Analyst

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

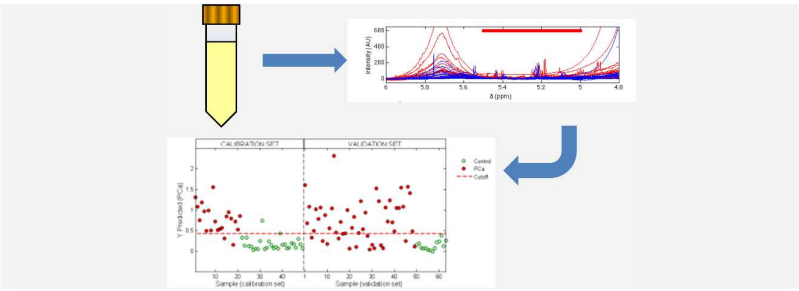
www.rsc.org/analyst

Entry for the Table of Contents

Catch Phrase

Patricia Zaragozá, Jose Luis Ruiz-Cerdá, Guillermo Quintás, Salvador Gil, Ana M. Costero, Zacarías León, José-Luis Vivancos, Ramón Martínez-Máñez

An multivariate approach based on ^{1}H NMR spectra profiles of urine samples to detect patients with prostate cancer



Journal Name

RSCPublishing

COMMUNICATION

Towards the Potential use of ^{1}H NMR Spectroscopy in Urine Samples for Prostate Cancer Detection

Cite this: DOI: 10.1039/xoxxoooooox

Patricia Zaragozá,[a] Jose Luis Ruiz-Cerdá,[b] Guillermo Quintás,[c] Salvador Gil,[a,d] Ana M. Costero,[a,d] Zacarías León,[e] José-Luis Vivancos,[a,f,g] Ramón Martínez-Máñez[a,g,h]

Received ooth January 2014,
Accepted ooth January 2014

DOI: 10.1039/xoxxoooooox

www.rsc.org/

A simple method based in multivariate analysis of ^{1}H NMR spectra profiles of urine samples can be used to detect patients with prostate cancer.

Prostate cancer (PCa) is the most frequent noncutaneous malignancy in the male population and only in the United States, it has an incidence of 241,740 cases per year.[1] Although prostate specific antigen (PSA) blood testing remains the most widely used tool for PCa detection, this suffers from a number of problems, including low specificity, inability to specify a cut-point below which cancer is unlikely, non-trivial false-negative rate for prostate biopsy, over-diagnosis and over-treatment of relatively indolent tumours with low potential for morbidity or death if left untreated. In this context, important efforts have been conducted in the last decade to investigate new biomarkers in blood or urine to improve early PCa detection and risk prediction.[2]

On the other hand it is known that neoplastic transformation necessitates metabolic alterations to provide the bioenergetic and synthetic requirements of malignancy[3]. It is on this basis that alterations to the metabolic signatures within biofluids or tissues may reflect changes in phenotype and function and are the key to differentiating tumors from normal tissue. In fact the study of metabolites has recently emerged as a novel discipline to the discovery of clinical relevant cancer biomarkers and pathways associated with numerous cancers including PCa[4]. In most cases this approach involves the study of global variations of metabolites in an attempt to evaluate malignant conditions by profiling the entire measurable metabolome instead of focusing only on certain metabolites or on isolated metabolic pathways.[5]

In the particular case of PCa, metabolites for PCa detection have been measured using a large variety of techniques in a range of different samples including tissue extracts, prostatic fluid, serum and urine. Following this approach distinct single metabolites and metabolite ratios have long been investigated as biomarker candidates in PCa.[6] Although these studies have increased the biological understanding of PCa, they do not provide a holistic picture of the malignant status. Moreover there is not a single metabolite that can be directly related with the presence of PCa. Moreover the current consensus is that entire metabolic profiles are more sensitive in identifying and characterizing prostate cancer.[7] Additionally, most of these studies have been reported in prostate tissues extracts[8] and, indeed it is apparent from the literature that studies on body fluids, especially urine, in relation to PCa are very rare.

