NMR-Based Metabolic Profiling: Methods and Application in Cancer Biomarker Discovery

by

Wencheng Ge

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2014**

Doctoral Committee:

  Professor Ayyalusamy Ramamoorthy, Chair
  Professor Zhan Chen
  Professor Ari Gafni
  Professor Michael D. Morris

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA — analysis of variance

AUC — Area Under the Curve

BCa — bladder cancer

BMRB — Biological Magnetic Resonance Data Bank

CPMG — Carr-Purcell-Meiboom-Gill

CSA — chemical shift anisotropy

CV — cross validation

FID — free induction decay

HMDB — Human Metabolome Database

MAS — magic angle spinning

MS — mass spectrometry

NMR — nuclear magnetic resonance

OPLS-DA — orthogonal partial least squares discriminant analysis

PCa — prostate cancer

PCA — principal component analysis

PLS-DA — partial least squares discriminant analysis

PQN — probabilistic quotient normalization

ROC — receiver operating characteristics

TOCSY — total correlation spectroscopy

WATERGATE — WATER suppression by GrAdient-Tailored Excitation

CHAPTER 1

Introduction to metabolomics

Metabolomics is the characterization of low molecular weight compounds (molecular weight < 1500 Da) in biofluids[1, 2, 3], tissues[4, 5, 6, 7] and cells[8, 9, 10] to get insights of the perturbation to an organism caused by diseases, aging[11] or external stimuli[12, 13, 14, 15, 16] (e.g. drug treatment). It is the most relevant to phenotypes as compared with other "Omics" studies[17] (Figure. 1.1). Research fields of metabolomics include sample preparation and extraction for metabolic profiling, disease biomarker discovery, drug toxicity study, bioinformatics study (e.g. application of classification and prediction models) and data processing approaches related to specific analytical techniques such as Nuclear Magnetic Resonance spectroscopy (NMR).

The "Omics" Cascade

What can happen → GENOME

What appears to be happening → TRANSCRIPTOME

What makes it happen → PROTEOME

What has happened and is happening → METABOLOME
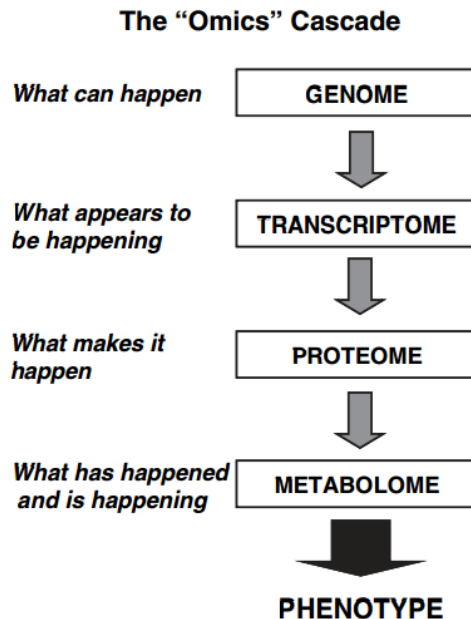
→ PHENOTYPE

Figure 1.1 The correlation of different "omics" studies. Metabolomics shows the closest relevance to the phenotype.[17]

1.1 Metabolic profiling by NMR

NMR is widely used in metabolic profiling to identify and quantify metabolites. This is because it allows non-destructive measurement of a variety of structurally different metabolites. Moreover, quantification is easily achieved based on the peak area of the specific metabolite related to the peak area of a known concentration of the internal standard. External reference compounds can also be used for the quantification of metabolites by inserting a co-axial capillary containing the solution of the reference compound into the NMR tube during measurement[18]. Moreover, NMR exhibits high reproducibility of measurement, and requires minimum sample preparation as compared with other techniques such as mass spectrometry (MS).

The drawback of NMR lies in the low sensitivity of detection. The sensitivity is related to the gyromagnetic ratio (γ) of the nuclear spins, experiment acquisition time and the metabolite concentrations that correspond to the number of nuclear spins being

2

observed etc. Metabolite concentrations ranging from µM to mM can be readily detected by 1D NMR in ~10 min. Increased signal/noise and spectral resolution can be achieved by the application of a higher magnetic field instrument and cryoprobe to reduce thermal noise[19]. Commonly, isotopic enrichment of the metabolites through chemical reactions can be applied to enhance the sensitivity of heteronuclear 2D NMR experiment[20, 21, 22].

1.1.1 NMR experiments for different sample types

Biofluids such as serum, plasma, urine, amniotic fluid, cerebrospinal fluid, saliva and other digestive fluids can be analyzed by liquid state NMR spectroscopy, usually with a 1D ^1H NMR experiment for high throughput metabolic profiling. Solvent suppression is required during the measurement due to the large water signal dominating the spectrum. Water presaturation by applying a weak radiofrequency (RF) during the relaxation/mixing time and WATERGATE (WATER suppression by GrAdient-Tailored Excitation) are the two most frequently used approaches for water suppression[23, 24]. Another method, excitation sculpting by water-selective suppression using pulsed field gradients is also reported to be convenient when detecting small molecules[25]. For samples containing minimum proteins and lipids (e.g. urine), phosphate buffer made in D_2O can be added to the sample and measured by a single pulse NMR experiment with water suppression. Whereas for biofluids containing macromolecules (e.g. blood serum), T_2 relaxation editing by Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence in NMR experiment is utilized to attenuate broad resonances from proteins and lipids which have short T_2 relaxation time[26]. Alternatively, sample preparation such as deproteinization by reaction with organic solvents or ultrafiltration can be used and the resulting metabolite extract measured by a single pulse NMR experiment.

Metabolic profiling of intact tissues and cells is important in obtaining biochemical information with minimum perturbation to the sample as well as retaining the sample for subsequent histopathological analysis. The nondestructive detection can be achieved using solid state magic-angle spinning (MAS) NMR with CPMG and water suppression pulse sequence to minimize the spectral line broadening[5, 27]. Standard processes for sample preparation are freezing the tissues in liquid nitrogen to quench the enzymatic reactions, followed by grinding the tissue in a cooled mortar or with an electric homogenizer. Then the tissue sample is packed into the rotor with the addition of D_2O for signal locking during NMR measurement. Inserting the sample rotor into the NMR probe at an angle of 54.7° relative to the external magnetic field B_0 and spinning the rotor will result in narrow linewidth due to the cancelation of the orientation dependent chemical shift anisotropy (CSA) and dipolar coupling effects in the solid samples. Moreover, experiments should be conducted at a low temperature (e.g. 4 °C) to avoid biological degradation of the sample.

1.1.2 Data preprocessing

The acquired free induction decay (FID) data can be processed using ACD/NMR Processor (ACD/Labs, Toronto, Canada) in which 2x zero-filling and an exponential window function equivalent to 0.3 Hz line broadening is usually applied before Fourier transformation. The spectra will be phased, baseline corrected, referenced to standard compound signal of a specific chemical shift, and exported as an ASCII file. The ASCII file is imported into Excel and the spectral region from -0.02 ppm to 10 ppm is chosen for subsequent statistical analysis. Other software such as topspin, NMRlab, ProMetab,

matNMR and MetaboLab can also be applied to process the raw FID data to visualize the spectra.

1.1.3 NMR spectral signal assignment

1D ^{1}H NMR spectra are routinely acquired due to the high NMR sensitivity of the hydrogen nucleus. However, the 1D spectra of samples containing complex mixture of metabolites can be congested due to the resonance appeared at the same chemical shift, making the resonance assignment a challenge. 2D homonuclear / heteronuclear correlation and J-resolved (JRES) experiments can be applied to alleviate the congestion of 1D spectrum. Spectral resonance assignment is usually performed by searching chemical shift lists, for example, in HMDB, BMRB, Chenomx software (Chenomx Inc., Edmonton, Candada), and published literatures along with the analysis of specific J-coupling patters of the metabolite spin systems. Moreover, this process is more efficient now with the help of MetaboID[28], a graphical user interface that offers comparison of experimental spectra with authentic metabolite spectra to guide assignment.   Usually, confident resonance assignment is achieved by combining the assignment of 1D ^{1}H spectra along with the knowledge obtained from the 2D experiments of selected samples.

1.2 Data processing and statistical analysis

1.2.1 Data pretreatment

Normalization is routinely used to account for the dilution effect of each sample as well as variation from different batch of measurement[29]. It is especially beneficial to urine sample data processing since urinary metabolite concentrations are dependent on the different water intake of individuals.   Normally, each metabolite peak integral is

normalized to the creatinine level based on the assumption that the creatinine level is an indicator of the metabolite concentrations in urine. However, creatinine normalization is questioned since factors such as muscle mass, age and gender can influence the normal urine creatinine level. Another commonly used integral normalization method is the total area normalization, which assumes the total integral of all metabolites in a sample to be a constant throughout all samples. However, this method is not applicable if drug metabolites also appeared in the sample. Thus far, probabilistic quotient normalization (PQN) is brought to attention[30]. It calculates a most probable dilution factor (e.g. the median of the quotients) among all the metabolite variables which is then utilized to perform normalization of all variables in the sample.

Scaling is a variable-based processing approach which is performed on the intensity of each metabolite across all samples, and influences the result of the multivariate analysis[31]. It eliminates the bias towards high concentration metabolites in statistical analysis and the scaling method chosen is dependent on the specific statistical methods used in the analysis. For instance, principal component analysis (PCA) which is performed on mean-centered data (e.g. metabolite variable mean is subtracted across all samples) is equivalent to performing PCA on a covariance matrix. Whereas applying PCA on standardized variable (e.g. based on correlation matrix of the variables) requires the data to be scaled with unit variance scaling, dividing the intensity of each variable by the standard deviation of the variable intensities across all samples. The advantages and disadvantages of various data pretreatment methods such as pareto scaling, log transformation and so forth were reported in Table.1.1.[32]

Table 1.1 Comparison of various data pretreatment methods as described by Van der burg et al[32].

Class	Method	Formula	Unit	Goal	Advantages	Disadvantages
I	Centering	$\tilde{x}_{ij} = x_{ij} - \bar{x}_i$	0	Focus on the differences and not the similarities in the data	Remove the offset from the data	When data is heteroscedastic, the effect of this pretreatment method is not always sufficient
II	Autoscaling	$\tilde{x}_{ij} = \dfrac{x_{ij} - \bar{x}_i}{s_i}$	(-)	Compare metabolites based on correlations	All metabolites become equally important	Inflation of the measurement errors
	Range scaling	$\tilde{x}_{ij} = \dfrac{x_{ij} - \bar{x}_i}{\left(x_{i_{max}} - x_{i_{min}}\right)}$	(-)	Compare metabolites relative to the biological response range	All metabolites become equally important. Scaling is related to biology	Inflation of the measurement errors and sensitive to outliers
	Pareto scaling	$\tilde{x}_{ij} = \dfrac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$	0	Reduce the relative importance of large values, but keep data structure partially intact	Stays closer to the original measurement than autoscaling	Sensitive to large fold changes
	Vast scaling	$\tilde{x}_{ij} = \dfrac{\left(x_{ij} - \bar{x}_i\right)}{s_i} \cdot \dfrac{\bar{x}_i}{s_i}$	(-)	Focus on the metabolites that show small fluctuations	Aims for robustness, can use prior group knowledge	Not suited for large induced variation without group structure
	Level scaling	$\tilde{x}_{ij} = \dfrac{x_{ij} - \bar{x}_i}{\bar{x}_i}$	(-)	Focus on relative response	Suited for identification of e.g. biomarkers	Inflation of the measurement errors
III	Log transformation	$\tilde{x}_{ij} = {}^{10}\log\left(x_{ij}\right)$ $\hat{x}_{ij} = \tilde{x}_{ij} - \bar{\tilde{x}}_i$	Log 0	Correct for heteroscedasticity, pseudo scaling. Make multiplicative models additive	Reduce heteroscedasticity, multiplicative effects become additive	Difficulties with values with large relative standard deviation and zeros
	Power transformation	$\tilde{x}_{ij} = \sqrt{\left(x_{ij}\right)}$ $\hat{x}_{ij} = \tilde{x}_{ij} - \bar{\tilde{x}}_i$	$\sqrt{0}$	Correct for heteroscedasticity, pseudo scaling	Reduce heteroscedasticity, no problems with small values	Choice for square root is arbitrary.

