

# NMR-Based Metabolomic Profiling of Urine: Evaluation for Application in Prostate Cancer Detection

Natural Product Communications  
May 2019: 1–13  
© The Author(s) 2019  
Article reuse guidelines:  
[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)  
DOI: 10.1177/1934578X19849978  
[journals.sagepub.com/home/npx](http://journals.sagepub.com/home/npx)



Neil MacKinnon<sup>1,2</sup>, Wencheng Ge<sup>2</sup>, Peisong Han<sup>3</sup>, Javed Siddiqui<sup>4</sup>, John T. Wei<sup>5,6</sup>, Trivellore Raghunathan<sup>3,7</sup>, Arul M. Chinnaiyan<sup>4,5,6,8</sup>, Thekkelnaycke M. Rajendiran<sup>4</sup>, and Ayyalusamy Ramamoorthy<sup>1,2</sup>

## Abstract

Detection of prostate cancer (PCa) and distinguishing indolent versus aggressive forms of the disease is a critical clinical challenge. The current clinical test is circulating prostate-specific antigen levels, which faces particular challenges in cancer diagnosis in the range of 4 to 10 ng/mL. Thus, a concerted effort toward building a noninvasive biomarker panel has developed. In this report, the hypothesis that nuclear magnetic resonance (NMR)-derived metabolomic profiles measured in the urine of biopsy-negative versus biopsy-positive individuals would nominate a selection of potential biomarker signals was investigated. <sup>1</sup>H NMR spectra of urine samples from 317 individuals (111 biopsy-negative, 206 biopsy-positive) were analyzed. A double cross-validation partial least squares-discriminant analysis modeling technique was utilized to nominate signals capable of distinguishing the two classes. It was observed that after variable selection protocols were applied, a subset of 29 variables produced an area under the curve (AUC) value of 0.94 after logistic regression analysis, whereas a “master list” of 18 variables produced a receiver operating characteristic (ROC) AUC of 0.80. As proof of principle, this study demonstrates the utility of NMR-based metabolomic profiling of urine biospecimens in the nomination of PCa-specific biomarker signals and suggests that further investigation is certainly warranted.

## Keywords

urinary biomarkers, prostate cancer, <sup>1</sup>H NMR, metabolomics

Received: November 4th, 2018; Accepted: January 18th, 2019.

Prostate cancer (PCa) is a disease that is estimated to be diagnosed in 164 690 American men in 2018.<sup>1</sup> As a disease that typically occurs late in life, morbidity rates are low, especially given that there are forms of PCa malignancies that progress slowly. However, this presents a critical challenge to clinicians responsible for diagnosing and delivering prognoses since there is a current lack of clinical standard for predicting aggressive versus indolent forms of PCa. The current clinical screening test is circulating serum prostate-specific antigen (PSA); however, there are several issues associated with the PSA test.<sup>2–5</sup> First is the low specificity of PCa, with multiple conditions associated with prostate swelling also resulting in elevated PSA levels.<sup>6</sup> Second is the sensitivity, with a clinically challenging range of 4 to 10 ng/mL that is difficult to diagnose as aggressive or indolent PCa. As a result of these drawbacks, there have been recent reports suggesting that PSA testing has potentially done greater harm than good, primarily through overdiagnosis and the associated deleterious treatment side effects.<sup>7–9</sup> Thus, there is

a critical requirement for clinically relevant screening tests

<sup>1</sup> Biophysics, University of Michigan, Ann Arbor, MI, USA

<sup>2</sup> Department of Chemistry, University of Michigan, Ann Arbor, MI, USA

<sup>3</sup> Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

<sup>4</sup> Michigan Center for Translational Pathology, Department of Pathology, University of Michigan, Ann Arbor, MI, USA

<sup>5</sup> Department of Urology, University of Michigan Medical School, Ann Arbor, MI, USA

<sup>6</sup> Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA

<sup>7</sup> Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

<sup>8</sup> Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI, USA

## Corresponding Author:

Neil MacKinnon, Institute of Microstructure Technology, Karlsruhe Institute of Technology, Karlsruhe, BW, Germany.  
Email: [neil.mackinnon@kit.edu](mailto:neil.mackinnon@kit.edu)



capable of distinguishing aggressive and indolent forms of PCa with high specificity.

Biomarker discovery in relation to PCa continues to be an active area of research, and several recent reviews are available on the topic.<sup>10–16</sup> In many cases, biomarkers are discovered via an “omics” approach, and thus there are candidate molecules representative of the genetic, protein, and metabolite levels. The standard clinical test relies on circulating PSA levels, a protein that exists both in free and bound forms. Thus in an effort to refine PSA testing, both fractions have been included, with the amount of free PSA (%fPSA) identified as an improved indicator.<sup>17</sup> Further examples of candidate protein markers include  $\alpha$ -methylacyl-CoA racemase<sup>18,19</sup> and anterior gradient 2.<sup>20</sup> Several peptides have also been nominated in urine as potential biomarkers of PCa, although care must be taken in sample collection and preparation.<sup>21</sup> Several genes have been identified as PCa-specific biomarkers, including PCA3 and the *TMPRSS2:ERG* gene fusion. These have been detected in both tissue<sup>22,23</sup> and urine,<sup>24–26</sup> and are thus attractive from the standpoint of development of noninvasive PCa biomarker panels. In addition, DNA methylation<sup>27</sup> and single-nucleotide polymorphisms<sup>28</sup> have been explored as urinary PCa markers. Finally, metabolite biomarkers have been identified, with a recent comprehensive profiling study nominating several metabolites capable of distinguishing biopsy-negative and biopsy-positive urine sediment samples.<sup>29</sup> Among the metabolites highlighted, sarcosine has received particular attention and several validation studies have since been reported, with a clear consensus on biomarker status yet to be established.<sup>30–35</sup>

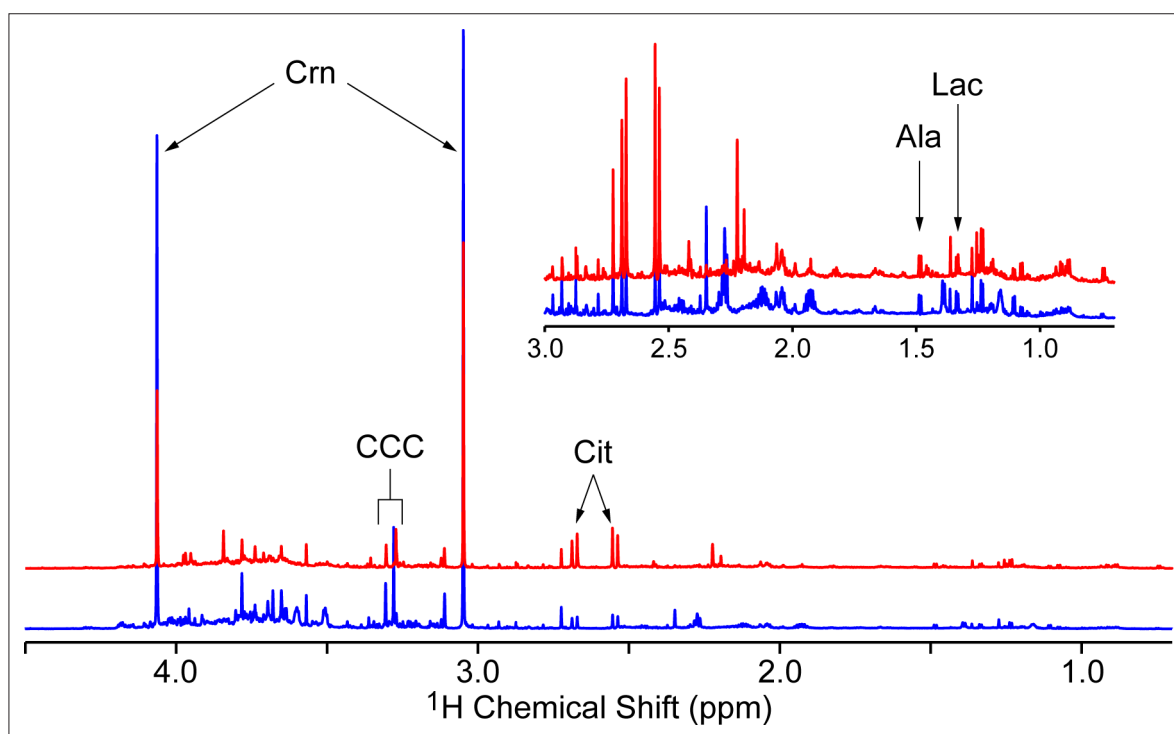
Metabolomic-based investigations are attractive in terms of sample processing, where, for example, extensive protein purification assays or gene amplification procedures are not necessary. Additionally, perturbed metabolite levels represent the end product of genetic dysfunction, and thus may be utilized to trace through the metabolic pathways to the dysfunctional enzymatic process. Analytical platforms commonly employed for metabolomic studies include mass spectrometry (MS) and nuclear magnetic resonance (NMR), each possessing complementary strengths and weaknesses.<sup>36</sup> PCa diagnosis continues to be a difficult clinical challenge, and new diagnostic techniques based on biomarker detection in noninvasive samples are under active research. Urine is perhaps the most attractive noninvasive biospecimen in terms of simplicity of collection, with biomarker detection possible in both urine supernatant and sediment. MS-based PCa metabolomic investigations of urine samples have been extensively reported; however, there is a noticeable absence of NMR-based reports. NMR of urine samples has been established in relation to various diseases,<sup>37–41</sup> thus this report intends to address biomarker panel detection for PCa based on the hypothesis that certain metabolites detected in urinary <sup>1</sup>H NMR profiles of biopsy-negative individuals will

differ from biopsy-positive individuals. Representative <sup>1</sup>H NMR spectra for biopsy-negative and biopsy-positive individuals are given in Figure 1.

