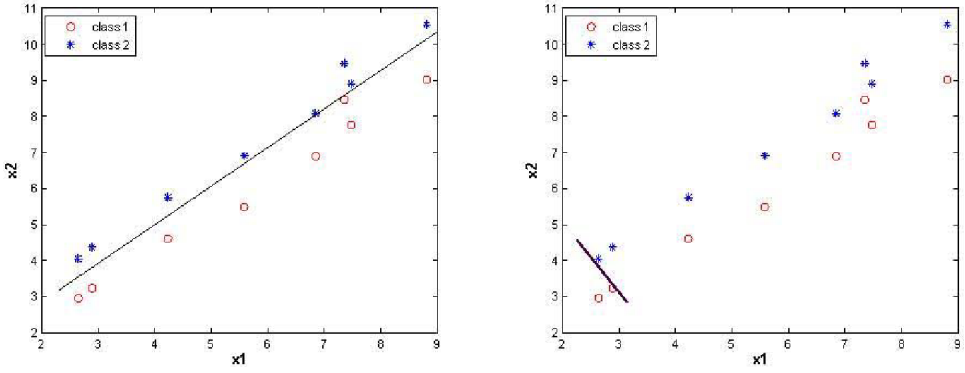
### *3.2 Dimension reduction based on PLS*

Logistic regression can not be applied to proteomic data directly, since the number of dimensions in proteomic data is far greater than its sample size. For the purpose of dimension reduction, PLS can be employed.

PLS is based on a linear transition from a large number of original descriptors to a new variable space formed by a small number of orthogonal factors (latent variables). This technique is especially useful in cases where the number of descriptors (independent variables) is comparable to or greater than the number of compounds (data points) and/or there exist other factors leading to correlations between variables. In these cases, the solution of classical least squares does not exist or is unstable and unreliable. On the other hand, the PLS approach usually leads to stable, correct and highly predictive models even for correlated descriptors (Martens and Naes, 1991). The latent variables are mutually independent (orthogonal) linear combinations of original descriptors. Unlike PCA, latent variables are chosen in such a way as to provide maximum correlation with the dependent variable. Thus, the PLS model contains the smallest necessary number of factors. With increasing number of factors, the PLS model converges to an ordinary multiple linear regression model (if one exists). In addition, the PLS approach allows one to detect relationship between activity and descriptors even if key descriptors have little contribution to the first few principal components. Figure 1 is used to illustrate the concept. In the figure, $x_1$ and $x_2$ are two independent variables. The first latent component given by PCA and PLS is the line in the right and left panel, respectively. The figure shows that even for this simple problem, PCA selects the poor

latent variable which can not be used to separate the two classes, because it only utilises the input feature matrix, while the PLS component make use of both input and output information and can be used to separate the two classes efficiently.

**Figure 1** The first latent component given by PLS (left) and PCA (right)




### 3.3 LPLS algorithm

Given a training dataset $\{z_i\}_{i=1}^{n}$ with class labels $\{y_i\}_{i=1}^{n}$ and a test dataset $\{z_t\}_{t=1}^{n_t}$ with labels $\{y_t\}_{t=1}^{n_t}$, the LPLS algorithm is described as follows:

1 Set matrix $Z = [z_i]$ for the training data with the label matrix $Y$, and the matrix $z_t = [z_t]$ for the test data.

2 Call PLS algorithm to find $k$ component directions (Rosipal and Trejo, 2001)

- a for $i = 1, \ldots, k$
- b initialise $u^i$
- c $w^i = Z'u^i$
- d $t^i = Zw^i$
- e $c^i = Y't^i$
- f $u^i = Yc^i$, $u^i \leftarrow u^i/\|u^i\|$
- g repeat steps (b)–(f) until convergence
- h deflate $Z$, $Y$ by $Z \leftarrow t^i t^{i'} Z$ and $Y \leftarrow Y - t^i t^{i'} Y$
- i obtain component matrix $W = [w^1, \cdots, w^k]$.

3 Find the projections $V = ZW$ and $V_{te} = Z_t W$ for the training and test data, respectively.

4 Build a logistic regression model using $V$ and $\{y_i\}_{i=1}^{n}$ and test the model performance using $V_{te}$ and $\{y_t\}_{t=1}^{n_t}$.

The dimension of projection (the number of components) $k$ used in the model can be selected using Akaike's Information Criteria (AIC):

$$\mathrm{AIC} = -2L(B/D) + 2(k+1), \tag{2}$$

where $L$ is the log likelihood. The log likelihood $L$ can be calculated using equation (1). We choose the best k that minimises AIC.

It is straightforward to extend LPLS using kernel functions. However, our experiments showed that the classification accuracy with the nonlinear version of LPLS did not improve significantly for MS data. The proposed algorithm LPLS is designed for multiclass classification and is efficient in dealing with proteomic data.