1.2.2 Univariate analysis

To compare each variable in the data containing two groups (e.g. benign and cancer), the student's t test which relies on the comparison of the two-sample means is commonly used. It requires each sample population to have a normal distribution, equal variance of the two populations and the data is independently sampled. Welch's t test can be applied to two sample populations with unequal variances. For data contains more than two groups, one-way analysis of variance (ANOVA) can be applied to test the

difference between group means of each variable. When the assumption of normal distribution is not met for the sample population, nonparametric analysis such as Wilcoxon rank sum test can be used to test the difference between two independent samples by comparing their medians. Kruskal-Wallis one-way analysis of variance is the nonparametric equivalent of ANOVA comparing data with multiple groups.

1.2.3 Multivariate analysis

To account for the impact of all metabolites on the outcome of measurement (e.g. cancer and normal), multivariate analysis can be utilized. Additionally, the multivariate model can be used to determine the metabolites contributing to the result as well as predicting the result.

1.2.3.1 Unsupervised analysis

Unsupervised analysis is the application of statistical models without the prior knowledge of the sample classification labels and it is usually the first step in data pattern exploration. Principal component analysis (PCA) is representative of the unsupervised method and is commonly applied to reduce the dimensionality and examine the structure of the data set. Scores plot is generated to assess the clustering of different samples, with the corresponding loadings plot demonstrating the variables accounting for the most variation in the specified principal component. In addition, the cluster analysis, which is useful for the visualization of subgroups of multivariate data by partitioning methods or hierarchical clustering, is also widely used in metabolomics data analysis.

1.2.3.2 Supervised analysis

In supervised analysis, information of sample class labels (e.g. disease and control) are also utilized in building the statistic models. One commonly used supervised analysis is partial least squares discriminant analysis (PLS-DA) which maximizes the covariance between predictor variables (metabolite intensities from NMR measurement) and the response variables (e.g. the classes of each sample)[33]. Namely, PLS-DA finds components (e.g. latent variables) in the predictor matrix that best predict the response variables. It uses variable importance to projection (VIP) scores to demonstrate the contribution of each variable to the model, with metabolites VIP scores> 1 considered important in classification. If class separation is not observed in scores plot of PLS-DA model, orthogonal partial least squares discriminant analysis (OPLS-DA) can be performed. The OPLS-DA is a modification of PLS-DA model, in which the systematic variation of data that is not related to the response variable (e.g. sample class labels) is removed[34]. It presents similar prediction ability to PLS-DA, and demonstrates improved model interpretability.[35] Diagnostic parameters such as the number of misclassifications, cross-validated explained variation Q^2 and the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) analysis are commonly used to indicate the model performance.

1.2.3.3 Model validation

To assess the predictive ability of the multivariate models, methods such as cross validation (CV), permutation or bootstrap can be conducted. Cross validation involves separating the data into a training set and a test set. The training set is used to build the

classification models (e.g. PLS-DA), and the resulting model is used to predict the classes of the test set. Depending on how the full data is partitioned into the two data sets, leave-one-out CV and k-fold CV are commonly applied. A permutation test can assess whether the classification based on true sample class is significantly better than classification based on randomly assigned sample class[36]. Morever, bootstrap is a resampling method used for model validation[37]. It generates a new data set the same size as the original by sampling with replacement from the original data set. This new data set is used to build the prediction model, and the validation is applied on the original data set.

1.3 Pathway analysis in disease biomarker discovery

Metabolic pathway analysis is essential for the understanding of cellular processes of specific diseases, providing insight in the development of treatment methods. After the identification of potential metabolite biomarkers, the particular pathway can be assessed using databases such as KEGG[38] and HMDB[39], or searching for literatures that contain pathway analysis of certain metabolite and disease. The enzymes controlling the metabolite biomarker levels in the cell play a critical role in the development of the disease, thus, studies focusing on the knockdown of certain enzyme in the cell can be applied to test its impact on the biomarker level and promote understanding of mechanisms associated with the specific disease.

1.4 References

1.   Barton, R. H.; Nicholson, J. K.; Elliott, P.; Holmes, E., High-throughput H-1 NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *International Journal of Epidemiology* **2008,** 37, 31-40.

2.   Fonville, J. M.; Maher, A. D.; Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Evaluation of Full-Resolution J-Resolved H-1 NMR Projections of Biofluids for Metabonomics Information Retrieval and Biomarker Identification. *Anal Chem* **2010,** 82, (5), 1811-1821.

3.   Schicho, R.; Nazyrova, A.; Shaykhutdinov, R.; Duggan, G.; Vogel, H. J.; Storr, M., Quantitative metabolomic profiling of serum and urine in DSS-induced ulcerative colitis of mice by (1)H NMR spectroscopy. *J Proteome Res* **2010,** 9, (12), 6265-73.

4.   Want, E. J.; Masson, P.; Michopoulos, F.; Wilson, I. D.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Loftus, N.; Holmes, E.; Nicholson, J. K., Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc* **2012,** 8, (1), 17-32.

5.   Beckonert, O.; Coen, M.; Keun, H. C.; Wang, Y.; Ebbels, T. M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues. *Nat Protoc* **2010,** 5, (6), 1019-32.

6.   Cheng, L. L.; Burns, M. A.; Taylor, J. L.; He, W.; Halpern, E. F.; McDougal, W. S.; Wu, C. L., Metabolic characterization of human prostate cancer with tissue magnetic resonance spectroscopy. *Cancer Res* **2005,** 65, (8), 3030-4.

7.   Somashekar, B. S., magic angle spinning NMR based metabolic profiling of head and neck squamous cell carcinoma tissues. *J Proteome Res* **2011**.

8.   Martineau, E.; Tea, I.; Loaec, G.; Giraudeau, P.; Akoka, S., Strategy for choosing extraction procedures for NMR-based metabolomic analysis of mammalian cells. *Anal Bioanal Chem* **2011,** 401, (7), 2133-2142.

9.   Sellick, C. A.; Hansen, R.; Stephens, G. M.; Goodacre, R.; Dickson, A. J., Metabolite extraction from suspension-cultured mammalian cells for global metabolite profiling. *Nat Protoc* **2011,** 6, (8), 1241-9.

10.   Ackerstaff, E.; Pflug, B. R.; Nelson, J. B.; Bhujwalla, Z. M., Detection of increased choline compounds with proton nuclear magnetic resonance spectroscopy subsequent to malignant transformation of human prostatic epithelial cells. *Cancer Res* **2001,** 61, (9), 3599-3603.

11.   Gu, H.; Pan, Z.; Xi, B.; Hainline, B. E.; Shanaiah, N.; Asiago, V.; Gowda, G. A.; Raftery, D., 1H NMR metabolomics study of age profiling in children. *NMR Biomed* **2009,** 22, (8), 826-33.

12.   Assfalg, M.; Bertini, I.; Colangiuli, D.; Luchinat, C.; Schafer, H.; Schutz, B.; Spraul, M., Evidence of different metabolic phenotypes in humans. *Proc Natl Acad Sci U S A* **2008,** 105, (5), 1420-4.

13.   Lindon, J. C.; Holmes, E.; Nicholson, J. K., Metabonomics and its role in drug development and disease diagnosis. *Expert Review of Molecular Diagnostics* **2004,** 4, (2), 189-199.

14.   Martin, F.-P. J.; Sprenger, N.; Montoliu, I.; Rezzi, S.; Kochhar, S.; Nicholson, J. K., Dietary Modulation of Gut Functional Ecology Studied by Fecal Metabonomics. *J Proteome Res* **2010,** 9, (10), 5284-5295.

15.   Yap, I. K. S.; Li, J. V.; Saric, J.; Martin, F.-P.; Davies, H.; Wang, Y.; Wilson, I. D.; Nicholson, J. K.; Utzinger, J.; Marchesi, J. R.; Holmes, E., Metabonomic and microbiological analysis of the dynamic effect of vancomycin-induced gut microbiota modification in the mouse. *J Proteome Res* **2008,** 7, (9), 3718-3728.

16.   Zhao, L.; Nicholson, J. K.; Lu, A.; Wang, Z.; Tang, H.; Holmes, E.; Shen, J.; Zhang, X.; Li, J. V.; Lindon, J. C., Targeting the Human Genome-Microbiome Axis for Drug Discovery: Inspirations from Global Systems Biology and Traditional Chinese Medicine. *J Proteome Res* **2012,** 11, (7), 3509-3519.

17.   Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **2007,** 26, (1), 51-78.

18.   Somashekar, B. S.; Ijare, O. B.; Nagana Gowda, G. A.; Ramesh, V.; Gupta, S.; Khetrapal, C. L., Simple pulse-acquire NMR methods for the quantitative analysis of calcium, magnesium and sodium in human serum. *Spectrochim Acta A Mol Biomol Spectrosc* **2006,** 65, (2), 254-60.

19.   Kovacs, H.; Moskau, D.; Spraul, M., Cryogenically cooled probes - a leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2005,** 46, (2-3), 131-155.

20.   Lewis, I. A.; Karsten, R. H.; Norton, M. E.; Tonelli, M.; Westler, W. M.; Markley, J. L., NMR method for measuring carbon-13 isotopic enrichment of metabolites in complex solutions. *Anal Chem* **2010,** 82, (11), 4558-63.

21.   Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A.; Zhang, S.; Raftery, D., Chemoselective 15N tag for sensitive and high-resolution nuclear magnetic resonance profiling of the carboxyl-containing metabolome. *Anal Chem* **2009,** 81, (12), 4882-8.

22.   Ye, T.; Zhang, S. C.; Mo, H. P.; Tayyari, F.; Gowda, G. A. N.; Raftery, D., C-13-Formylation for Improved Nuclear Magnetic Resonance Profiling of Amino Metabolites in Biofluids. *Anal Chem* **2010,** 82, (6), 2303-2309.

23.   Hoult, D. I., SOLVENT PEAK SATURATION WITH SINGLE-PHASE AND QUADRATURE FOURIER TRANSFORMATION. *Journal of Magnetic Resonance* **1976,** 21, (2), 337-347.