Urine is a particularly challenging biospecimen to work with in terms of biomarker discovery due to the large compositional variability depending on diet, time of collection, gender, and medication taken by the patient.<sup>42–45</sup> To overcome this challenge, large sample cohorts are necessary and multivariate statistical modeling is essential in identifying reliable candidate biomarkers. Principal components analysis (PCA) is a common technique utilized in metabolomic studies; however, PCA failed to separate biopsy-negative from biopsy-positive profiles (data not shown) since the urine compositional variation was dominant over the variation attributable to PCa. Thus a partial least squares-discriminant analysis (PLS-DA) approach was applied to the data.

*Auto-scaled data.* As argued by Westerhuis and coworkers,<sup>46</sup> model parameters reporting on performance characteristics are often difficult to judge without an idea of how the model operates under the null hypothesis. Thus, they proposed performing permutation analyses as part of the modeling process to generate expected parameter distributions for the same data set assuming the null hypothesis is true. The expected number of misclassifications, area under the curve (AUC) values, and  $Q^2$  values after permutation of class labels are summarized in Figure 2 and Table 1. Taking 30% of the data as the test set resulted in 96 samples comprising the test set. The number of misclassifications and AUC values were determined to be  $48 \pm 2$  and  $0.49 \pm 0.04$ , respectively, indicating as expected that the modeling performed no better than random in class prediction. Importantly, the parameter distributions are vastly different when a double cross-validation (2CV) PLS-DA is applied utilizing the correct class labels. Improvement of all parameters was observed (decreased number of misclassifications, increased  $Q^2$ , and increased AUC,  $P < 0.0001$  in all cases), suggesting the models were capable of identifying signals important in distinguishing biopsy-negative versus biopsy-positive individuals.

Application of 2CV PLS-DA resulted in a collection of 1000 models, each built using a randomly selected subset containing 70% of the samples from the full data set. Thus, while each model is different, all models are related.<sup>46</sup> Identification of a single model would facilitate the examination of standard PLS-DA parameters such as latent variable loadings and scores in order to gain information on the variables contributing to the class separation. In the case where many similar models are generated, selection of a single “best” model is not trivial.<sup>46</sup> Therefore, variable selection methods were applied to each model generated, resulting in a collection of important signals over all models. Variable selection continues to be an important problem in multivariate analyses,<sup>47,48</sup> and without a clear consensus on a particular strategy, three methods were



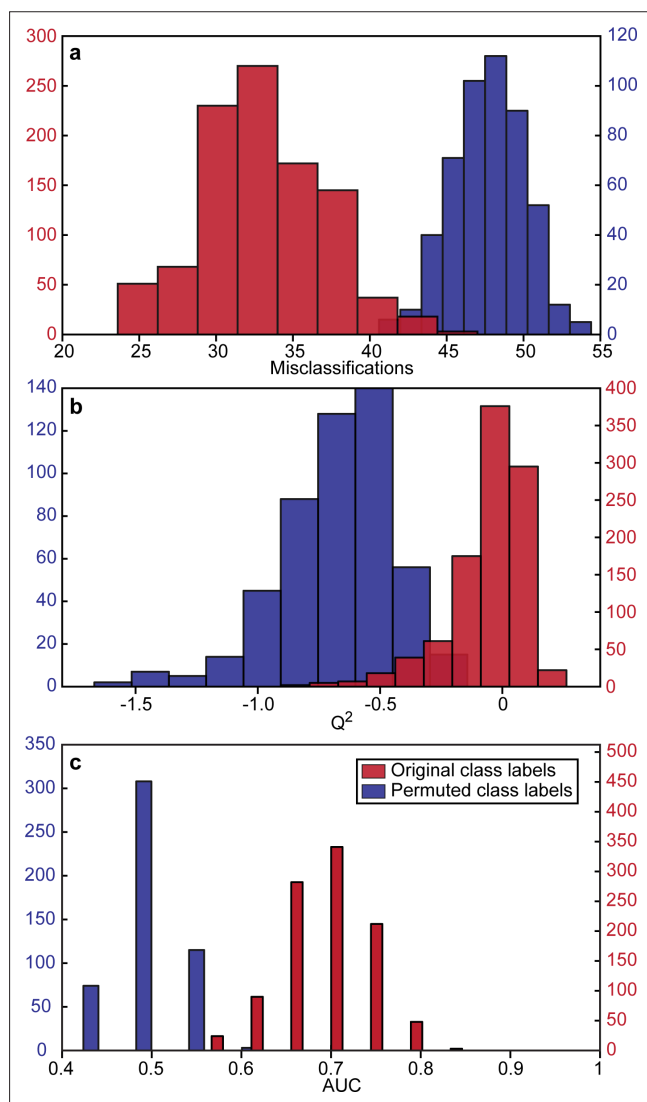
**Figure 1.** The 900 MHz  $^1\text{H}$  NMR spectra of a selected biopsy-negative (blue) and biopsy-positive (red) urine supernatant sample. Spectra were referenced to 3-(trimethylsilyl)propionic-2,2,3,3- $\text{d}_4$  acid sodium salt (TSP), probabilistic quotient normalized, and aligned. Inset: expansion of chemical shift region from 0.7 to 3.0 ppm, with vertical scale increased by a factor of 10. Selected assignments of common prostate cancer biomarkers reported in the literature are given (see text for references): Crn, creatinine; CCC, choline containing compounds region; Cit, citrate; Ala, alanine; Lac, lactate.

explored. The variable importance to projection (VIP) score is a value based on PLS loadings and thus reports on the importance of each variable in contributing to the overall model.<sup>49</sup> Since the average VIP score is 1, variables whose value is greater than 1 are then considered important. Alternatively, the absolute value of the regression coefficient calculated in the PLS algorithm may be used to rank each variable, which describes the relationship between the data loadings and the class label in DA.<sup>49</sup> Finally, a method recently proposed by Li and coworkers<sup>48</sup> termed CARS (competitive adaptive reweighted sampling) was explored as an alternative variable selection protocol. CARS aims to generate calibration models with variable subsets, with predictive power used to determine the optimal variable collection.<sup>48</sup> A variable subset is selected from the model with the best prediction performance.

Each variable selection protocol was applied to each model generated, and thus 1000 variable collections were obtained. All variables selected by CARS were examined and the frequency with which a variable was selected was calculated. In the case of VIP and regression coefficient-ranked variables, only the top 2% of variables from each model were collected (therefore, VIP score  $>4.26$ ), followed by calculation of the frequency of selection

(summarized in Table 2). Regardless of the variable selection protocol, variables that appeared in at least 50% of the models generated were considered important. This resulted in a collection of 21, 21, and 29 variables for the VIP, regression coefficient, and CARS variable selection protocols, respectively. It was observed that the 21 variables identified using VIP scores were identical to those selected using regression coefficients. Additionally, 12 variables were found to be common with the CARS variable list (all variable lists provided in the Supplemental Data). Selecting these variable subsets, logistic regression analysis was performed and receiver operating characteristic (ROC) curves generated, resulting in AUC values of 0.83 and 0.94 using all regression coefficient and CARS variables, respectively (Figure 3). Additionally, 2CV PLS-DA with permutation analysis was performed utilizing the regression coefficient and CARS variable subsets, with the results presented in Figure 4 and summarized in Table 1. Clearly, the predictive power of the models was enhanced when selecting variable subsets, with particular improvement using the CARS variable subset.