## 4 Experimental results

### *4.1 Ovarian cancer*

First we evaluate the performance of the proposed algorithm on the ovarian cancer data. This cancer dataset was downloaded directly from the website: http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp. The sample set includes 91 controls and 162 ovarian cancer cases. To evaluate the performance of the proposed methods, we merged the control and cancer data together and split the data with a ten-fold validation scheme. The data were divided randomly into ten roughly equal subsets, and then we applied the algorithm ten times, each time with nine subsets used for training and the remaining subset for performance evaluation. The averaged error over the ten times was reported as an overall performance. One output is given in Table 1, where *T*-test statistic, viewed as a pre-processing procedure applied to the original data, and DWT were compared. The number of features in Table 1 is the number of components used in the LPLS algorithm. This table clearly shows that LPLS performs better with DWT pre-processing method. Petricoin et al. (2002a) achieved a performance comparable to that of DWT on a slightly different ovarian data.

**Table 1** Performance of LPLS on ovarian cancer data with different data pre-processing methods

| *T-test* | | *DWT* |
|---|---|---|
| No. of features | 30 | 10 |
| Test error (%) | 1.75 ± 1.4 | 0 ± 0 |
| Sensitivity (%) | 99.03 ± 1.48 | 100 ± 0 |
| Specificity (%) | 96.64 ± 2.19 | 100 ± 0 |

After DWT data pre-processing, performances of different feature selection and classification methods are given in Table 2, where Fisher LD, *k* Nearest Neighbour (KNN), and neural networks were employed. This table indicates that LPLS, and PCA and PLS, both viewed as feature selection methods, can lead to the 100% accuracy. It is seen that neural networks have the worst performance.

**Table 2** Performance of LPLS on ovarian cancer data of different feature selection and classification methods

| *Feature and classification methods* | *Test accuracy* (%) |
|---|---|
| PCA and LD | 100 |
| PCA and logistic regression | 100 |
| PCA and KNN | 99.7 |
| PCA and neural network (15 nodes) | 99.5 |
| PLS and LD | 100 |
| PLS and KNN | 100 |
| PLS and neural network | 99.7 |
| LPLS | 100 |

### *4.2 Prostate cancer*

The prostate cancer data were downloaded from the same website as the ovarian data. The Surface Enhanced Laser Desorption/Ionisation (SELDI) time of flight method and a mass spectra analysis for this data set have been performed in Petricoin et al. (2002b). SELDI process is a relatively new medical technique that measures the content of different proteins in blood samples from patients. This dataset consists of four subsets:

- 63 samples with no evidence of disease and the Prostate Specific Antigen (PSA) level less than 1(ng/ml)
- 190 samples with benign prostate and PSA level greater than 4
- 26 samples with prostate cancer and PSA levels between 4 and 10
- 43 samples with prostate cancer and PSA levels greater than 4.

There are 322 samples in total and we treated them as coming from three classes: normal, benign, and cancer. Again, the ten-fold validation was used for the experiments. The 'one against all others' scheme was applied to separate each class against the other two. The experimental results are given in Tables 3 and 4. These results show that DWT performs better than *T* test statistic in data pre-processing and that after DWT is used for data pre-processing, LPLS gives a higher prediction accuracy than any other feature selection and classification method.

**Table 3** Performances of LPLS on prostate cancer data of different data and pre-processing methods

| *T-test* | | *DWT* |
|---|---|---|
| No. of features | 53 | 28 |
| Test error (%) | 11.8 ± 2.8 | 1.7 ± 1.4 |
| Sensitivity (%) | 87.1 ± 2.94 | 98.5 ± 1.3 |
| Specificity healthy (%) | 96.8 ± 2.69 | 100 ± 0 |
| Specificity benign (%) | 83.1 ± 1.72 | 94.7 ± 3.58 |

**Table 4** Performances of LPLS on prostate cancer data of different feature selection and classification methods

| *Feature and classification methods* | *Test accuracy* (%) |
|---|---|
| PCA and LD | 95.3 |
| PCA and logistic regression | 96.8 |
| PCA and KNN | 91.9 |
| PCA and neural network (15 nodes) | 93.5 |
| PLS and LD | 96.3 |
| PLS and KNN | 93.6 |
| PLS and neural network | 95.9 |
| LPLS | 98.3 |

## 5 Conclusion

Our limited experiments with two cancer datasets show that the proposed LPLS algorithm coupled with DWT data pre-processing procedure is promising in analysing MS data. The pre-processing of MS output is a crucial step in the overall analysis of MS data. Our proposed DWT for data pre-processing worked well for the two datasets investigated in this study. Feature selection (dimension reduction) and classification constitute two more steps in analysis. For these, LPLS, a combination of PLS and logistic regression, showed superior performance on the two datasets when compared with other feature selection and classification methods.