24.   Piotto, M.; Saudek, V.; Sklenar, V., Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **1992,** 2, (6), 661-5.

25.   Nguyen, B. D.; Meng, X.; Donovan, K. J.; Shaka, A. J., SOGGY: solvent-optimized double gradient spectroscopy for water suppression. A comparison with some existing techniques. *J Magn Reson* **2007,** 184, (2), 263-74.

26.   Carr, H. Y.; Purcell, E. M., EFFECTS OF DIFFUSION ON FREE PRECESSION IN NUCLEAR MAGNETIC RESONANCE EXPERIMENTS. *Physical Review* **1954,** 94, (3), 630-638.

27.	Chan, E. C.; Koh, P. K.; Mal, M.; Cheah, P. Y.; Eu, K. W.; Backshall, A.; Cavill, R.; Nicholson, J. K.; Keun, H. C., Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J Proteome Res* **2009,** 8, (1), 352-61.

28.	Mackinnon, N.; Somashekar, B. S.; Tripathi, P.; Ge, W.; Rajendiran, T. M.; Chinnaiyan, A. M.; Ramamoorthy, A., MetaboID: A graphical user interface package for assignment of (1)H NMR spectra of bodyfluids and tissues. *J Magn Reson* **2013,** 226, 93-9.

29.	Torgrip, R. J. O.; Aberg, K. M.; Alm, E.; Schuppe-Koistinen, I.; Lindberg, J., A note on normalization of biofluid 1D H-1-NMR data. *Metabolomics* **2008,** 4, (2), 114-121.

30.	Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H., Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem* **2006,** 78, (13), 4281-90.

31.	Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C., Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Anal Chem* **2006,** 78, (7), 2262-7.

32.	van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006,** 7, 142.

33.   Barker, M.; Rayens, W., Partial least squares for discrimination. *Journal of Chemometrics* **2003,** 17, (3), 166-173.

34.   Stenlund, H.; Johansson, E.; Gottfries, J.; Trygg, J., Unlocking interpretation in near infrared multivariate calibrations by orthogonal partial least squares. *Anal Chem* **2009,** 81, (1), 203-9.

35.   Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E., Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. *Anal Chem* **2005,** 77, (2), 517-26.

36.   Golland, P.; Liang, F.; Mukherjee, S.; Panchenko, D., Permutation tests for classification. In *Learning Theory, Proceedings*, Auer, P.; Meir, R., Eds. Springer-Verlag Berlin: Berlin, 2005; Vol. 3559, pp 501-515.

37.   Shao, J., Bootstrap model selection. *Journal of the American Statistical Association* **1996,** 91, (434), 655-665.

38.   Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K. F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M., From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **2006,** 34, (Database issue), D354-7.

39.   Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.;

Scalbert, A., HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* **2013,** 41, (D1), D801-7.

CHAPTER 2

Deproteinization and Delipidation of Blood serum for Quantitative NMR-based Metabolomics

2.1 Abstract

In metabolomics studies, conclusions are drawn based on the assumption that the expressed metabolic profiles will vary between sample states (e.g. disease and normal states). Characterizing the concentrations of these differentially expressed metabolites is important for future application in clinical diagnosis of disease based on a threshold concentration level of the metabolite biomarkers.   Thus, sample preparation that enhances the detection of small molecular weight metabolites and facilitates their quantitation in nuclear magnetic resonance (NMR) spectroscopy is highly desired. Blood serum is a commonly used clinical sample that contains both macromolecules such as proteins (e.g. albumin) and lipids (e.g. fatty acids) as well as small molecular weight metabolites.   In this study, we investigate the preparation of biofluids which contain macromolecules for NMR analysis by using human blood serum samples.   Principal component analysis was applied to assess the reproducibility of different preparation protocols, with serum/cold methanol 1:1 (v/v) followed by delipidation using cold chloroform producing the most efficient protein and lipid removal with high reproducibility. We demonstrate that the addition of a known quantity of TSP (Trimethylsilyl propionate

with deuterated methylene groups) incorporated with serum deproteination and delipidation can provide a quantitative reference for reliable estimates of metabolite concentrations in NMR analysis, which was confirmed by authentic metabolites spike-in experiments. This sample preparation protocol results in improved metabolic profiling of serum metabolites and can be applied to quantitation of metabolites for NMR-based metabolomics.

2.2 Introduction

Nuclear magnetic resonance (NMR) spectroscopy has been employed widely in metabolomics research for its quantitative nature and broad coverage of metabolites, including those not amenable to mass spectrometry (e.g. non-ionizable metabolites). NMR measurements exhibit higher reproducibility as compared with mass spectrometry, generating data with less variation that could be attributed to instrumental variation[1]. Metabolic profiling using NMR can be applied to a variety of sample types including biofluids[2, 3, 4, 5, 6, 7, 8], tissues[9, 10, 11, 12] and cell lines[13, 14, 15]. In particular, blood serum, equivalent to blood plasma with clotting factors removed[16], is a commonly used biofluid for the biomarkers discovery of diseases[17, 18]. The problem with analyzing blood serum lies in the fact that NMR resonances of macromolecules (e.g. proteins and lipids) in the serum mask the signals of metabolites in a single pulse NMR experiment, renders the quantitative analysis of metabolites impossible.

Conventionally, for metabolic profiling of serum, T_2 relaxation editing by Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence in an NMR experiment[19, 20] is applied to attenuate signals from macromolecules with short T_2 relaxation times. However, signal intensities from small molecules that bind to proteins will also be reduced. Moreover,

pulse inaccuracies can lead to phase and baseline distortions[21], resulting in inaccurate depictions of the metabolite concentration in the serum.

Instead of CPMG NMR experiment of serum, sample preparation methods such as serum deproteinization by organic solvents or ultrafiltration can be used to retrieve information of the small molecular weight metabolites with improved data quality. For example, protein precipitation by methanol/ethanol (1:1, v/v) and methanol/acetonitrile/acetone (1:1:1, v/v/v) yield a large number of extracted metabolites with good quality data in UPLC/MS measurement [22]. Using 100% methanol has also been shown to be very reproducible and effective in deproteination for metabolic profiling by LC/MS[23]. However, the study of serum sample preparation for quantitative measurement of metabolites by NMR-based analysis is lacking. One method uses ultrafiltration, whereby the serum is applied to a microcentrifuge filter with a specific molecular weight cutoff (e.g. 3 kDa) in order to filter out large molecular weight components. Tiziani et.al[24] have evaluated both deproteinization and ultrafiltration of serum for NMR analysis and proposed that ultrafiltration is more efficient at retaining metabolites after the sample processing and offers higher reproducibility. However, ultrafiltration results in the loss of the hydrophobic fractions of the sample which is useful for lipidomics studies. Moreover, sample processing is time consuming due to multiple washing cycles needed to clean the glycerol in the microcentrifuge filter before applied to the serum.

Targeted profiling was also reported for the quantitative analysis of 1D NMR data[25] where a library containing metabolite spectra obtained in similar experimental conditions

were used to characterize metabolite concentrations in the sample. But this approach requires the Chenomx software which is not readily available.

Though various methods were reported for the serum preparation by MS analysis, there is no optimized and cost-effective preparation protocol available for metabolic profiling as well as metabolite quantification for NMR measurement. Thus, serum preparation using organic solvents for the absolute quantitation of metabolites in 1D ^{1}H NMR spectroscopy was evaluated in this study. In addition, spike-in experiments with five authentic compounds (valine, alanine, glucose, tyrosine and phenylalanine) were performed, showing high accuracy of the protocol for serum metabolites quantification with TSP as an internal standard.   We demonstrate that the absolute quantification of serum metabolites can be achieved by introducing a known quantity of TSP prior to sample deproteinization and delipidation. This protocol allows the utility of simple 1D single pulse NMR experiments for characterizing the concentration of metabolites and can be applied in the determination of cancer biomarker levels in serum or other proteinaceous biofluids (e.g. blood plasma and saliva).

2.3 Experimental design

2.3.1 Materials.

All chemicals including alanine, valine, glucose, tyrosine, phenylalanine, potassium phosphate (monobasic, dibasic), Deuterium oxide (D_2O), methanol, ethanol, chloroform, 3-(Trimethylsilyl)-propionic-2,2,3,3-d4 acid sodium salt (TSP) and human blood sera were purchased from Sigma/Aldrich (Milwaukee, USA).

2.3.2 Protein removal and metabolite extraction.

      To evaluate the optimal protein removal solvent conditions, serum were thawed and transferred to separate Eppendorf tubes with different volumes (300, 400, 500 and 600 μL) and subject to the addition of methanol (4 °C) or mixture of equal volume of methanol and ethanol (4 °C) in 1:1 and 1:2 ratios as shown in Table 2.1. The solutions were vortexed for 30 s and incubated at   -20 0C for an hour. Following the protein precipitation, the solutions were centrifuged at 10,000 RPM for 10 min and equal volume of the supernatant were obtained for further addition of chloroform (4 °C). The mixtures were vortexed briefly and centrifuged (10,000 RPM, 10 min) in order to separate the aqueous and lipid fractions. The aqueous phases were transferred into glass vials for drying in the vacuum concentrator SpeedVac.

Table 2.1 Protocols for serum sample preparation.
Each protocol was tested with different serum quantity represented by experiment 1-4. All volumes were at µL scale.

Protocol a		Serum	MeOH	CHCl$_3$
1:1:1	**Exp. 1**	**300**	**300**	**300**
(v/v/v)	**Exp. 2**	**400**	**400**	**400**
	Exp. 3	**500**	**500**	**500**
	Exp. 4	**600**	**600**	**600**
Protocol b		Serum	MeOH	CHCl$_3$
1:2:2	**Exp. 1**	**300**	**600**	**600**
(v/v/v)	**Exp. 2**	**400**	**800**	**800**
	Exp. 3	**500**	**1000**	**1000**
	Exp. 4	**600**	**1200**	**1200**
Protocol c		Serum	MeOH+EtOH 1:1 (v/v)	CHCl$_3$
1:1:1	**Exp. 1**	**300**	**300**	**300**
(v/v/v)	**Exp. 2**	**400**	**400**	**400**
	Exp. 3	**500**	**500**	**500**
	Exp. 4	**600**	**600**	**600**

2.3.3 TSP as an internal standard for serum extraction.

To test the effectiveness of TSP as an internal standard for serum extraction, two methods were compared using 400 µL of serum separately. In the first method, 50 **µL** TSP solution (1.42 M) was added to the serum before methanol (4°C) 1:1 (v/v)

processing, whereas in the second method, serum protein was removed first by reaction with methanol (4 °C), then 50 μL of TSP was added to the supernatant. Both samples were then extracted with chloroform and the aqueous layer was obtained and dried using SpeedVac.