*Variable changes relative to biopsy-negative data.* The effect of centering and scaling all variables with respect to the mean and standard deviation of the normal samples (i.e., biopsy-negative) is a common approach to gain insight into



**Figure 2.** Distributions of the number of misclassifications,  $Q^2$ , and AUC after 2CV-PLS-DA modeling of the full NMR data set. Data were auto-scaled prior to modeling; 1000 models were generated using the correct class labels (red), and 500 models using permuted class labels (blue). In all cases, the distributions using the correct class labels were significantly different compared to the permuted models ( $P < 0.0001$  for each parameter). The data are summarized in Table 1. AUC, area under the curve; 2CV-PLS-DA, double cross-validation partial least squares-discriminant analysis; NMR, nuclear magnetic resonance.

variables whose levels are different specifically in comparison with normal samples. Holding all modeling procedures otherwise constant, the resulting performance parameters are reported in Table 3. As in the case of auto-scaling, performance after class permutation yielded results consistent with random sample class prediction. Utilizing the correct class labels, the predictive performance was observed to slightly increase when  $z$ -score changes with respect to biopsy-negative variables were used. Variable selection resulted in 34, 31, and 9 signals

appearing in at least 50% of the models, ranked according to VIP scores (resulting in scores  $>2.46$  collected), regression coefficients, and CARS, respectively (Table 2). In contrast to the auto-scaled data, all VIP-ranked variables were unique in comparison with the regression coefficient-ranked signals, while 5 of the 9 CARS-selected variables were also selected utilizing regression coefficient ranking.

Interestingly, only modest improvement in modeling parameters was observed when selecting variable subsets (Table 3). In particular, CARS failed to improve the modeling parameters as observed with the auto-scaled data, perhaps a result of the severe reduction in modeled variables. This is in contrast to the VIP score variable subset, where the increase in modeled variables degraded the classification parameters in comparison with the auto-scaled data. Close examination of the correct classification rates for each class (biopsy-negative and biopsy-positive) revealed a high rate for the biopsy-negative samples, but a poor rate for the biopsy-positive samples, contributing to an overall poor classification rate (Table 4). This is particularly true for the VIP variable subset which had the lowest modeling performance.

*Variable subset comparisons—univariate statistics and correlations.* The variable selection protocols identified signals important in multivariate 2CV-PLS-DA modeling. Once these signals are identified, it is interesting to determine their univariate capability in distinguishing PCa-positive versus PCa-negative groups (variable lists and  $P$ -values for each value are available in the Supplemental Data). Wilcoxon rank-sum tests were done for each variable along with the appropriate data scaling procedure. The number of variables deemed significant (i.e.,  $P < 0.05$ ) for each model are summarized in Figure 5. A potential explanation for the poor performance of the VIP-selected variable subset after  $z$ -score calculation relative to normal samples is evident. Despite having the largest variable subset, only 4 of the signals identified from the multivariate modeling were found to be significant using univariate statistics. In contrast, the variable subset with the best multivariate modeling parameters (auto-scaled data, CARS variable selection) was comprised of signals where greater than 50% were significant using univariate statistics.

With the development of variable subsets containing a relatively large number of signals, it is reasonable to question whether each signal is acting independently of the others. For example, since these variables are NMR signals it might be assumed that multiple resonances from a single molecule may be contributing to the model. To address this point, the two scaled data sets were subjected to a statistical total correlation spectroscopy analysis and the correlation between each signal within each subset of variables was examined.<sup>50</sup> In general there were a small number of correlations greater than 0.8, and even fewer correlations with  $r > 0.9$ , above which correlations suggest molecular structure connectivity<sup>51</sup> (Figure 6). However, the VIP data set subjected to normalization and scaling relative to biopsy-negative samples strongly deviates from this trend. A set of strong

**Table 1.** Summary of Model Performance Parameters for 2CV-PLS-DA Modeling Based on NMR Data That Was Auto-Scaled; 1000 Models Were Generated Using Original Labels and 500 Models Were Generated Using the Permuted Labels. In All Cases, the Test Data Set Contained 96 Randomly Selected Samples (30% of the Total Data Set). All Values Are Reported as Mean  $\pm$  Standard Deviation. In the Case of Permuted Data, Mean Parameter Values Were First Calculated Over the 20 Models Per Label Permutation Followed by Mean  $\pm$  Standard Deviation Calculation Over All 500 Permutations. The List of Variables Used in These Models Is Provided in the Supplementary Data.

Labels	Data set	Misclassification	$Q^2$	AUC
<b>Original Labels</b>	Full	33 $\pm$ 4	-0.06 $\pm$ 0.16	0.69 $\pm$ 0.05
	CARS	20 $\pm$ 4	0.35 $\pm$ 0.08	0.86 $\pm$ 0.03
	Reg. coeff.	28 $\pm$ 4	-0.01 $\pm$ 0.14	0.76 $\pm$ 0.04
	VIP	28 $\pm$ 4	0.01 $\pm$ 0.14	0.76 $\pm$ 0.04
<b>Permuted Labels</b>	Full	48 $\pm$ 2	-0.68 $\pm$ 0.24	0.49 $\pm$ 0.03
	CARS	48 $\pm$ 3	-1.6 $\pm$ 0.4	0.49 $\pm$ 0.04
	Reg. coeff.	47 $\pm$ 3	-2.8 $\pm$ 0.9	0.49 $\pm$ 0.04
	VIP	47 $\pm$ 3	-2.9 $\pm$ 0.8	0.49 $\pm$ 0.01

2CV-PLS-DA, double cross-validation partial least squares-discriminant analysis; NMR, nuclear magnetic resonance; CARS, competitive adaptive reweighted sampling; VIP, variable importance to projection; Reg. coeff., regression coefficient; AUC, area under the curve.

correlations was observed, with a high percentage possessing  $r > 0.9$ . The correlated variables primarily involved signals in the range of 3.2 to 3.9 ppm and most likely originate from a sugar molecule (e.g., glucose). This model contains a large number of highly correlated signals which may further explain the poor predictive performance since a large amount of redundant information is included. Interestingly the CARS model with auto-scaled data (i.e., the best multivariate model) had the lowest ratio of correlated signals to total number of variables modeled. This is an indication that the CARS method selected the greatest number of independent variables and thus had the maximum information content for reliable modeling.

*An optimal biomarker panel for PCa?* In total, 6 sets of signals were nominated as potential biomarker panels. From these variables subsets, 82 unique variables out of 145 were identified and 18 were selected in at least 50% of the 6 data sets. Boxplots of a selection of these variables are given in Figure 7, where biopsy-positive responses are varied relative

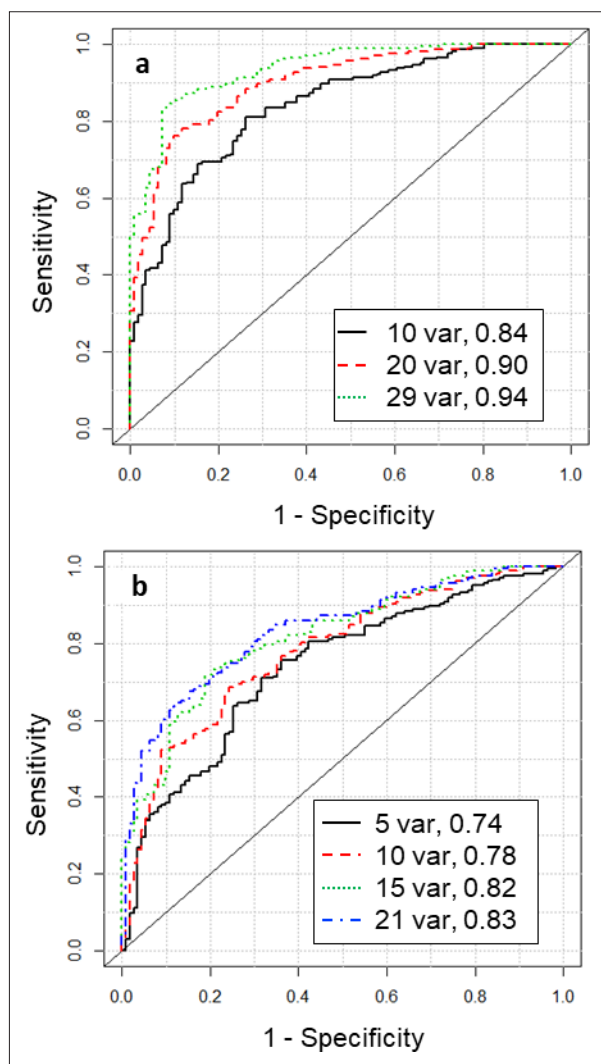
to biopsy-negative. It should be noted that of the 18 common variables, none were found uniquely in the data centered and scaled relative to biopsy-negative samples. This is somewhat unexpected since such data processing would be thought to amplify the differences between the negative and positive cases. This suggests that auto-scaling may be a sufficient data processing protocol. Nevertheless, the top 18 variables were taken as a final variable subset, a so-called “master list” of differential signals (Table S3, Supplemental Data). The modeling procedure was repeated with this master list and the performance characteristics are summarized in Table 5. In the case of the auto-scaled data, the results are intermediate of those obtained for the regression coefficient and CARS model data sets. This is not entirely unexpected since a large proportion of the “master” variables are common to both of these lists. Also encouraging are the results for the biopsy-negative centered and scaled data, which are essentially equivalent to the auto-scaled results. Thus through the analysis of two data processing techniques and three variable selection procedures, a final list of potential biomarkers has been identified which provides stable modeling results independent of the data processing technique.