However the potential use of simple urine samples to detect PCa is highly appealing as urine is collected non-invasively and requires minimal sample preparation prior to analysis.[9] Detection of PCa-related metabolites in urine relies on the presence of products that are release either directly into the urine or carried within prostatic cells that are shed into urine.[10]

Among different analytical techniques, Nuclear Magnetic Resonance (NMR) spectroscopy is a typical method used in the characterization of metabolites. NMR is highly quantitative and reproducible and it has the advantage of requiring minimal sample preparation and it is a rapid, single-step procedure with a high specificity. In fact ^{1}H NMR provides a number of peaks which can be related with the presence of organic compounds in solution typically in the mM range.[11] In this context, although there is a number of reports in the literature on *in vivo* Magnetic Resonance Imaging (MRI)[12,13] and Magnetic Resonance Spectroscopy (MRS)[6,13] studies of PCa, there are a very limited number of ^{1}H NMR studies in *ex vivo* body fluid samples such as urine.

Based on these concepts, we focused our attention on the potential use of urine as suitable fluid and ^{1}H NMR as an easy technique for PCa detection. The aim of this work was not to perform a complete analysis of metabolites (that will be carried out in due course) but to assess the prospective use of ^{1}H NMR spectra profiles of urine samples combined with multivariate analysis as a simple diagnostic tool for the potential correct classification of patients with PCa.

For this work, a total of 113 samples were used and split into calibration and validation subsets. The set of patients with PCa contained both, patients with tumours that were further operated to remove the prostate gland (radical prostatectomy) and patients with very low risk tumours with minimal percentage of tumour in biopsy that were left untreated and that were included in an active surveillance program. As control (without PCa), patients after radical prostatectomy and patients

diagnosed benign prostatic hyperplasia (BPH) were used. The randomly selected calibration subset included a total of 49 samples collected from 21 patients with PCa and a total of 28 control samples (17 after radical prostatectomy and 11 diagnosed with BPH). The validation set was formed by 64 samples including 50 PCa samples, and 14 samples classified as control (9 after radical prostatectomy and 5 diagnosed with BPH). All samples were collected at La Fe Hospital, Valencia (Spain). The collected urine samples were frozen and stored at −80 ^{o}C until analyses. Samples were centrifuged at 2500 rpm for 5 minutes to eliminate solids and other insoluble material, and then aliquoted. ^{1}H NMR studies, with a presaturation sequence on the water signal, was recorded in the urine samples using a Bruker DRX-500 provided with a BBOF prove (^{1}H (500 MHz)).

NMR spectra acquired were imported into MATLAB®. The interval correlation shifting (icoshift) algorithm developed by Savorani et al.[14] was used for initial spectral alignment to overcome shifts of pH dependent signals found in the data set. The icoshift algorithm aligns each NMR feature to a target (in this work, the median spectrum of the whole spectral data set) by maximizing the cross correlation between user defined intervals. Here, the NMR spectra were split into 51 intervals selected after visual inspection of the regions according to common spectral features among samples. Results from the alignment are depicted in Figure 1. For multivariate analysis, the spectral intervals highlighted in Figure 1 were used.

Initially, a principal component analysis (PCA) model was built using the calibration set and auto-scaling as data pre-treatment. From the scores plot of the first versus the second principal components and from the Q-residual values (see Supplementary Information for details), no samples included on the calibration set were classified as outliers.

The calibration set was used for model development and feature selection. Supervised discriminant analysis was performed using partial least squares (PLSDA) and a maximum number of 5 latent variables (LVs). The X-block (i.e. NMR data) was auto-scaled and the y vector containing class labels (i.e. -1 and +1 for control and PCa samples, respectively) was mean centred.

Initial PLSDA figures of merit were obtained by after 11 iterations of a random 5-fold cross validation. From cross validation data, 3 latent variables were retained. Then, a selection of the most differentiating spectral features was carried out based on the variable importance scores vector (VIP) calculated from the initial PLSDA model. Since the average of the squared VIP scores equals 1, the greater than 1 criteria is typically used as a rule of thumb for variable elimination.[15] The effect of using VIP cutoff values in the 0-7 range was evaluated by leave one out cross validation using the discriminant Q^2 (dQ^2) statistic and the number of misclassified (NMC) samples as target. The evolution of the dQ^2 and NMC values as a function of the VIP cutoff shown in Figure 2 indicated that although a first maximum was found at VIP ≈ 1, a further reduction in the number of retained features did not negatively affect the predictive performance of the model. Therefore, based on a VIP cutoff value of 2.28 a total of 1627 variables were retained and used for the calculation of a second PLSDA model (see Supplementary Information for details).