2.3.4 Metabolite spike-in experiment.

In order to test the validity of serum sample preparation for the quantification of metabolites, spike-in experiments were conducted by the addition of five metabolites in varying weights with TSP as the internal standard. Specifically, a metabolite stock solution was made so that 0.109, 0.091, 0.566, 0.057 and 0.062 mg of valine, alanine, glucose, tyrosine and phenylalanine respectively will present in every 100 uL of the stock solution. Equal volume of the serum sample (400 uL) was taken and each was separately spiked with a known and different amount of the metabolite stock solution (40, 60, 80 and 100 uL). Control sample was prepared using 400 uL of serum without spike-in. We then added 50 uL of TSP solution (1.42 M) to each sample and the samples were all diluted to the same volume (550 uL) by the addition of MilliQ water. The resulting samples were then subject to protein precipitation by methanol (4 °C) in 1:1 (v/v) and further extraction as described in protocol a. Mixture of 100 uL of metabolite stock solution and 50 uL of TSP solution (1.42 M) were prepared to provide the theoretical quantification of the spiked in metabolites. All experiments described above were performed in triplicate.

2.3.5 NMR experiments and data processing.

The dried aqueous extracts were reconstituted in 540 μL of phosphate buffer (0.1 M, pH 7.2, prepared in D_2O). Samples were vortexed briefly and transferred to 5 mm NMR tubes for measurement. For the comparison of extraction protocols, a co-axial

capillary containing 60 μL of TSP (1.16 M, prepared in D_2O) was used to provide an external reference signal.

All measurements were performed at 298 K on a Bruker AVANCE 500 MHz spectrometer (Bruker BioSpin, Karlsruhe, Germany) equipped with a BBI probe. For each 1D ^1H spectrum, a single pulse experiment with a 45° flip angle and water suppression were used. Experiments were conducted with a 17 ppm spectral width, 5 s recycle delay and 256 transients each containing 32K data points. The acquired free induction decay (FID) data were processed using ACD/NMR Processor 12.01 (ACD/Labs, Toronto, Canada) in which 2x zero-filling and an exponential window function equivalent to 0.3 Hz line broadening were applied before Fourier transformation. The spectra were phased, baseline corrected, referenced to TSP at 0 ppm, and exported as an ASCII file. The ASCII files were imported into Excel and the spectral region from -0.1 ppm to 10 ppm was chosen for subsequent analysis. Spectral alignment was performed in SpecAlign and the spectral data were normalized to the integral of TSP signal. Additionally, water (4.6-5.15 ppm) and methanol (3.33-3.4 ppm) intensities were set to zero.

The signal assignment of the 1D spectra were confirmed by the 2D ^1H–^1H total correlation spectroscopy (TOCSY) NMR experiments. The data were acquired with the MLEVPRTP pulse sequence (from the Bruker pulse program library) using a spin-lock mixing time of 70 ms. A spectral width of 8503.40 Hz was used in both dimensions, and 256 t_1 increments were acquired, each containing 2048 complex data points. The number of transients per t_1 increment was 64, and the relaxation delay was set to 5 s. All 2D spectra were processed in TopSpin (version 2.0, Bruker Biospin).

For quantitative NMR, it is necessary to re-establish the magnetization of nuclei in between FID acquisition and the application of a next pulse, which is affected by the setting of relaxation delay in the experiment. This parameter is usually determined at 5 times the longest longitudinal proton relaxation time T_1 in the sample and also is dependent on the flip angle of the excitation pulse. To ensure sufficient relaxation delay was given in the 1D ^{1}H experiment, the relaxation time constant T_1 was measured for all protons in a serum extract. An inversion recovery experiment with water suppression was applied and 17 variable delays ranging from 0.0001 to 10 s were utilized. Peaks from non-overlapped regions were integrated and subjected to T_1 calculation.

2.3.6 Statistical analysis.

To reduce the dimensionality and examine the structure of the data set, principal component analysis (PCA) was performed on the mean-centered data set of serum extraction using protocols a, b and c. Scores plots were generated to assess the clustering of samples in different extraction conditions. The statistical tests were performed in Matlab (R2010a).

2.3.7 Quantitation of serum metabolites.

Metabolite signal integrals were determined from characteristic peak regions of the five metabolites presented in the stock solution. Especially, chemical shift region from 3.1650-3.2615 ppm was used to calculate the glucose integral since glucose level from both α and β anomeric formations were demonstrated in this region. TSP and lactate integrals were obtained as quantitation references. Furthermore, we calculated the integral ratio of metabolite signal with respect to TSP and subtracted the ratio calculated

from the 400 µL serum control samples to estimate the observed quantity of the spiked in metabolites.

Quantification of serum metabolites may be determined according to Equation 2.1:

$$Metabolite \text{ Wt. (mg)} = \frac{TSP\ Wt.(mg)}{TSP\ MW} \times \frac{Metabolite\ peak\ integral}{TSP\ peak\ integral}$$
$$\times \frac{No.\ of\ proton\ (TSP)}{No.\ of\ proton\ (metabolite\ peak)} \times Metabolite\ MW$$

Equation 2.1

where the No. of proton from the metabolite peak are the ones that contribute to the determination of the peak integral.

2.4 Results and Discussion

2.4.1 TSP as an internal quantitative reference for serum preparation and analysis by NMR.

Absolute quantification of metabolites in NMR analysis requires the addition of an internal standard that neither interacts with the sample system nor appears at a chemical shift that overlaps with other metabolite signals[26]. Formate has been reported as a quantitative reference for serum analysis by NMR[27]. However, the varied endogenous quantity of formate in serum of different individuals renders this compound inappropriate for quantitative analysis. External quantitative reference can also be used to avoid reaction with the samples by putting the external reference compound in a coaxial capillary which then inserted into the NMR tube[28]. The drawback of using an external reference is that the loss of metabolites during sample processing cannot be revealed due to reference to a fixed compound concentration in the capillary.

TSP is a common chemical shift reference for NMR measurement due to its non-reactivity with metabolites and resonance in the high field region isolated from metabolite signals in the spectrum. It also serves as a quantitative reference for urinary metabolite quantification[29] in ^{1}H NMR study. However, as for an internal quantitative reference compound, TSP cannot be added to samples containing high protein levels such as tissue and blood serum with no further processing of the mixture due to the interaction of TSP with proteins[30]. We tested serum sample preparation by the addition of a known quantity of TSP before continuing to deproteinization and delipidation to assess whether this sample processing can remove the effect of the interaction between TSP and the macromolecules. As illustrated in Figure. 2.1, no loss of TSP was observed during protein removal, indicating the effect of weak binding of TSP with serum proteins were removed, which can be explained by the high efficiency of protein removal, causing the dissociation of the bound TSP. Thus, TSP can be used as an internal standard to monitor the recovery of extracted metabolites.
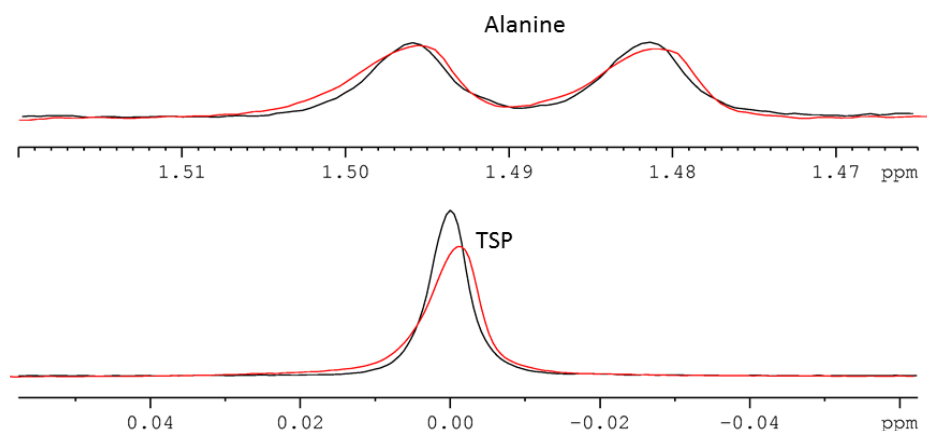
Figure 2.1 TSP as a quantitative reference for serum metabolite quantification.
1D ^{1}H NMR spectra of selected regions. Red spectrum: from a sample with the addition of TSP before protein removal by cold methanol. Black spectrum: from a sample with the addition of equal quantity of TSP after protein removal. The peak area of alanine is the same since equal amount of serum was used in both experiments. TSP peak areas matched well with a less than 10% difference.

To ensure quantitative NMR analysis, the proper relaxation delay need to be given that allows the magnetization to relax to equilibrium before the execution of a next pulse. Thus, longitudinal magnetization recovery experiments were conducted to characterize the T_1 relaxation time of various metabolites. The calculated T_1 values (data not shown) proved the efficacy of using a relaxation delay of 5 s when a 45° pulse was used in our NMR experiment.

2.4.2 Comparison of serum preparation protocols.

To assess the efficiency of protein removal and metabolite extraction, Figure 2.2 demonstrated the resulting NMR spectra after serum preparation by protocols a, b and c. Protocol a applies 100% methanol for protein removal with serum to organic solvent volume ratio of 1:1, whereas protocol c utilizes 1:1 (v/v) mixture of methanol and ethanol in replacement of the pure methanol. Similar metabolite features were observed for protocols a and c, with slightly reduced peak intensity of histidine in protocol c. Protocol b assessed protein precipitation with serum to methanol in 1:2 (v/v), yielding broad peaks around 0.7-0.9 and 1.1-1.3 ppm due to the insufficient removal of lipids and proteins. This result is predictable based on the observation during sample processing where protein was further extracted out with a layer in between solvents after the addition of chloroform for protocols a and c but not for protocol b.
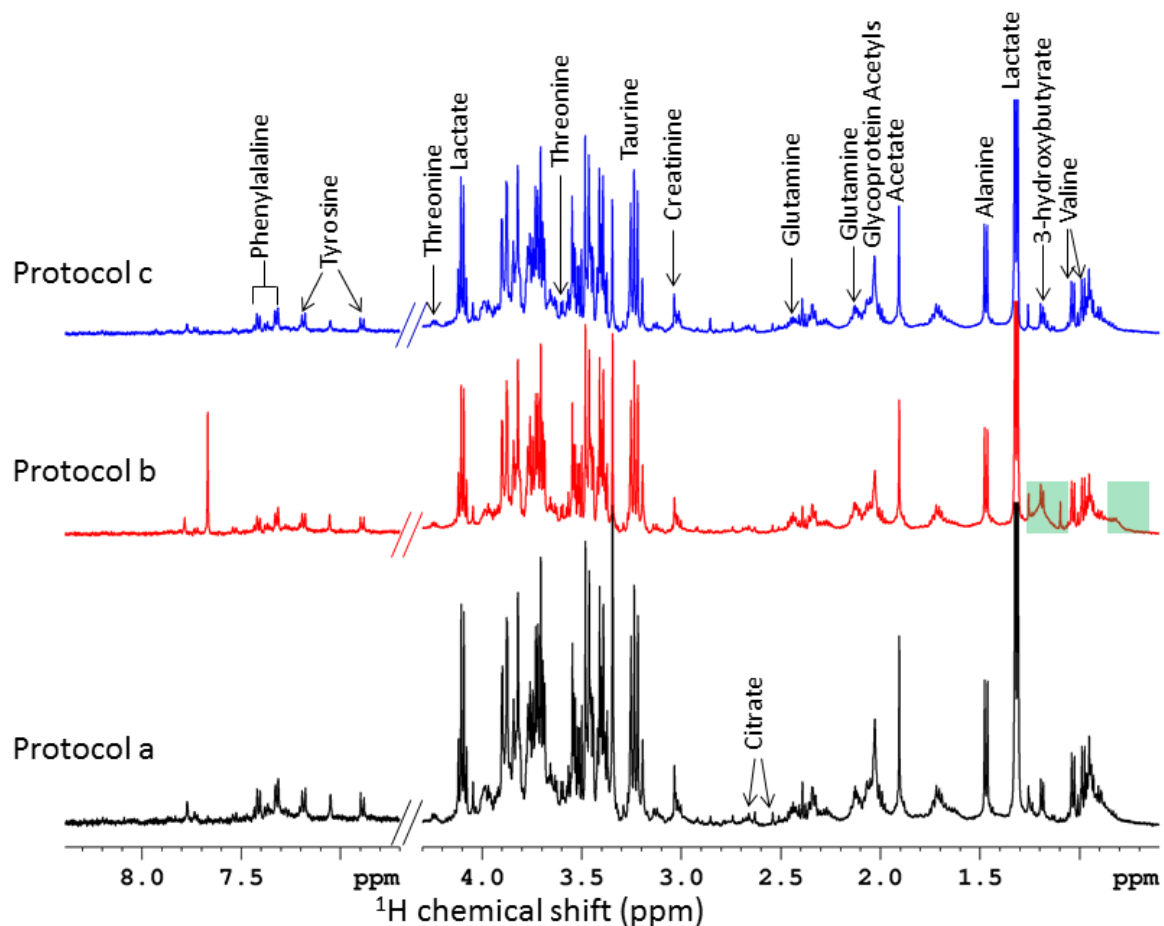
Figure 2.2 Metabolic profiles based on different extraction protocols.
Representative 1D ^{1}H NMR spectra from samples processed by protocol a, b and c. The green regions highlight signals originate from large molecular weight components.