NMR detection of metabolite markers of PCa has been primarily performed on prostate tissue (ex vivo or in vivo), prostatic secretions, or seminal fluid. From these studies (and others), citrate, choline containing compounds, spermine, myo-inositol, lactate, alanine, omega-6 fatty acids, creatine, and various signal ratios of these metabolites have been reported as potential markers of PCa.<sup>53–68</sup> Interestingly, while each of these metabolites possesses a clearly resolved NMR signal, not a single one was nominated in our analysis of urine supernatant. Urine-based MS has identified several metabolites correlated with PCa; however, these studies were performed on urine sediment. The sediment from urine

**Table 2.** The Number of Variables Selected for Each Data Treatment Procedure and Variable Selection Protocol. The Value Indicates the Number of Variables That Appeared in at Least 50% of the Models Generated, and Values in Parentheses Are the Total Number of Unique Variables Identified Over the 1000 Models Generated.

Data set	VIP	Regression coefficient	CARS
<b>Auto-scaled</b>	21 (723)	21 (692)	29 (1891)
<b>Scaled wrt biopsy-negative</b>	34 (169)	31 (342)	9 (1773)

VIP, variable importance to projection; CARS, competitive adaptive reweighted sampling.



**Figure 3.** Logistic regression analysis performed on the auto-scaled NMR data set taking the (a) CARS variables and (b) regression coefficient-ranked variable subsets identified from 2CV-PLS-DA analysis. ROC curves were generated using the first 10, 20, and 29 variables of the CARS subset, and the first 5, 10, 15, and 21 variables of the regression coefficient-ranked subset. The corresponding AUC values for each ROC curve are given in the legend. NMR, nuclear magnetic resonance; CARS, competitive adaptive reweighted sampling; 2CV-PLS-DA, double cross-validation partial least squares-discriminant analysis; ROC, receiver operating characteristic; AUC, area under the curve.

collected post-digital rectal exam (DRE) is expected to contain prostate cells which may be the source for these signals. Thus, urine supernatants, which may carry a greater fraction of hydrophilic water-soluble metabolites, appear to be exhibiting a different metabolomic profile compared to tissue and urine sediment profiles and may be a useful orthogonal route to PCa detection.

Such a biomarker panel is important in the context of PCa since there remains a population of PCa cases that are clinically challenging to diagnose. In an attempt to address

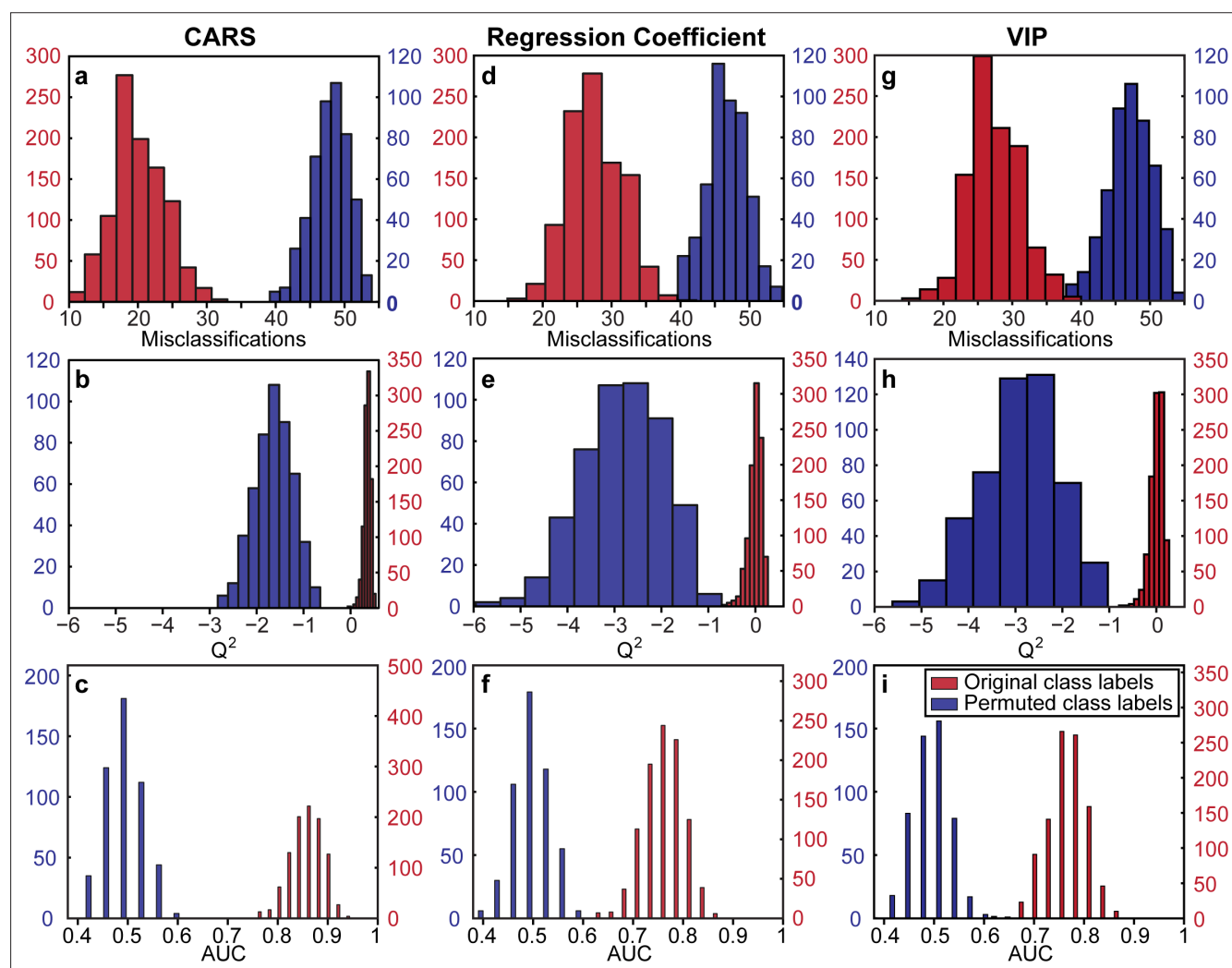
this issue, it is worth re-emphasizing that “normal cases” in this study referred to patients experiencing non-cancerous prostate disorders. It is common for patients possessing such disorders to display elevated levels of PSA, the current clinical standard for PCa diagnosis. This therefore complicates the diagnosis of PCa, and further invasive procedures are required. This study demonstrated the capability of NMR data to distinguish PCa patients from the biopsy-negative population. This suggests that a metabolomic analysis of urine (supernatant and sediment separately), a noninvasive biospecimen, may be further explored for PCa diagnostic potential. In combination with genetic and protein urinary biomarker diagnostics under active investigation, a completely noninvasive clinical test may be on the horizon capable of outperforming PSA levels for PCa identification.

It is encouraging that taken as potential biomarker panels the variable subsets perform exceptionally well in differentiating biopsy-negative versus biopsy-positive profiles. While it is certainly the case that the variable selection protocols utilized here are not completely independent (particularly VIP and regression coefficient ranking), a certain level of confidence is gained by more than one method selecting common variables. It is envisioned that a single variable selection protocol will not dominate, and instead a combination of techniques will generate the most reliable variable nominations. The next challenge becomes signal identification so that the biological significance of these variables may be investigated. Currently, the signal assignments remain unknown (attempted using Chenomx and MetaboID<sup>69</sup>), pointing to a primary challenge in NMR-based urinary metabolomics. Due to the complexity of the samples, assignment is a significant challenge which is slightly alleviated by utilizing high magnetic field strengths (900 MHz in this case). Unfortunately, the signals identified appear at low concentration (thus poor SNR) resulting in increased difficulty in obtaining suitable multidimensional NMR data, which is essential for confident assignment of the otherwise unknown resonances. Speculation of the most likely chemical functionality is possible. For example, the signal centered at 1.164 ppm most likely arises from small organic acid molecules containing CH<sub>2</sub> moieties. Nevertheless, this study demonstrates the utility of NMR-based metabolomics of urine for the nomination of a PCa metabolite biomarker panel.

## Experimental

### Materials

3-(Trimethylsilyl)propionic-2,2,3,3-d<sub>4</sub> acid sodium salt (TSP), D<sub>2</sub>O, and potassium phosphate (monobasic, dibasic) were obtained from Sigma/Aldrich (Milwaukee, USA) and used as received. All water was of MilliQ purity.