In that second PLSDA model two latent variables were used based on results obtained after 11 iterations of a random 5-fold cross validation (data not shown). In order to avoid overoptimistic results[16], the statistical significance of the model was evaluated using the external validation set which consisted in a total of 64 samples of which 50 were urine samples from patients with PCa and 14 samples that were used as control. The results are summarized in Table 1, whereas predicted values are also shown in Figure 3. In this testing phase the use of ^{1}H NMR spectra in urine samples was able to correctly classify 36 samples out of 50 for patients with PCa, whereas it classifies 14 samples out of 14 correctly from control patients. This results in a sensitivity of 72% and a specificity of 100%.

Moreover, a closer look at the data showed that 57% of the samples that were predicted as false negatives corresponded to patients that fulfilled criteria of indolent prostate cancer in biopsy which have minimal prostatic gland involvement. In this context, if only patients bearing prostate cancer with indication of active treatment were taken into account in the validation set, a sensitivity of 79% was reached.

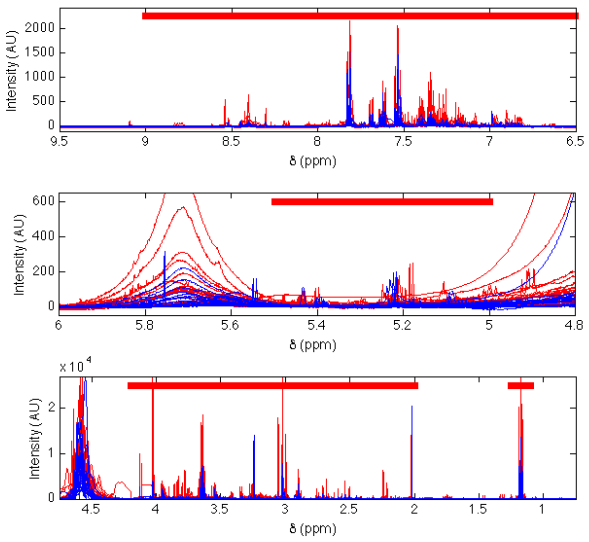


Figure 1. Calibration spectra after icoshift alignment. For a better visualization only the PCa (red lines) and control (blue lines) samples included in the calibration spectra are depicted. Spectral regions used for PLSDA are indicated by the red bar included in each subplot.

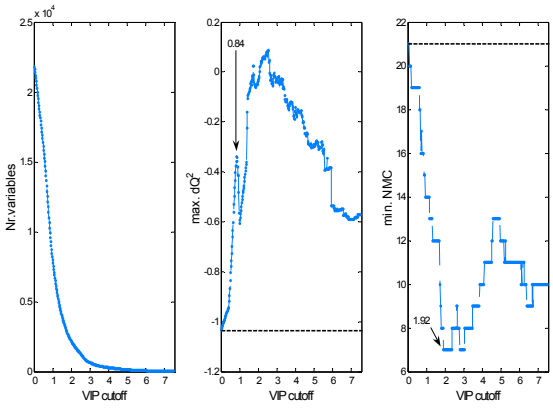


Figure 2. Effect of the VIP cutoff level on the number of retained variables (left), the discriminant Q^2 statistic (middle) and the number of misclassified samples (NMC) estimated by leave-one-out cross validation in the calibration set (right).

We are aware that further analysis need to be completed related with the characterization of metabolites in urine samples of patients with PCa. However it was apparent from preliminary studies that the presence of PCa could not be

related with the presence of a unique analyte (unique set of signals in the NMR spectra) but is most likely linked to the presence, or changes in the concentration, of a certain number of metabolites. In particular, we developed a PLSDA model using the interval covering a set of metabolites that have been shown to be related to PCa in a previous work.[17] The set of metabolites included: myo-inositol, phosphocholine, spermine, citrate, glutamine, spermine, alanine, lactate, OH-butyrate and Valine-Leucine. However this procedure did not improve the results shown in Table 1 (see Supplementary Information for details). A more complete metabolomics study is being currently carried out with the additional help of mass spectrometry (MS) studies. Moreover it is noteworthy that potential biases related with food or drug consumption or associated with other diseases have not been taken into account in our study. Despite this, ^{1}H NMR profiles of urine samples appear as a powerful tool for PCa prediction.