2.4.3 PCA analysis of serum sample preparation methods.

NMR data of different sample preparation protocols a, b and c were normalized to TSP integral and subject to multivariate statistical analysis PCA to assess the data structure and protocol reproducibility. As shown in Figure 2.3, the triplicate samples of protocol a grouped together tightly, symbolizing good reproducibility with the difference along PC 1 arising from the various serum sample quantity used in the sample preparation.

30

However, the lack of reproducibility of samples of protocol b indicates that the volume ratio of serum to organic solvents is crucial for the variation in the data. Interestingly, the addition of more organic solvent does not necessarily improve the deproteination and delipidation efficiency of serum as demonstrated in the comparison of samples from protocol a and b. Protocol c demonstrated a separation trend along PC 1 and the clustering of triplicate samples, which is in agreement with the study of the preparation of plasma for metabolic profiling by UPLC-MS[22], showing good reproducibility for methanol/ethanol deproteinization.
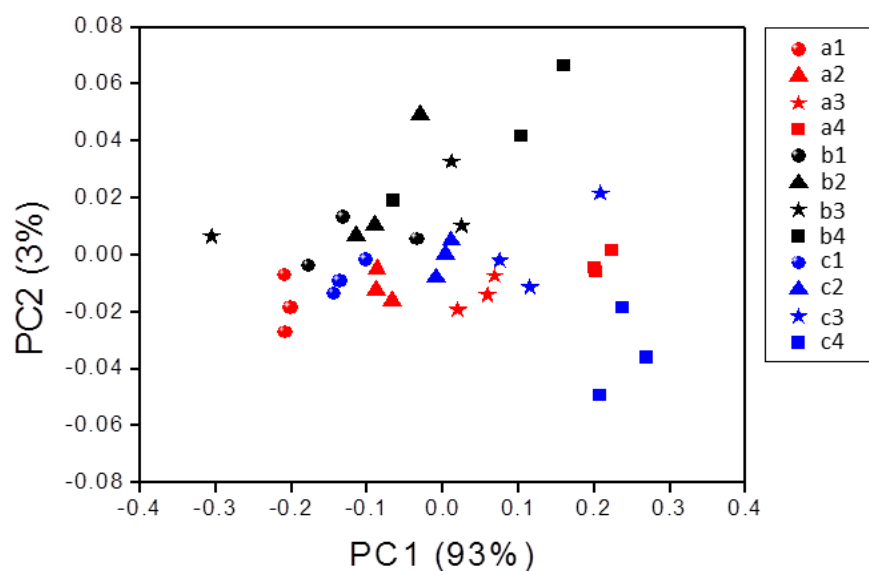
Figure 2.3 PCA scores plot of 1D 1H NMR data of serum samples prepared with protocols a, b and c.

2.4.4 Spike-in experiment proved the accuracy of quantitative analysis.

NMR peak integrals were obtained for the five spiked-in metabolites in samples containing the 400 μL of serum (control samples) and samples consisting of equal volume of serum (400 μL) but different volumes of the metabolite stock solution. Subtracting

from the controls yields the observed quantity due to the amount of metabolite stock solution added.   Theoretical spiked-in metabolite levels were obtained by calculating the metabolite peak integrals of the 100 µL stock solution prepared with the addition of equal quantity of TSP (50 µL ,1.42 M).   All integrals were further characterized relative to TSP and lactate in each spectrum since equal concentration of lactate and TSP were present across all samples.   Figure 2.4 shows the plot of the spiked-in metabolite quantity observed by NMR with respect to the theoretical quantity added.   Great linearity was observed with R^2 > 0.99 for all the spiked metabolites, which reflected the high accuracy of this sample preparation method for the metabolite quantification in NMR-based analysis.
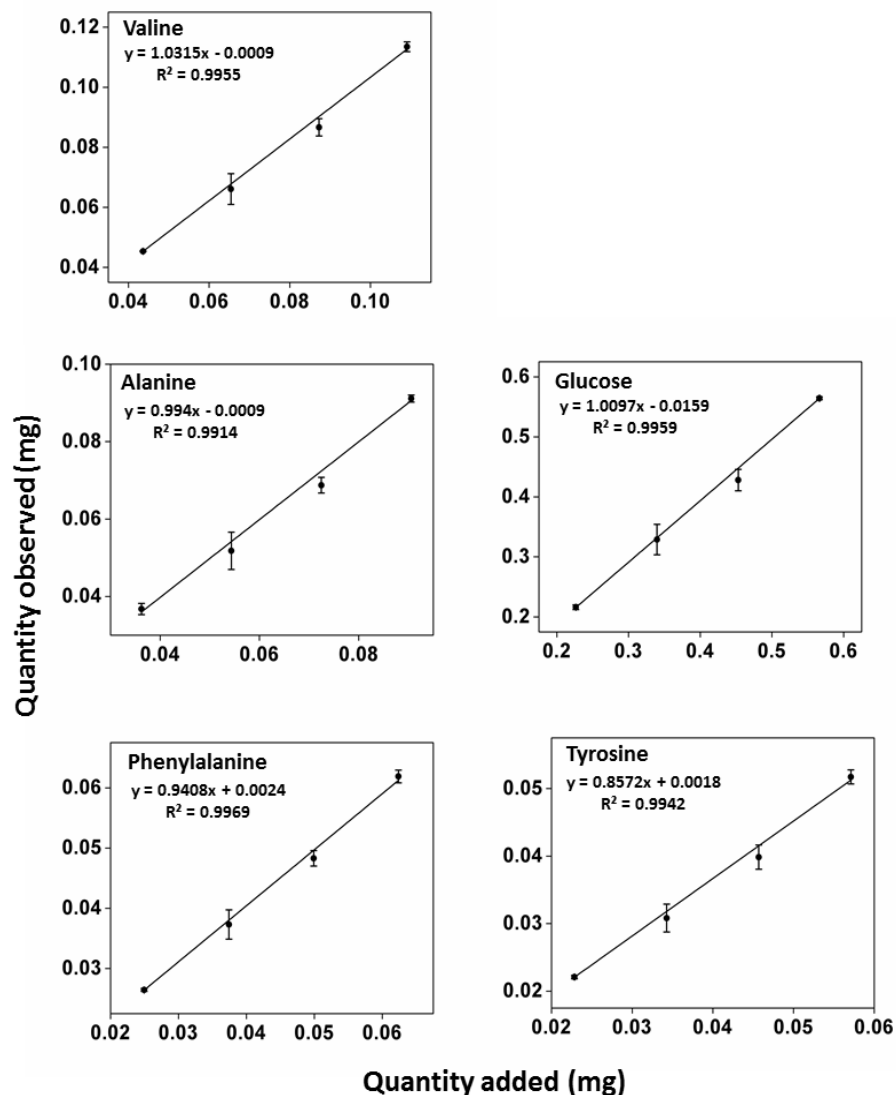
Figure 2.4 Calibration curves obtained for the authentic metabolites spike-in experiment.
Observed metabolite quantity in the y-axis represents the spiked-in metabolite quantity resulted from subtracting the observed quantity of the control samples from the true sample metabolite concentrations in NMR spectra. The high R^2 values demonstrate the validity of the sample preparation for the absolute quantification of metabolites. Error bars represent the standard deviation of three replicate sample measurements with the same amount of spiked in metabolites.

2.5 Conclusion

Quantitation of serum metabolites by metabolic profiling in NMR is commonly applied in studies such as biomarker discover of specific diseases. However, due to the presence of macromolecules in the serum that give rise to broad peaks in the single pulse

33

1D ^{1}H NMR spectrum, CPMG NMR is often used to suppress the broad peaks to facilitate the quantification of small molecular weight metabolites. The quantitative result is arbitrary due to the simultaneous T_2 relaxation processes of both the macromolecules and the metabolites. And the signal to noise no longer serves as an accurate measurement of the metabolite concentrations. Thus, to obtain the absolute quantitative information of serum metabolites, we propose deproteination and delipidation of serum with TSP as an internal standard for NMR-based metabolic profiling (Figure 2.5). Our result shows that serum proteins can be effectively removed by 1:1 (v/v) mixture with cold methanol and further delipidation by the addition of cold chloroform in volume ratio of 1:1 relative to serum. The sample should be left in -20 °C for at least 30 min to allow fully protein precipitation. It is proved by the spike-in experiments that the absolute concentration of metabolites can be obtained with respect to TSP due to the fact that a known quantity of TSP was added and any loss of metabolites during the sample processing will also be reflected in the loss of TSP. This serum sample preparation method allows the application of 1D ^{1}H single pulse NMR experiment with water suppression, which demonstrates accurate quantification of metabolites, thus it can be applied systematically in quantitative NMR analysis of samples containing macromolecules.
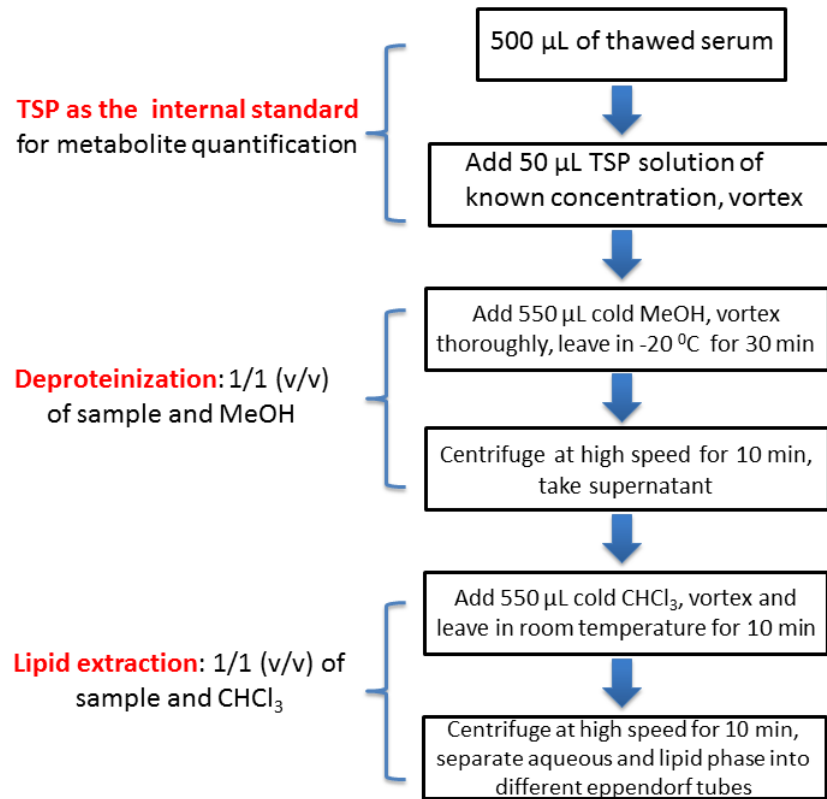
Figure 2.5 Proposed method of serum preparation for quantitation of metabolites by NMR.