**Figure 4.** Distributions of the number of misclassifications,  $Q^2$ , and AUC after 2CV-PLS-DA modeling using variable subsets. The NMR data was auto-scaled prior to modeling. The subsets are the CARS variable subset (a-c), regression coefficient-ranked subset (d-f), and the VIP-ranked subset (g-i). In all cases, 1000 models were generated using the correct class labels (red) and 500 models were generated using permuted class labels (blue). The distributions using the correct class labels were all significantly different in comparison with the permuted distributions ( $P < 0.0001$  for each parameter and each subset). The data are summarized in Table 3, and the number of variables included in each model is given in Table 2. The full variable lists are available in the supplemental data. AUC, area under the curve; 2CV-PLS-DA, double cross-validation partial least squares-discriminant analysis; NMR, nuclear magnetic resonance; CARS, competitive adaptive reweighted sampling; VIP, variable importance to projection.

### Urine Specimens and Sample Treatment

Urine samples were collected according to a protocol approved by the Institutional Review Board. Briefly, after administering a DRE, urine samples were collected in preservative-free urine collection cups (minimum 10 mL) and centrifuged at 4000 rpm for 15 minutes at 4°C. The resulting supernatants were carefully separated from the pellets and stored separately at -80°C until used for analysis. NMR sample preparation involved aliquoting 450  $\mu$ L of neat urine supernatant with 50  $\mu$ L of phosphate buffer (pH 7.2) containing TSP as a chemical shift and concentration reference ( $[\text{phosphate}]_f = 100 \text{ mM}$ ,  $[\text{TSP}]_f = 3.7 \text{ mM}$ ). The sample was subjected to centrifugation for 1 minute to remove insoluble

material and transferred to a 5 mm NMR tube. All samples were prepared on the day of measurement.

### NMR Spectroscopy

All NMR measurements were performed on a Bruker AVANCE™ 900 MHz (Bruker Biospin, Rheinstetten, Germany) at the Michigan State University (East Lansing) with a TCI cryoprobe operating at a  $^1\text{H}$  frequency of 889.79 MHz at 298 K. Each 1D  $^1\text{H}$  spectrum was collected using a 30° flip angle, 128 scans containing 32 K data points, with a spectral width of 16 ppm and recycle delay of 2 seconds. Water suppression by gradient-tailored excitation was utilized.

**Table 3.** Summary of Model Performance Parameters for 2CV-PLS-DA Modeling Based on NMR Data That Was Centered and Scaled Relative to the Biopsy-Negative Samples; 1000 Models Were Generated Using Original Labels and 500 Models Were Generated Using the Permuted Labels. In All Cases, the Test Data Set Contained 96 Randomly Selected Samples (30% of the Total Data Set). All Values Are Reported as Mean  $\pm$  Standard Deviation. In the Case of Permuted Data, Mean Parameter Values Were First Calculated Over the 20 Models Per Label Permutation Followed by Mean  $\pm$  Standard Deviation Calculation Over All 500 Permutations. The List of Variables Used in These Models Is Provided in the Supplemental Data.

Labels	Data set	Misclassification	$Q^2$	AUC
<b>Original Labels</b>	Full	32 $\pm$ 4	0.01 $\pm$ 0.11	0.74 $\pm$ 0.05
	CARS	27 $\pm$ 4	-0.03 $\pm$ 0.13	0.74 $\pm$ 0.04
	Reg. coeff.	27 $\pm$ 4	0.06 $\pm$ 0.11	0.80 $\pm$ 0.04
	VIP	41 $\pm$ 4	-0.4 $\pm$ 0.04	0.71 $\pm$ 0.04
<b>Permuted Labels</b>	Full	47 $\pm$ 5	-4 $\pm$ 6	0.49 $\pm$ 0.04
	CARS	48 $\pm$ 3	-5 $\pm$ 3	0.49 $\pm$ 0.03
	Reg. coeff.	49 $\pm$ 3	-3 $\pm$ 1	0.49 $\pm$ 0.03
	VIP	46 $\pm$ 6	-6 $\pm$ 3	0.49 $\pm$ 0.03

2CV-PLS-DA, double cross-validation partial least squares-discriminant analysis; NMR, nuclear magnetic resonance; CARS, competitive adaptive reweighted sampling; VIP, variable importance to projection; Reg. coeff., regression coefficient.

### Data Preparation

The raw Free induction decay were imported in ACD/Labs NMR Processor (ACD/Labs, Toronto, Canada) and were subjected to zero filling to 64 K points and exponential multiplication equivalent to 0.3 Hz line broadening prior to Fourier transformation. Each spectrum was manually phase- and baseline-corrected and referenced to the TSP signal at 0 ppm prior to exporting the data as an ASCII file. All further processing, including statistical analyses, was performed in Matlab R2010b (V 7.11.0.584, The Mathworks, Natick, MA, USA).

Spectra were normalized using the probabilistic quotient normalization method, which has been shown to effectively reduce dilution variation.<sup>52</sup> The reference spectrum for normalization was selected according to a closeness index,<sup>70</sup> a scaled correlation coefficient which

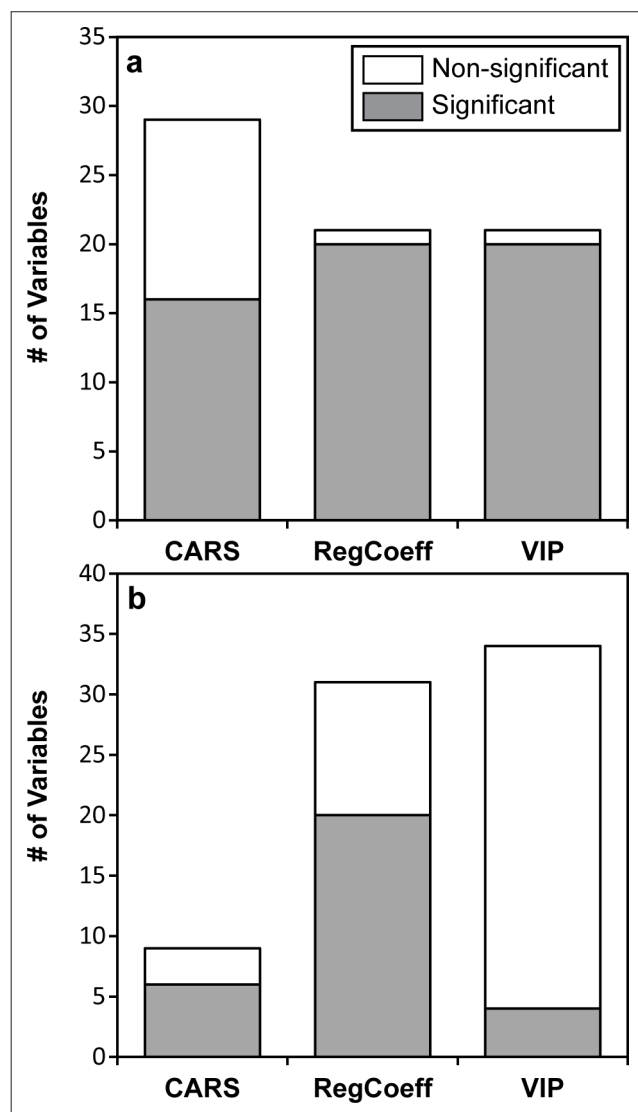
identifies the spectrum most similar to all other spectra. In order to reduce data variation arising from small chemical shift variation, spectral alignment was performed using the variable reference alignment method.<sup>71</sup> Prior to statistical modeling, the data set was binned (bin size = 0.005 ppm) in order to reduce the number of variables to a computationally manageable number, and scaled. Two scaling procedures were tested: first, each variable had the mean subtracted and scaled to unit variance ( $z$ -scores, auto-scaling<sup>72</sup>). Second, each variable had the mean of the biopsy-negative samples subtracted followed by division by the standard deviation of the biopsy-negative samples. In the latter scaling method, changes in variable  $z$ -score were then relative to the biopsy-negative cases.

**Table 4.** Correct Classification Rates Calculated for Each Model. After Each Model Was Determined, It Was Used to Predict the Classes of the Test Data Set. This Was Done Over the 1000 Models Generated, and the Number of Times Each Sample Was Part of the Test Set and Classified Correctly Was Determined. Classification Rates Were Then Calculated for Biopsy-Negative and Biopsy-Positive Samples as Individual Groups, and Over All Samples.

Data treatment	Data set	Biopsy-negative (%)	Biopsy-positive (%)	Overall (%)
<b>Auto-scaled</b>	Full	69 $\pm$ 34	63 $\pm$ 39	65 $\pm$ 37
	CARS	86 $\pm$ 27	75 $\pm$ 37	79 $\pm$ 34
	Reg. coeff.	69 $\pm$ 36	72 $\pm$ 40	71 $\pm$ 38
	VIP	70 $\pm$ 37	72 $\pm$ 40	71 $\pm$ 39
<b>Biopsy-negative centered and scaled</b>	Full	80 $\pm$ 32	59 $\pm$ 41	66 $\pm$ 39
	CARS	74 $\pm$ 41	71 $\pm$ 41	72 $\pm$ 41
	Reg. coeff.	88 $\pm$ 28	64 $\pm$ 45	72 $\pm$ 41
	VIP	87 $\pm$ 27	41 $\pm$ 43	57 $\pm$ 44

CARS, competitive adaptive reweighted sampling; VIP, variable importance to projection; Reg. coeff., regression coefficient.