Table 1. Cross validation and external validation results obtained by PLSDA using 1627 variables retained after the initial variable selection.

	CV		Validation set	
	PCa	Control	PCa	Control
Predicted as PCa	19	2	36	0
Predicted as control	2	26	14	14

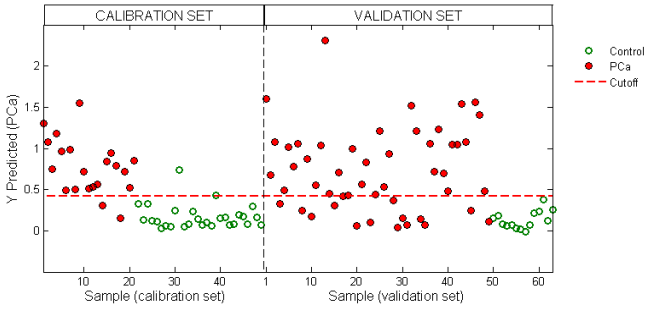


Figure 3. Predicted y values by PLSDA using 1627 retained variables, 2 LVs and autoscaling as X-block pretreatment.

Conclusions

Metabolite profiles with potential relevance to PCa biology have been traditionally identified using in vivo ^{1}H MRS and ^{1}H NMR in tissue extracts. In contrast studies in body fluids related with PCa, especially urine, are very rare. In this communication we have shown that application of multivariate analysis to ^{1}H NMR spectra profiles of urine samples can be a suitable and promising method for PCa detection. In particular 113 urine samples were used and split into calibration and validation subsets. In the validation set the model correctly classifies 36 samples out of 50 for patients with PCa, and 14 samples out of 14 correctly from control patients. These results suggest that ^{1}H NMR data from urine samples can be a potential easy-to-use non-invasive tool for the evaluation of PCa. In a wider context, given the relatively low cost and easy of ^{1}H NMR and the expected parallel advances being made in magnetic resonance equipment and in the automatization of multivariate statistical analysis, we believe that this simple approach of using ^{1}H NMR profiles of urine as "fingerprints" has great potential for applications in the detection of certain clinically significant diseases (i.e. PCa) and for combining diagnosis and further simple monitoring after therapy.

Financial support from the Spanish Government (Project MAT2012-38429-C04), the IIS La Fe (Project 2013_0509) and the Generalitat Valenciana (Project PROMETEO/2009/016) is gratefully acknowledged. P.Z. is grateful to the IDM for her grant. SCIE from the Universitat de Valencia is gratefully acknowledged for NMR facilities.

Notes and references

[a] Centro de Reconocimiento Molecular y Desarrollo Tecnológico (IDM), Unidad Mixta Universitat Politècnica de València - Universitat de València.
[b] Servicio de Urología, Hospital Universitario y Politécnico La Fe, Valencia, España
[c] Leitat Technological Center, Bio in vitro Division, Valencia, Spain.
[d] Departamento de Química Orgánica, Facultad de Químicas, Universidad de Valencia, Doctor Moliner 50, 46100 Burjassot, Valencia, Spain
[e] Unidad Analítica, Instituto de Investigación Sanitaria – Fundación Hospital La Fe, Valencia, Spain
[f] Departamento de Proyectos de Ingeniería, Universitat Politècnica de València, Camino de Vera s/n., 46022 Valencia, Spain.
[g] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BNN), Spain.
[h] Departamento de Química, Universitat Politècnica de València, Camino de Vera s/n., 46022 Valencia, Spain.
Electronic Supplementary Information (ESI) available: details of the set of samples, NMR experimental data and data analysis. See DOI: 10.1039/c000000x/