2.6 References

1.   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 (R Coll Radiol)* **2012,** 24, (8), 545-55.

2.   Diaz, S. O.; Barros, A. S.; Goodfellow, B. J.; Duarte, I. F.; Carreira, I. M.; Galhano, E.; Pita, C.; Almeida Mdo, C.; Gil, A. M., Following healthy pregnancy by nuclear magnetic resonance (NMR) metabolic profiling of human urine. *J Proteome Res* **2013,** 12, (2), 969-79.

3.   Gronwald, W.; Klein, M. S.; Zeltner, R.; Schulze, B. D.; Reinhold, S. W.; Deutschmann, M.; Immervoll, A. K.; Boger, C. A.; Banas, B.; Eckardt, K. U.; Oefner, P. J., Detection of autosomal dominant polycystic kidney disease by NMR spectroscopic fingerprinting of urine. *Kidney Int* **2011,** 79, (11), 1244-53.

4.   Fathi, F.; Kyani, A.; Darvizeh, F.; Mehrpour, M.; Tafazzoli, M.; Shahidi, G., Relationship Between Serum Level of Selenium and Metabolites Using (HNMR)-H-1-Based Metabonomics in Parkinson's Disease. *Applied Magnetic Resonance* **2013,** 44, (6), 721-734.

5.   Brindle, J. T.; Nicholson, J. K.; Schofield, P. M.; Grainger, D. J.; Holmes, E., Application of chemometrics to H-1 NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. *Analyst* **2003,** 128, (1), 32-36.

6.   Barton, R. H.; Nicholson, J. K.; Elliott, P.; Holmes, E., High-throughput H-1 NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *International Journal of Epidemiology* **2008,** 37, 31-40.

7.   Holmes, E.; Tsang, T. M.; Huang, J. T. J.; Leweke, F. M.; Koethe, D.; Gerth, C. W.; Nolden, B. M.; Gross, S.; Schreiber, D.; Nicholson, J. K.; Bahn, S., Metabolic profiling of CSF: Evidence that early intervention may impact on disease progression and outcome in schizophrenia. *Plos Medicine* **2006,** 3, (8), 1420-+.

8.   Wei, S.; Liu, L.; Zhang, J.; Bowers, J.; Gowda, G. A. N.; Seeger, H.; Fehm, T.; Neubauer, H. J.; Vogel, U.; Clare, S. E.; Raftery, D., Metabolomics approach for predicting response to neoadjuvant chemotherapy for breast cancer. *Molecular oncology* **2013,** 7, (3), 297-307.

9.   Jimenez, B.; Mirnezami, R.; Kinross, J.; Cloarec, O.; Keun, H. C.; Holmes, E.; Goldin, R. D.; Ziprin, P.; Darzi, A.; Nicholson, J. K., H-1 HR-MAS NMR Spectroscopy of Tumor-Induced Local Metabolic "Field-Effects" Enables Colorectal Cancer Staging and Prognostication. *Journal of Proteome Research* **2013,** 12, (2), 959-968.

10.   Somashekar, B. S.; Kamarajan, P.; Danciu, T.; Kapila, Y. L.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., Magic Angle Spinning NMR-Based Metabolic Profiling of Head and Neck Squamous Cell Carcinoma Tissues. *Journal of Proteome Research* **2011,** 10, (11), 5232-5241.

11.   Chan, E. C. Y.; Koh, P. K.; Mal, M.; Cheah, P. Y.; Eu, K. W.; Backshall, A.; Cavill, R.; Nicholson, J. K.; Keun, H. C., Metabolic Profiling of Human Colorectal Cancer Using High-Resolution Magic Angle Spinning Nuclear Magnetic Resonance (HR-MAS NMR)

Spectroscopy and Gas Chromatography Mass Spectrometry (GC/MS). *Journal of Proteome Research* **2009,** 8, (1), 352-361.

12.   Rocha, C. M.; Barros, A. S.; Gil, A. M.; Goodfellow, B. J.; Humpfer, E.; Spraul, M.; Carreira, I. M.; Melo, J. B.; Bernardo, J.; Gomes, A.; Sousa, V.; Carvalho, L.; Duarte, I. F., Metabolic Profiling of Human Lung Cancer Tissue by H-1 High Resolution Magic Angle Spinning (HRMAS) NMR Spectroscopy. *Journal of Proteome Research* **2010,** 9, (1), 319-332.

13.   Tripathi, P.; Kamarajan, P.; Somashekar, B. S.; MacKinnon, N.; Chinnaiyan, A. M.; Kapila, Y. L.; Rajendiran, T. M.; Ramamoorthy, A., Delineating metabolic signatures of head and neck squamous cell carcinoma: phospholipase A2, a potential therapeutic target. *Int J Biochem Cell Biol* **2012,** 44, (11), 1852-61.

14.   Bayet-Robert, M.; Loiseau, D.; Rio, P.; Demidem, A.; Barthomeuf, C.; Stepien, G.; Morvan, D., Quantitative Two-Dimensional HRMAS H-1-NMR Spectroscopy-Based Metabolite Profiling of Human Cancer Cell Lines and Response to Chemotherapy. *Magnetic Resonance in Medicine* **2010,** 63, (5), 1172-1183.

15.   MacKinnon, N.; Khan, A. P.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., Androgen receptor activation results in metabolite signatures of an aggressive prostate cancer phenotype: an NMR-based metabonomics study. *Metabolomics* **2012,** 8, (6), 1026-1036.

16.   Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner,

R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S., The Human Serum Metabolome. *Plos One* **2011,** 6, (2).

17.   Odunsi, K.; Wollman, R. M.; Ambrosone, C. B.; Hutson, A.; McCann, S. E.; Tammela, J.; Geisler, J. P.; Miller, G.; Sellers, T.; Cliby, W.; Qian, F.; Keitz, B.; Intengan, M.; Lele, S.; Alderfer, J. L., Detection of epithelial ovarian cancer using H-1-NMR-based metabonomics. *International Journal of Cancer* **2005,** 113, (5), 782-788.

18.   Bertini, I.; Calalbro, A.; De Carli, V.; Luchinat, C.; Nepi, S.; Porfirio, B.; Renzi, D.; Saccenti, E.; Tenori, L., The Metabonomic Signature of Celiac Disease. *Journal of Proteome Research* **2009,** 8, (1), 170-177.

19.   Carr, H. Y.; Purcell, E. M., EFFECTS OF DIFFUSION ON FREE PRECESSION IN NUCLEAR MAGNETIC RESONANCE EXPERIMENTS. *Physical Review* **1954,** 94, (3), 630-638.

20.   Meiboom, S.; Gill, D., MODIFIED SPIN-ECHO METHOD FOR MEASURING NUCLEAR RELAXATION TIMES. *Review of Scientific Instruments* **1958,** 29, (8), 688-691.

21.   Song, Y. Q., Categories of coherence pathways for the CPMG sequence. *J Magn Reson* **2002,** 157, (1), 82-91.

22.   Bruce, S. J.; Tavazzi, I.; Parisod, V.; Rezzi, S.; Kochhar, S.; Guy, P. A., Investigation of Human Blood Plasma Sample Preparation for Performing Metabolomics Using Ultrahigh Performance Liquid Chromatography/Mass Spectrometry. *Analytical Chemistry* **2009,** 81, (9), 3285-3296.

23.   Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G., Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal Chem* **2006,** 78, (3), 743-752.

24.   Tiziani, S.; Einwas, A.-H.; Lodi, A.; Ludwig, C.; Bunce, C. M.; Viant, M. R.; Guenther, U. L., Optimized metabolite extraction from blood serum for H-1 nuclear magnetic resonance spectroscopy. *Analytical Biochemistry* **2008,** 377, (1), 16-23.

25.   Kovacs, H.; Moskau, D.; Spraul, M., Cryogenically cooled probes - a leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2005,** 46, (2-3), 131-155.

26.   Lewis, I. A.; Karsten, R. H.; Norton, M. E.; Tonelli, M.; Westler, W. M.; Markley, J. L., NMR method for measuring carbon-13 isotopic enrichment of metabolites in complex solutions. *Anal Chem* **2010,** 82, (11), 4558-63.

27.   Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K. F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M., From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **2006,** 34, (Database issue), D354-7.

28.   Prensner, J. R.; Rubin, M. A.; Wei, J. T.; Chinnaiyan, A. M., Beyond PSA: the next generation of prostate cancer biomarkers. *Sci Transl Med* **2012,** 4, (127), 127rv3.

29.   Roobol, M. J.; Haese, A.; Bjartell, A., Tumour markers in prostate cancer III: biomarkers in urine. *Acta Oncol* **2011,** 50 Suppl 1, 85-9.

30.   Armstrong, A. J.; Eisenberger, M. A.; Halabi, S.; Oudard, S.; Nanus, D. M.; Petrylak, D. P.; Sartor, A. O.; Scher, H. I., Biomarkers in the management and treatment of men with metastatic castration-resistant prostate cancer. *Eur Urol* **2012,** 61, (3), 549-59.