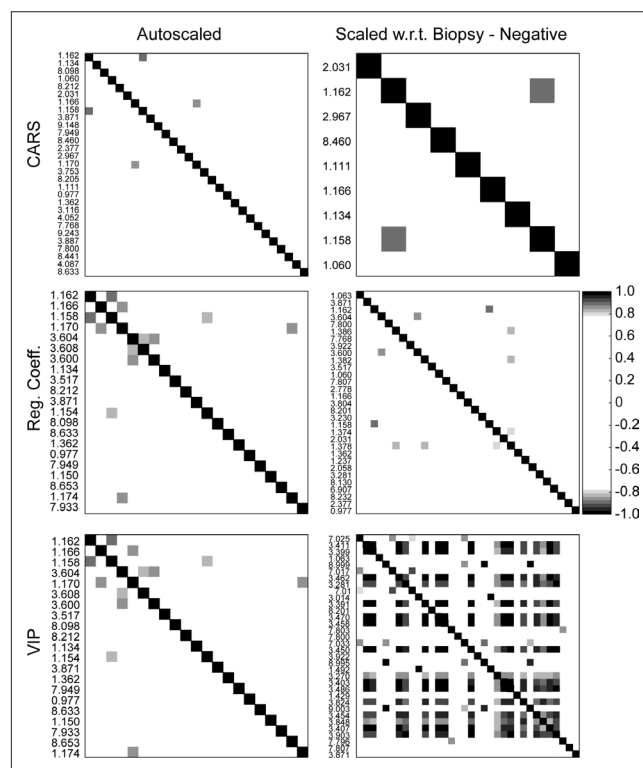


**Figure 5.** Bar plots representing the number of significant and non-significant variables identified in each variable subset using univariate statistics (Wilcoxon rank-sum test). (a) NMR data was auto-scaled; (b) NMR data was centered and scaled relative to the biopsy-negative samples. All *P*-values were corrected for multiple comparisons using the Bonferroni correction. Significance was tested at  $\alpha = 0.05$ . The shaded bar indicates the number of significant variables and the white bar the number of non-significant variables for each subset. NMR, nuclear magnetic resonance; CARS, competitive adaptive reweighted sampling; VIP, variable importance to projection; Reg. coeff., regression coefficient.

## Statistical Analyses

### Model development

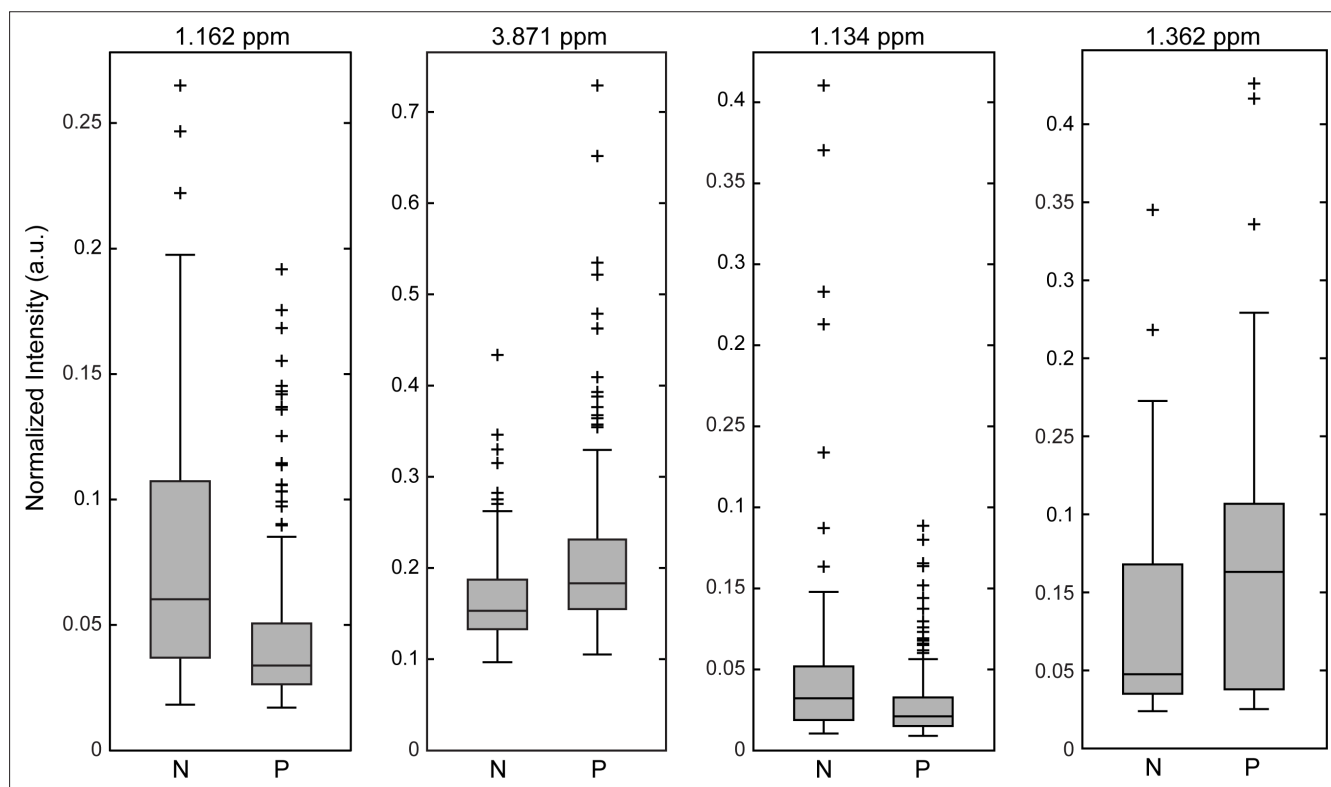
A 2CV-PLS-DA method<sup>46</sup> was used in order to develop models capable of classifying biopsy-negative versus biopsy-positive individuals. Briefly, the full data set was first randomly divided into a calibration and test set. The



**Figure 6.** Correlations between variables within each subset for each data normalization/scaling procedure. A Pearson correlation coefficient greater than 0.80 was chosen as a threshold. In the case of the auto-scaled data, the number of correlations with  $r > 0.80$  was 2, 6, and 6 for the CARS, regression coefficient-ranked, and VIP-ranked variable subsets, respectively. In the case of normalization and scaling relative to the biopsy-negative data, the number of correlations with  $r > 0.80$  were 1, 5, and 131 for the CARS, regression coefficient-ranked, and VIP-ranked variable subsets, respectively. CARS, competitive adaptive reweighted sampling; VIP, variable importance to projection.

calibration set is utilized to develop a PLS-DA model according to the traditional single cross-validation (1CV), whereby the data is randomly divided into training and validation sets. The training sets are used to generate PLS-DA models which are used to predict the classifications of the validation set. All prediction errors are stored and the PLS-DA model with optimal prediction performance is selected. After 1CV is performed on the calibration data, an optimal number of PLS components is identified and a final model is generated using the full calibration data set. Finally, this model is used to predict the classifications of the test set and all prediction errors are stored. The entire procedure is repeated several times, generating a series of different, but related, models.

Model performance was measured by three parameters: the number of misclassifications, a prediction error measure ( $Q^2$ ), and the AUC of an ROC curve. Reference distributions for the null hypothesis of the three parameters were generated by performing 2CV on the full data set after class label permutation. Significant values are expected only when using the correct



**Figure 7.** Representative signal boxplots of selected variables identified in the “master list.” The intensity was integrated at the signal indicated for the biopsy-negative (N) and biopsy-positive (P) spectra after all NMR data were normalized according to the PQN procedure.<sup>52</sup> The Wilcoxon *P*-values were  $P < 0.0001$ ,  $P < 0.0001$ ,  $P = 0.0002$ , and  $P = 0.01$  for the signals at 1.162, 3.871, 1.134, and 1.362 ppm, respectively. All *P*-values were corrected for multiple comparisons using the Bonferroni method. PQN, probabilistic quotient normalization; NMR, nuclear magnetic resonance.

class labels, which are directly comparable to the null hypothesis distributions using the permutation strategy.