Though there are recent debates in the literature regarding the reproducibility of MS data (Baggerly et al., 2004, 2005; Master, 2005), we believe that the proposed methods in this paper have their own values as they can be implemented easily and can have high prediction accuracies for data with high quality.

## Acknowledgements

D. Chen was supported by the National Science Foundation grant CCR-0311252. J.P. Tian would like to acknowledge the grant support by the National Science Foundation agreement 0112050.

## References

Adam, B.L., Qu, Y., Davis, J.W., Ward, M.D., Clements, M.A., Cazares, L.H., Semmes, O.J., Schellhammer, P.F., Yasui, Y., Feng, Z. and Wright Jr., G.L. (2002) 'Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men', *Cancer Res*., Vol. 62, pp.3609–3614.

Baggerly, K.A., Morris, J.S. and Coombes, K.R. (2004) 'Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments', *Bioinformatic*s, Vol. 20, pp.777–785.

Baggerly, K.A., Morris, J.S., Edmonson, S.R. and Coombes, K.R. (2005) 'Signal in noise: evaluating reported reproducibility of serum proteomic test for ovarian cancer', *Journal of National Cancer Institute*, Vol. 97, pp.307–309.

Burrus, C.S., Gopinath, R.A. and Guo, H. (1998) *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice-Hall, Inc., Upper Saddle River, New Jersey 07458, USA.

Chen, D., Liu, Z., Ma, X. and Hua, D. (2005) 'Selecting genes by test statistics', *Journal of Biomedicine and Biotechnology*, Vol. 2, pp.132–138.

Diamandis, E.P. (2003) 'Proteomic patterns in biological fluids: Do they represent the future of cancer diagnostics?', *Clin. Chem.*, Vol. 49, pp.1272–1275.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) 'Least angle regression', *Ann. Statist.*, Vol. 32, No. 2, pp.407–499.

Jaeger, J., Sengupta, R. and Ruzzo, W.L. (2003) 'Improved gene selection for classification of microarrays', *Pacific Symposium on Biocomputing*, Kauai, Hawaii, Vol. 8, pp.53–64.

Krishnapuram, B., Figueiredo, M., Carin, L. and Hartemink, A. (2005) 'Sparse multinomial logistic regression: fast algorithms and generalization bounds', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, pp.957–968.

Levner, I. (2005) 'Feature selection and nearest centroid classification for protein mass spectrometry', *BMC Bioinformatics*, Vol. 6, p.68.

Lilien, R.H., Farid, H. and Donald, B.R. (2003) 'Probabilistic disease classification of expression-dependent proteomic data from mass spectrometry of human serum', *Journal of Computational Biology*, Vol. 10, No. 6, pp.925–946.

Liu, Z. and Chen, D. (2004) 'Gene expression data classification with revised kernel partial least squares algorithm', *Proceedings of the 17th International FLAIRS Conference*, South Beach, Florida, USA, pp.104–108.

Martens, H. and Naes, T. (1991) *Multivariate Calibration*, Wiley, Chichester.

Master, S.R. (2005) 'Diagnostic proteomics: Back to basics?', *Clinical Chemistry*, Vol. 51, pp.1333, 1334.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C. and Liotta, L.A. (2002a) 'Use of proteomic patterns in serum to identify ovarian cancer', *Lancet*, Vol. 359, No. 9306, pp.572–577.

Petricoin, E.F., Ornstein, D.K., Paweletz, C.P., Ardekani, A., Hackett, P.S., Hitt, B.A., Velassco, A., Trucco, C., Wiegand, L., Wood, K., Simone, C.B., Levine, P.J., Linehan, W.M., Emmert-Buck, M.R., Steinberg, S.M., Kohn, E.C. and Liotta, L.A. (2002b) 'Serum proteomic patterns for detection of prostate cancer', *J. Natl. Cancer Inst.*, Vol. 94, pp.1576–1578.

Qu, Y., Adam, B.L., Yasui, Y., Ward, M.D., Cazares, L.H., Schellhammer, P.F., Feng, Z., Semmes, O.J. and Wright Jr., G.L. (2002) 'Boosted decision tree analysis of SELDI mass spectral serum profiles discriminates prostate cancer from non-cancer patients', *Clinical Chemistry*, Vol. 48, pp.1835–1843.

Rosipal, R. and Trejo, L.J. (2001) 'Kernel partial least squares regression in reproducing kernel Hilbert space', *Journal of Machine Learning Research*, Vol. T2, December, pp.97–123.

Tipping, M.E. (2001) 'Sparse Bayesian learning and the relevance vector machine', *Journal of Machine Learning Research*, Vol. 1, pp.211–244.

Wu, B., Abbott, T., Fishman, D., McMurray, W., Mor, G., Stone, K., Ward, D., Williams, K. and Zhao, H. (2003) 'Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data', *Bioinformatics*, Vol. 19, No. 13, pp.1636–1643.