CHAPTER 3

Optimized urinary metabolite separation into two-phase solvent system for improved signal assignment in NMR-based metabolomics

3.1 Abstract

Urine metabolic profiling is increasingly used for disease and drug metabolism research by nuclear magnetic resonance (NMR) spectroscopy which has the unique advantage of nonselective and nondestructive detection of metabolites. However, the constituent complexity of urine imposes challenges to NMR-based urine metabolomics studies with difficulties arising from accurate signal assignment of peaks from the highly congested ^{1}H NMR spectra. In this study, we demonstrate an approach to overcome the spectral crowding issue by employing a two-phase solvent extraction (methanol/chloroform/water extraction) on urine before NMR analysis for the purpose of separating urinary metabolites into either polar or non-polar phases. Measurement of each phase individually is then hypothesized to yield simplified metabolic profiles of urine, leading to improved spectral assignment. Extraction conditions were compared with varying water volumes and water pH in 2:1 (v/v) chloroform/methanol mixtures. Our results demonstrate that the efficiency of metabolite extraction was dependent on water volume while independent of water pH, and the protocol for optimal distribution of metabolites into the two phases was to use a minimal water volume for extraction. This separation approach is beneficial as it enhances the ability to assign NMR resonances of

urinary metabolites for subsequent analyses and applications such as cancer biomarker identification.

3.2 Introduction

Characterizing low molecular weight metabolites [1, 2] in biofluids (e.g. serum, plasma, urine, amniotic fluid, cerebrospinal fluid, saliva and other digestive fluids) [3, 4, 5, 6], tissues [7, 8, 9, 10] and cells [11, 12, 13] have been extensively studied to get insights of the perturbation to an organism which may have been caused by diseases [14], aging [15] or external stimuli (e.g. drug treatment) [16, 17, 18, 19, 20]. Among these sample types, urine, the metabolic end product that reflects the physiological conditions of the organism, is a favored choice both in clinical screening and metabolic phenotype research for its noninvasive nature and ease of collection. One of the analytical techniques for metabolomics study, nuclear magnetic resonance (NMR) spectroscopy has been employed widely for its quantitative nature and broad coverage of metabolites, including those not amenable to mass spectrometry (e.g. non-ionizable metabolites). NMR measurements exhibit high reproducibility as compared with mass spectrometry, generating data with less variation that could be attributed to instrumental variation. Commonly one-dimensional ^{1}H NMR spectra are sufficient and acquired for subsequent statistical analysis including both univariate and multivariate methods. Metabolite signal assignment is usually performed by searching chemical shift lists, for example, in HMDB [21], BMRB [22], Chenomx software (Chenomx Inc., Edmonton, Candada) and published literatures [23, 24] combined with the analysis of specific *J*-coupling patters of the spin systems of the compound. Moreover, this process is more efficient now with the help of MetaboID [25], a graphical user interface that offers comparison of experimental spectra

with authentic metabolite spectra to guide assignment. However, the complexity of biological sample components, in particular urine, which contains a large variety of compounds (e.g. aromatic, aliphatic compounds, amino acids, sugars, and osmolytes) existing in a broad range of concentrations result in complicated NMR spectra and thus spectral assignment is a significant challenge. One approach to cope with sample complexity is separation of metabolites prior to NMR analysis. Multiple studies using various separation techniques have been reported for NMR-based metabolomics. For example, two-phase solvent extraction of tissues [26] and cells [27, 28] for the assessment of differences in metabolic phenotype between normal and malignant states of cancer, and high performance liquid chromatography (HPLC) [29] and solid phase extraction (SPE) for urine purification in drug metabolite analysis [30]. In particular, extraction methods such as MeOH/H_2O, ACN/H_2O, MeOH/$CHCl_3$/H_2O and ACN/$CHCl_3$/H_2O have been compared for liver tissue by NMR and LC-MS analysis [31]. The results of Beltran and coworkers indicates MeOH/H_2O treatment followed by MeOH/$CHCl_3$/H_2O extraction is the most efficient protocol, with the extract NMR spectrum having an enhanced signal to noise ratio (S/N) as compared with the raw tissue spectrum. A comparison of various extraction methods on mammalian cells for NMR analysis also demonstrated the efficiency and reliability of MeOH/$CHCl_3$/H_2O extraction of intracellular metabolites [11]. Critically, there is little data currently available that describes the extraction of urine, which would be expected to analogously reduce the complexity and improve the sensitivity for NMR-based metabolomics.

We propose a two-phase extraction protocol of urine for metabolic profiling by NMR and investigate how the separation of metabolites can aid in simplifying the spectra

and enhance the ability of spectral assignment. Our results suggest that the integration of a two-phase urine extraction protocol in the standard sample preparation for metabolomic analyses may be beneficial for the identification of otherwise difficult to assign NMR signals.

3.3 Experimental design

3.3.1 Materials and urine sample.

All chemicals including potassium phosphate (monobasic, dibasic), sodium carbonate, sodium bicarbonate, Deuterium oxide (D_2O), methanol, chloroform and 3-(Trimethylsilyl)-propionic-2,2,3,3-d4 acid sodium salt (TSP) were purchased from Sigma/Aldrich (Milwaukee, USA). All the solutions were prepared in water of MilliQ quality. Urine samples were collected based on a protocol approved by the Institutional Review Board, and urine of 10 different individuals were pooled together and stored at -80 °C. This pooled urine sample was utilized in extraction. Since individual sampling error was eliminated by using this pooled sample, the main source of variation is expected to be caused by the various extraction protocols.

3.3.2 Two-phase urinary metabolite separation.

Three different protocols (Table 3.1) were used to optimize the efficient separation of metabolites in two-phase solvent systems. Before the separation, urinary proteins were removed by adding cold methanol to 500 µL of urine in a 1:1 (v/v) ratio and keeping the mixture at -20 °C for at least 30 min.   Following the protein precipitation, the mixture was centrifuged (10,000 rpm, 10 min), and the supernatant was transferred into glass vials for drying in SpeedVac. Each separation protocol was performed in triplicate. Next,

CHCl$_3$ and CH$_3$OH were added at a 2:1 (v/v) ratio to each of the triplicate sample, mixed thoroughly and kept at room temperature for about 60 min. Two conditions of extraction water were compared in which either the water volume or pH were chosen as variables as described below.

Extraction H$_2$O – varied volume. The amount of water added played an important role in determining the polarity of the extraction solvent system. In this protocol, 120 µL, 90 µL and 60 µL of water were added separately as shown by protocol 1, 2 and 3 to study the effect of water volume on the extraction efficiency.

Extraction H$_2$O – varied pH. This experiment was performed to compare the extraction efficiency of metabolites at multiple pH of extraction water. Typical simple carboxylic acids have an approximate pKa = 5, while primary amine groups have an approximate pKa = 9. In this study, H$_2$0 buffered at pH 3, 7 and 10 were chosen for the purpose of decreasing the aqueous solubility of the carboxylate and amino containing metabolites through protonation or deprotonation, respectively, thereby reducing the congestion of metabolite signals in the polar phase NMR spectra. Extraction water with specified pH was obtained by using 0.1 M phosphate (pH 3, 7) or carbonate (pH 10) buffered solution. The water volume was held at 60 µL for all extractions involving buffered H$_2$O (protocol 4, 5 and 6).

Table 3.1 Extraction protocols of urine samples.
Protocols 1, 2 and 3 test the effect of varied volume of H_2O, whereas protocols 4, 5 and 6 compare the impact of varied pH of H_2O on the extraction.

Protocol	Urine*	MeOH	CHCl$_3$	H$_2$O	MeOH/CHCl$_3$/H$_2$O (v/v/v)
1	500 µL	300 µL	600 µL	120 µL	1: 2: 0.4
2	500 µL	300 µL	600 µL	90 µL	1: 2: 0.3
3	500 µL	300 µL	600 µL	60 µL	1: 2: 0.2
4	500 µL	300 µL	600 µL	60 µL pH 3	1: 2: 0.2
5	500 µL	300 µL	600 µL	60 µL pH 7	1: 2: 0.2
6	500 µL	300 µL	600 µL	60 µL pH 10	1: 2: 0.2

* Urine was treated with 500 µL cold methanol and kept at -20 °C for 1 h. Samples were then centrifuged and the supernatant was dried in SpeedVac.

Regardless of the extraction protocol, after the addition of H_2O, samples were vortexed and centrifuged at 10,000 RPM for 10 minutes. The polar and non-polar phases were collected separately. The non-polar extracts were evaporated to dryness under a stream of N_2 and all extracts were then subjected to SpeedVac drying. Dried samples were stored at -80 °C until NMR measurements.

3.3.3 NMR spectroscopy.

Both polar and non-polar extracts were reconstituted separately in 560 µL of phosphate buffer (0.1 M, pH 7.2, prepared in D_2O) containing TSP ([TSP] = 1.5 mM) as the NMR reference signal. Samples were vortexed and centrifuged briefly with 550 µL of the sample collected for NMR analysis.

All measurements were performed at 293 K on a Bruker AVANCE 500 MHz spectrometer (Bruker BioSpin, Karlsruhe, Germany) equipped with a 5 mm TXI SB probe. For each 1D ^{1}H spectrum, a single pulse experiment with a 45° flip angle and water suppression were used. Experiments were performed with a 14 ppm spectral width, 3 s recycle delay and 128 transients each containing 32K data points. The acquired free induction decay (FID) data were processed using ACD/NMR Processor 12.01 (ACD/Labs, Toronto, Canada) in which 2x zero-filling and an exponential window function equivalent to 0.3 Hz line broadening were applied before Fourier transformation. The spectra were phased, baseline corrected, referenced to TSP at 0 ppm, and exported as an ASCII file. The ASCII files were imported into Excel and the spectral region from -0.02 ppm to 10 ppm was chosen for subsequent analysis. The intensity of each spectrum was normalized to the TSP signal. Additionally, water (4.6-5.15 ppm) and methanol (3.33-3.4 ppm) intensities were set to zero. All 1D spectra were aligned based on the variable-reference interval correlated shifting (vr-icoShift) method [32].

To further reduce the signal overlap and compare metabolic patterns of polar and non-polar extracts, two-dimensional ^{1}H–^{1}H total correlation spectroscopy (TOCSY) NMR experiments were performed on the extracted sample processed according to protocol 3. 2D TOCSY spectra were acquired with the MLEVPRTP pulse sequence (from the Bruker pulse program library) using a spin-lock mixing time of 70 ms. A spectral width of 7002.80 Hz was used in both dimensions, and 256 t_1 increments were acquired, each containing 2048 complex data points. The number of transients per t_1 increment was 64, and the relaxation delay was set to 1.5 s. All 2D spectra were processed in TopSpin (version 2.0, Bruker Biospin).

3.3.4 Statistical analysis.

In order to reduce the dimensionality and examine the structure of the data set, principal component analysis (PCA) was performed on the mean-centered data set. Scores plots were generated to assess the clustering of samples in different extraction conditions, with the corresponding loadings plot examined to identify the variables accounting for the most variation in the specified principal component. The statistical tests were performed in Matlab (R2010a).