In this work, null hypothesis distributions for the three performance parameters were generated by performing 500 class label permutations, with 20 models developed for each permutation (i.e., 2CV was performed 20 times for each permutation). For each permutation, 70% of the samples were randomly selected as the calibration set and the average value

of the number of misclassifications,  $Q^2$ , and AUC was calculated over the 20 models. The 2CV procedure was then repeated 1000 times using the correct class labels (each model taking a different random selection of the calibration set), resulting in a collection of 1000 different, but related models. Both the distribution over all 1000 models and mean  $\pm$  standard deviation of the three performance parameters are reported.

**Table 5.** Summary of Model Performance Parameters for 2CV-PLS-DA Modeling Based on the “Master List” Variable Subset; 1000 Models Were Generated Using Original Labels and 500 Models Were Generated Using the Permuted Labels. In All Cases, the Test Data Set Contained 96 Randomly Selected Samples (30% of the Total Data Set). All Values Are Reported as Mean  $\pm$  Standard Deviation. In the Case of Permuted Data, Mean Parameter Values Were First Calculated Over the 20 Models Per Label Permutation Followed by Mean  $\pm$  Standard Deviation Calculation Over All 500 Permutations. The List of Variables Used in These Models Is Provided in the Supplemental Data.

Labels	Data set	Misclassification	$Q^2$	AUC
<b>Original</b>	Auto-scaled	26 $\pm$ 4	0.15 $\pm$ 0.11	0.80 $\pm$ 0.04
	BNCS	27 $\pm$ 4	0.10 $\pm$ 0.11	0.79 $\pm$ 0.04
<b>Permuted</b>	Auto-scaled	48 $\pm$ 3	-2.8 $\pm$ 0.9	0.49 $\pm$ 0.04
	BNCS	48 $\pm$ 3	-5 $\pm$ 2	0.49 $\pm$ 0.03

BNCS, biopsy-negative centered and scaled; AUC, area under the curve.

## Variable selection

Three methods were explored for their ability to select variables strongly contributing to the PLS-DA models developed. The first method involved ranking variables according to their absolute regression coefficient value calculated for each PLS-DA model. The frequency with which a variable appeared in the top 2% ranking was calculated across the models. Variables that appeared in the top 2% in at least 50% of the models were considered important. The second method involved the calculation of the VIP score of each variable for all models calculated.<sup>49</sup> The variables were ranked according to their VIP score, and a frequency was calculated in the same fashion as for the regression coefficient analysis. The final method involved a more rigorous calculation of variable importance, known as CARS, described by Li and coworkers.<sup>48</sup> Briefly, in this method a series of PLS-DA models, built upon a data subset selected by Monte Carlo sampling of the calibration data, are generated and variables are exponentially filtered and subjected to competitive weighting according to their ability to affect model performance. A subset of variables is finally realized after selecting the model with best performance characteristics (lowest prediction error). The CARS method was applied to each model developed using 100 Monte Carlo samplings, and variables that were selected in at least 50% of the models were considered important.

*Logistic regression* analysis was performed using the selected variables with ROC curves generated and AUC values calculated.

*Univariate statistics* were performed on the selected variables as an alternative measure of variable importance. The Wilcoxon *U*-test was used to compare the normalized intensities of the selected variables, and all *P*-values were subjected to the Bonferroni correction for multiple comparisons.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: N.M., W.G., A.R., and the NMR studies were supported by funds from the University of Michigan (to A.R.).

## Supplemental Material

Supplemental material for this article is available online.

## References

1. Website. <https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html>. Accessed October 14, 2018.
2. Killick E, Bancroft E, Kote-Jarai Z, Eeles R. Beyond prostate-specific antigen — future biomarkers for the early detection and management of prostate cancer. *Clin Oncol*. 2012;24(8):545-555.
3. Djulbegovic M, Beyth RJ, Neuberger MM, et al. Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2010;341(sep14 1):c4543.
4. Lee YJ, Park JE, Jeon BR, Lee SM, Kim SY, Lee YK. Is prostate-specific antigen effective for population screening of prostate cancer? A systematic review. *Ann Lab Med*. 2013;33(4):233-241.
5. Pezaro C, Woo HH, Davis ID. Prostate cancer: measuring PSA. *Intern Med J*. 2014;44(5):433-440.
6. Barry MJ. Clinical practice. Prostate-specific-antigen testing for early diagnosis of prostate cancer. *N Engl J Med*. 2001;344(18):1373-1377.
7. Chou R, Crosswell JM, Dana T, et al. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2011;155(11):762-771.
8. Moyer VA, ; U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;157(2):120-134.
9. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst*. 2009;101(6):374-383.
10. Armstrong AJ, Eisenberger MA, Halabi S, et al. Biomarkers in the management and treatment of men with metastatic castration-resistant prostate cancer. *Eur Urol*. 2012;61(3):549-559.
11. Roobol MJ, Haese A, Bjartell A. Tumour markers in prostate cancer III: biomarkers in urine. *Acta Oncol*. 2011;50 Suppl 1(sup1):85-89.
12. Prensner JR, Rubin MA, Wei JT, Chinnaiyan AM. Beyond PSA: the next generation of prostate Cancer biomarkers. *Sci Transl Med*. 2012;4(127):127rv3.
13. Roberts MJ, Schirra HJ, Lavin MF, Gardiner RA. Metabolomics: a novel approach to early and noninvasive prostate cancer detection. *Korean J Urol*. 2011;52(2):79-89.
14. Gamagedara S, Kaczmarek AT, Jiang Y, et al. Validation study of urinary metabolites as potential biomarkers for prostate cancer detection. *Bioanalysis*. 2012;4(10):1175-1183.
15. Kumar D, Gupta A, Nath K. NMR-based metabolomics of prostate cancer: A protagonist in clinical diagnostics. *Expert Rev Mol Diagn*. 2016;16(6):651-661.
16. Eidelman E, Twum-Ampofo J, Ansari J, Siddiqui MM. The metabolic phenotype of prostate cancer. *Front Oncol*. 2017;7:131.
17. Sasaki M, Ishidoya S, Ito A, et al. Low percentage of free prostate-specific antigen (PSA) is a strong predictor of later detection of prostate cancer among Japanese men with serum levels of total PSA of 4.0 ng/mL or less. *Urology*. 2014;84(5):1163-1167.
18. Sreekumar A, Laxman B, Rhodes DR, et al. Humoral immune response to alpha-methylacyl-CoA racemase and prostate cancer. *J Natl Cancer Inst*. 2004;96(11):834-843.
19. Maraldo D, Garcia FU, Mutharasan R. Method for quantification of a prostate cancer biomarker in urine without sample preparation. *Anal Chem*. 2007;79(20):7683-7690.