1 S. Rebecca, D. Naishadham and A. Jemal, *CA: A Cancer Journal for Clinicians*, 2012, **62**, 10.
2 W.J. Catalona, A.V. D'Amico, W.F. Fitzgibbons, O. Kosoko-Lasaki, S.W. Leslie, H.T. Lynch, J.W. Moul, M.S. Rendell and P.C. Walsh, , *Annals of Internal Medicine*, 2012, **157**, 137; V.A. Moyer, *Annals of Internal Medicine*, 2012, **157**, 120.
3 M.J. Roberts, H.J. Schirra, M.F. Lavin and R.A. Gardiner, Metabolomics: a novel approach to early and noninvasive prostate cancer detection, *Korean Journal of Urology,* 2011, **52**, 79.
4 J.L. Spratlin, N.J. Serkova and S.G. Eckhardt, Clinical applications of metabolomics in oncology: a review, *Clinical Cancer Research,* 2009, **15**, 431.
5 L.L. Cheng and U. Pohl, *The handbook of metabonomics and metabolomics*, Amsterdam, 2007, 345; B.J. Trock, *Urologic Oncology: Seminars and Original Investigations,* 2011, **29**, 572.
6 M.G. Swanson, D.B. Vigneron, Z.L. Tabatabai, R.G. Males, L. Schmitt, P.R. Carroll, J.K. James, R.E. Hurd and J. Kurhanewicz, *Magnetic Resonance in Medicine,* 2003, **50**, 944; L.L. Cheng, M.A. Burns, J.L. Taylor, W. He, E.F. Halpern, W.S. McDougal and C.L., *Cancer Research,* 2005, **65**, 3030.
7 E.M. Spur, E.A. Decelle and L.L. Cheng, *European Journal of Nuclear Medicine and Molecular Imaging*, 2013, **40(S1)**, 60.
8 A. Sreekumar, L.M. Poisson, T.M. Rajendiran, A.P. Khan, Q. Cao, J. Yu, B. Laxman, R. Mehra, R.J. Lonigro, Y. Li, M.K. Nyati, A. Ahsan, S. Kalyana-Sundaram, B. Han, X. Cao, J. Byun, G.S. Omenn, D. Ghosh, S. Pennathur, D.C. Alexander, A. Berger, J.R. Shuster, J.T. Wei, S. Varambally, C. Beecher and AM.Chinnaiyan, *Nature*, 2009, **457**, 910.
9 M. Truong, B. Yang and D. Jarrard, *Journal of Urology*, 2013, **189**, 422.
10 R.R. Drake, K.Y. White, T.W. Fuller, E. Igwe, M.A. Clements, J.O. Nyalwidhe, R.W. Given, R.S. Lance and O.J. Semmes, *Journal of Proteomics*, 2009, **72**, 907.
11 V. Kumar, D.K. Dwivedi and N.R. Jagannathan, *NMR in Biomedicine*, 2014, **27**, 80.
12 S.A. Reinsberg, G.S. Payne, S.F. Riches, S. Ashley, J.M. Brewster, V.A. Morgan, and N.M. de Souza, *American Journal of Roentgenology*, 2007, **188**, 91.
13 H Hricak, *The British Journal of Radiology*, 2005, **78**, S103; V. Panebianco, A. Sciarra, A. Marcantonio, V Forte, T. Biondi, A. Laghi, and C. Catalano, *The quarterly journal of nuclear medicine and molecular imaging*, 2012, **56(4)**, 331.
14 F. Savorani, G. Tomasi and S.B. Engelsen, *Journal of Magnetic Resonance,* 2010, **202**, 190.
15 I.G. Chong and C.H. Jun, *Chemometrics and Intelligent Laboratory Systems*, 2005, **78**, 103.
16 P. Filzmoser, B. Liebmann and K. Varmuza, *Journal of Chemometrics*, 2009, **23**, 160; J. Kuligowski, D. Pérez-Guaita, J.

Escobar, M. de la Guardia, M. Vento, A. Ferrer and G. Quintás, *Talanta*, 2013, **116**, 835; C.M. Rubingh, S. Bijlsma, E.P.P.A. Derks, I. Bobeldijk, E.R. Verheij, S. Kochhar and A.K. Smilde, *Metabolomics*, 2006, **2**, 53.

17 N.J. Serkova, E.J. Gamito, R.H. Jones, C. O'Donnell, J.L. Brown, S. Green, H. Sullivan, T. Hedlund and E.D. Crawford, *Prostate,* 2008, **68**, 620.