20. Wayner EA, Quek S-I, Ahmad R, et al. Development of an ELISA to detect the secreted prostate cancer biomarker AGR2 in voided urine. *Prostate*. 2012;72(9):1023-1034.
21. Theodorescu D, Fliser D, Wittke S, et al. Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis*. 2005;26(14):2797-2808.
22. Ramos CG, Valdevenito R, Vergara I, Anabalón P, Sánchez C, Fulla J. PCA3 sensitivity and specificity for prostate cancer detection in patients with abnormal PSA and/or suspicious digital rectal examination. First Latin American experience. *Urologic Oncology: Seminars and Original Investigations*. 2013;31(8):1522-1526.
23. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and Ets transcription factor genes in prostate cancer. *Science*. 2005;310(5748):644-648.
24. Laxman B, Tomlins SA, Mehra R, et al. Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia*. 2006;8(10):885-888.
25. de Kok JB, Verhaegh GW, Roelofs RW, et al. DD3<sup>(PCA3)</sup>, a very sensitive and specific marker to detect prostate tumors. *Cancer Res*. 2002;62(9):2695-2698.
26. Tomlins SA, Aubin SMJ, Siddiqui J, et al. Urine TMPRSS2:ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. *Sci Transl Med*. 2011;3(94):94ra72.
27. Payne SR, Serth J, Schostak M, et al. DNA methylation biomarkers of prostate cancer: confirmation of candidates and evidence urine is the most sensitive body fluid for non-invasive detection. *Prostate*. 2009;69(12):1257-1269.
28. Whitaker HC, Kote-Jarai Z, Ross-Adams H, et al. The rs10993994 risk allele for prostate cancer results in clinically relevant changes in microseminal protein-beta expression in tissue and urine. *PLoS One*. 2010;5(10):e13363.
29. Sreekumar A, Poisson LM, Rajendiran TM, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*. 2009;457(7231):910-914.
30. Lucarelli G, Fanelli M, Larocca AMV, et al. Serum sarcosine increases the accuracy of prostate cancer detection in patients with total serum PSA less than 4.0 ng/ml. *Prostate*. 2012;72(15):1611-1621.
31. Wu H, Liu T, Ma C, et al. GC/MS-based metabolomic approach to validate the role of urinary sarcosine and target biomarkers for human prostate cancer by microwave-assisted derivatization. *Anal Bioanal Chem*. 2011;401(2):635-646.
32. Jiang Y, Cheng X, Wang C, Ma Y. Quantitative determination of sarcosine and related compounds in urinary samples by liquid chromatography with tandem mass spectrometry. *Anal Chem*. 2010;82(21):9022-9027.
33. Jentzmik F, Stephan C, Müller K, et al. Sarcosine in urine after digital rectal examination fails as a marker in prostate cancer detection and identification of aggressive tumours. *Eur Urol*. 2010;58(1):12-18.
34. Cao D-L, Ye D-W, Zhu Y, Zhang H-L, Wang Y-X, Yao X-D. Efforts to resolve the contradictions in early diagnosis of prostate cancer: a comparison of different algorithms of sarcosine in urine. *Prostate Cancer Prostatic Dis*. 2011;14(2):166-172.
35. Lima AR, Bastos MdeL, Carvalho M, Guedes de Pinho P, de Bastos M, de Pinho P. Biomarker discovery in human prostate cancer: an update in metabolomics studies. *Transl Oncol*. 2016;9(4):357-370.
36. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem*. 2007;387(2):525-527.
37. Deja S, Barg E, Młynarz P, Basiak A, Willak-Janc E. 1H NMR-based metabolomics studies of urine reveal differences between type 1 diabetic patients with high and low HbA1c values. *J Pharm Biomed Anal*. 2013;83:43-48.
38. Nadal-Desbarats L, Aïdoud N, Emond P, et al. Combined 1H-NMR and 1H-13C HSQC-NMR to improve urinary screening in autism spectrum disorders. *Analyst*. 2014;139(13):3460-3468.
39. Austdal M, Tangerås LH, Skråstad RB, et al. First trimester urine and serum metabolomics for prediction of preeclampsia and gestational hypertension: a prospective screening study. *Int J Mol Sci*. 2015;16(9):21520-21538.
40. Dawiskiba T, Deja S, Mulak A, et al. Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. *World J Gastroenterol*. 2014;20(1):163-174.
41. Yang B, Liao G-Q, Wen X-F, et al. Nuclear magnetic resonance spectroscopy as a new approach for improvement of early diagnosis and risk stratification of prostate cancer. *J Zhejiang Univ Sci B*. 2017;18(11):921-933.
42. Edmands WMB, Beckonert OP, Stella C, et al. Identification of human urinary biomarkers of cruciferous vegetable consumption by metabolomic profiling. *J Proteome Res*. 2011;10(10):4513-4521.
43. Lenz EM, Bright J, Wilson ID, et al. Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *J Pharm Biomed Anal*. 2004;36(4):841-849.
44. Zhang S, Liu L, Steffen D, Ye T, Raftery D. Metabolic profiling of gender: Headspace-SPME/GC-MS and 1H NMR analysis of urine. *Metabolomics*. 2012;8(2):323-334.
45. Saude EJ, Adamko D, Rowe BH, Marrie T, Sykes BD. Variation of metabolites in normal human urine. *Metabolomics*. 2007;3(4):439-451.
46. Westerhuis JA, Hoefsloot HCJ, Smit S, et al. Assessment of PLS-DA cross validation. *Metabolomics*. 2008;4(1):81-89.
47. Brereton RG. Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. *TrAC Trends in Analytical Chemistry*. 2006;25(11):1103-1111.
48. Li H, Liang Y, Xu Q, Cao D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal Chim Acta*. 2009;648(1):77-84.
49. Chong I-G, Jun C-H. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*. 2005;78(1-2):103-112.

50. Cloarec O, Dumas M-E, Craig A, et al. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic  $^1\text{H}$  NMR data sets. *Anal Chem*. 2005;77(5):1282-1289.
51. Coen M, Hong Y-S, Cloarec O, et al. Heteronuclear  $^1\text{H}$ - $^{31}\text{P}$  statistical total correlation NMR spectroscopy of intact liver for metabolic biomarker assignment: application to galactosamine-induced hepatotoxicity. *Anal Chem*. 2007;79(23):8956-8966.
52. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in  $^1\text{H}$  NMR metabolomics. *Anal Chem*. 2006;78(13):4281-4290.
53. Kurhanewicz J, Vigneron DB, Nelson SJ, et al. Citrate as an in vivo marker to discriminate prostate cancer from benign prostatic hyperplasia and normal prostate peripheral zone: detection via localized proton spectroscopy. *Urology*. 1995;45(3):459-466.
54. Lynch MJ, Nicholson JK. Proton MRS of human prostatic fluid: correlations between citrate, spermine, and myo-inositol levels and changes with disease. *Prostate*. 1997;30(4):248-255.
55. Costello LC, Franklin RB, Narayan P. Citrate in the diagnosis of prostate cancer. *Prostate*. 1999;38(3):237-245.
56. van der Graaf M, Schipper RG, Oosterhof GON, Scholken JA, Verhofstad AAJ, Heerschap A. Proton MR spectroscopy of prostatic tissue focused on the detection of spermine, a possible biomarker of malignant behavior in prostate cancer. *Magnetic Resonance Materials in Physics, Biology and Medicine*. 2000;10(3):153-159.
57. Cheng LL, Wu C, Smith MR, Gonzalez RG. Non-destructive quantitation of spermine in human prostate tissue samples using HRMAS  $^1\text{H}$  NMR spectroscopy at 9.4 T. *FEBS Lett*. 2001;494(1-2):112-116.
58. Swanson MG, Vigneron DB, Tabatabai ZL, et al. Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted postsurgical prostate tissues. *Magn Reson Med*. 2003;50(5):944-954.
59. Averna TA, Kline EE, Smith AY, Sillerud LO. A decrease in  $^1\text{H}$  nuclear magnetic resonance spectroscopically determined citrate in human seminal fluid accompanies the development of prostate adenocarcinoma. *J Urol*. 2005;173(2):433-438.
60. Kline EE, Treat EG, Averna TA, Davis MS, Smith AY, Sillerud LO. Citrate concentrations in human seminal fluid and expressed prostatic fluid determined via  $^1\text{H}$  nuclear magnetic resonance spectroscopy outperform prostate specific antigen in prostate cancer detection. *J Urol*. 2006;176(5):2274-2279.
61. Swanson MG, Zektzer AS, Tabatabai ZL, et al. Quantitative analysis of prostate metabolites using  $^1\text{H}$  HR-MAS spectroscopy. *Magn Reson Med*. 2006;55(6):1257-1264.
62. Swanson MG, Keshari KR, Tabatabai ZL, et al. Quantification of choline- and ethanolamine-containing metabolites in human prostate tissues using  $^1\text{H}$  HR-MAS total correlation spectroscopy. *Magn Reson Med*. 2008;60(1):33-40.
63. Tessem M-B, Swanson MG, Keshari KR, et al. Evaluation of lactate and alanine as metabolic biomarkers of prostate cancer using  $^1\text{H}$  HR-MAS spectroscopy of biopsy tissues. *Magn Reson Med*. 2008;60(3):510-516.
64. van Asten JJA, Cuijpers V, Hulsbergen-van de Kaa C, et al. High resolution magic angle spinning NMR spectroscopy for metabolic assessment of cancer presence and Gleason score in human prostate needle biopsies. *Magnetic Resonance Materials in Physics, Biology and Medicine*. 2008;21(6):435-442.
65. Serkova NJ, Gamito EJ, Jones RH, et al. The metabolites citrate, myo-inositol, and spermine are potential age-independent markers of prostate cancer in human expressed prostatic secretions. *Prostate*. 2008;68(6):620-628.
66. Stenman K, Hauksson JB, Gröbner G, Stattin P, Bergh A, Riklund K. Detection of polyunsaturated omega-6 fatty acid in human malignant prostate tissue by 1D and 2D high-resolution magic angle spinning NMR spectroscopy. *MAGMA*. 2009;22(6):327-331.
67. DeFeo EM, Cheng LL. Characterizing human cancer metabolomics with *ex vivo*  $^1\text{H}$  HRMAS MRS. *Technol Cancer Res Treat*. 2010;9(4):381-391.
68. Vandergrift LA, Decelle EA, Kurth J, et al. Metabolomic prediction of human prostate cancer aggressiveness: magnetic resonance spectroscopy of histologically benign tissue. *Sci Rep*. 2018;8(1):4997.
69. MacKinnon N, Somashekar BS, Tripathi P, et al. MetaboID: a graphical user interface package for assignment of  $^1\text{H}$  NMR spectra of bodyfluids and tissues. *J Magn Reson*. 2013;226:93-99.
70. Veselkov KA, Lindon JC, Ebbels TMD, et al. Recursive segment-wise peak alignment of biological (1)h NMR spectra for improved metabolic biomarker recovery. *Anal Chem*. 2009;81(1):56-66.
71. MacKinnon N, Ge W, Khan AP, et al. Variable reference alignment: an improved peak alignment protocol for NMR spectral data with large intersample variation. *Anal Chem*. 2012;84(12):5372-5379.
72. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*. 2006;7(1):142.