3.4 Results and Discussion

3.4.1 Complexity of NMR spectra of urine.

NMR-based metabolomics of urine have been applied to diverse fields of research such as the study of metabolic alterations with drug treatments to delineate the adaptive response to drugs associated with toxicity[33], identification of urinary metabolic biomarkers for early detection of prenatal disorders[34], autosomal dominant polycystic kidney disease[35] and lung cancer[36], and assessment of the effect of physical exercise protocols on the metabolic changes in urine[37]. The complexity of 1D ^{1}H NMR spectra of urine is both beneficial in terms of the rich information content available, and challenging with respect to resonance assignment, where the objective would be to extract the full information content. Spectral complexity of biofluids can be alleviated by chemoselective isotope tagging of carboxyl containing metabolites and amino metabolites, with the limit of detection to µM level. For example, ^{15}N tagging of carboxyl containing metabolites with ^{15}N-ethanolamine[38] and ^{13}C formylation of amino metabolites using ^{13}C formic acid [39]. These methods greatly help in the identification and quantification of only certain class (carboxylic/amine containing) of metabolites but fails to produce the information of

metabolites of the other classes. Due to the various steps involved for the tagging reactions followed by quantification using 2D ^{1}H-^{13}C/^{15}N HSQC, the routine application on the urine is limited for high throughput metabolic profiling by NMR-based analysis. To address the spectral complexity of urine, we performed MeOH/CHCl$_3$/H$_2$O extraction with the goal of separating polar and non-polar metabolites into a two-phase solvent system, which was reported to yield high extraction efficiency of tissues as compared with other extraction methods[31]. Individually, the polar and non-polar phases would then be expected to yield two simplified ^{1}H NMR spectra, potentially aiding in the resonance assignment. We explored the MeOH/CHCl$_3$/H$_2$O extraction with different volumes and pH of H$_2$O since the extraction efficiency and reproducibility may be affected by the pH and polarity of the extraction solvent system. This extraction method is easy to perform and results in a broad coverage of metabolites that may otherwise lost with other processing methods.

3.4.2 Sample pre-processing and optimization of the two-phase solvent systems.

Before the extraction, cold MeOH was used for urinary protein removal with a 1:1 volume ratio of MeOH to urine, which reduces the effective concentration of water, resulting in less hydration of the protein in the urine. To examine the impact of the protein removal step, a control experiment without the protein removal was performed where the native urine was centrifuged and the supernatant was collected for NMR analysis. No distinct differences of the metabolite resonances were observed in the comparison of the native urine with the cold MeOH processed urine, which reveals that the addition of methanol to urine will not result in the loss of any urinary metabolites. Moreover, the protein removal facilitates the separation of the polar and non-polar phases in the

MeOH/CHCl$_3$/H$_2$O extraction by reducing the protein content at the interface of the two layers. Thus all the urine samples were subjected to cold methanol treatment prior to extraction. Notably, methanol treatment for protein removal is reported to be highly effective for tissue [31] and serum [40] sample analysis.

3.4.3 NMR analysis of non-polar and polar solvent phases.

^{1}H NMR spectrum of non-polar part (Figure 3.1 a) obtained with protocol 1(1:2:0.4) showed peaks corresponding to only few metabolites since most of the metabolites were retained in the polar part (Figure 3.1 b). However, non-polar part (Figure 3.2 a) obtained with protocol 2 (1:2:0.3) showed the presence of considerable amount of metabolites. This clearly indicates that reducing the water content in the solvent mixture enhances the chances of distributing the metabolites into the non-polar phase. Therefore, with further reduction of water ratio in protocol 3 (1:2:0.2), non-polar part (Figure 3.3 c) showed >90% of relatively non-polar metabolites whereas highly polar metabolites retained in the polar extract (Figure 3.3 b). In Figure 3.3, selected regions in the black dotted squares highlight metabolites that are dominant in either the non-polar (blue spectrum) or polar (red spectrum) phase. For instance, the citrate signal (2.5 – 2.7 ppm) only appeared in the polar phase, as might be expected of a highly polar small molecule such as citrate which contains a hydroxyl and three carboxyl groups. In contrast, metabolites such as aromatic amino acids (phenylalanine, hippurate), modified aromatic amino acids (phenylacetylglycine, phenylacetylglutamine), branched chain amino acids (valine, leucine, isoleucine) and other relatively non-polar metabolites (dimethylamine, succinic acid) can be easily extracted into the non-polar part, leaving the majority of polar metabolites in the polar extract. Further reduction of water in MeOH/CHCl$_3$ mixture may

result in better separation of polar and non-polar metabolites but is limited due to practical difficulties in separating the two solvent layers. In ^{1}H NMR spectrum of urine, usually the spectral range between 3-4.5 ppm is highly complex in which αCH peaks of amino acids appear along with other metabolites.
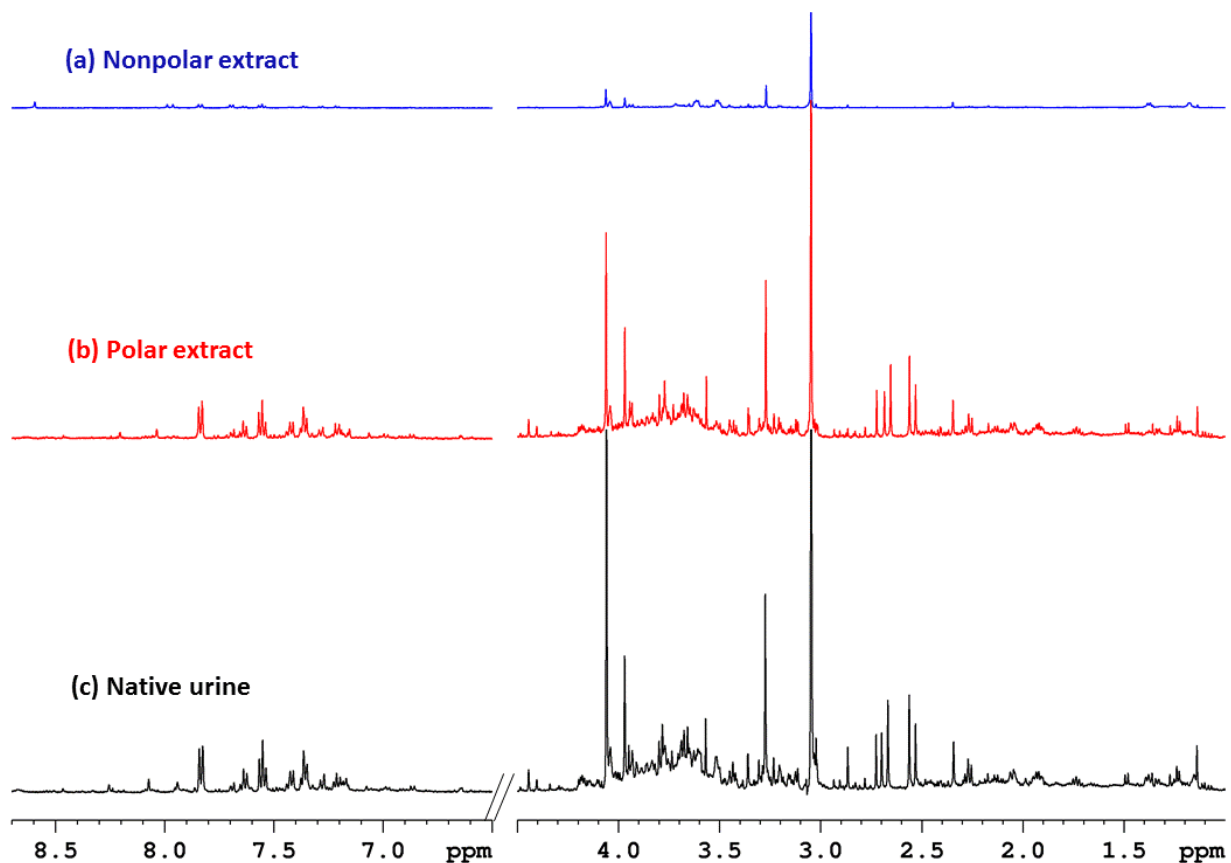
Figure 3.1 Representative 1D ^{1}H NMR spectra of native urine and extracts using protocol 1 (MeOH/CHCl$_3$/H$_2$O volume ratio of 1: 2: 0.4).
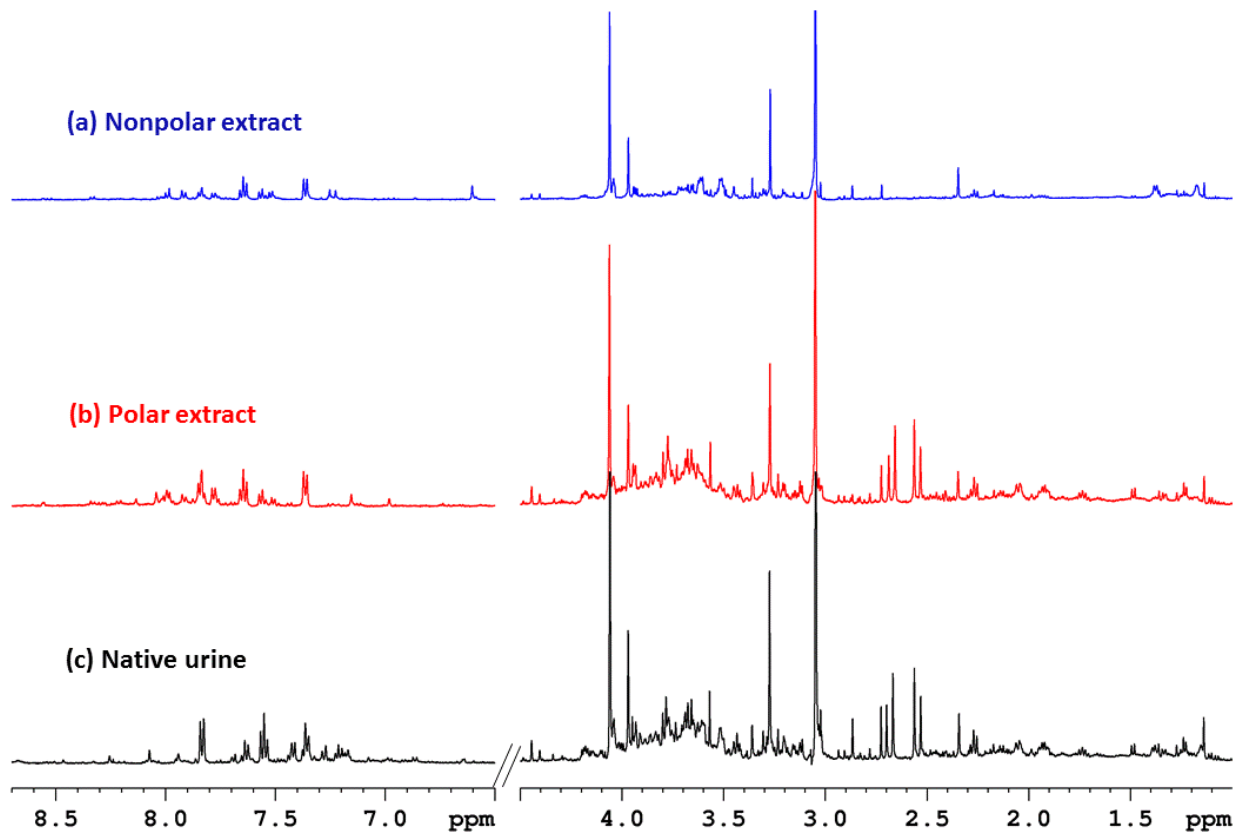
Figure 3.2 Representative 1D ^{1}H NMR spectra of native urine and extracts using protocol 2 (MeOH/CHCl$_3$/H$_2$O volume ratio of 1: 2: 0.3).
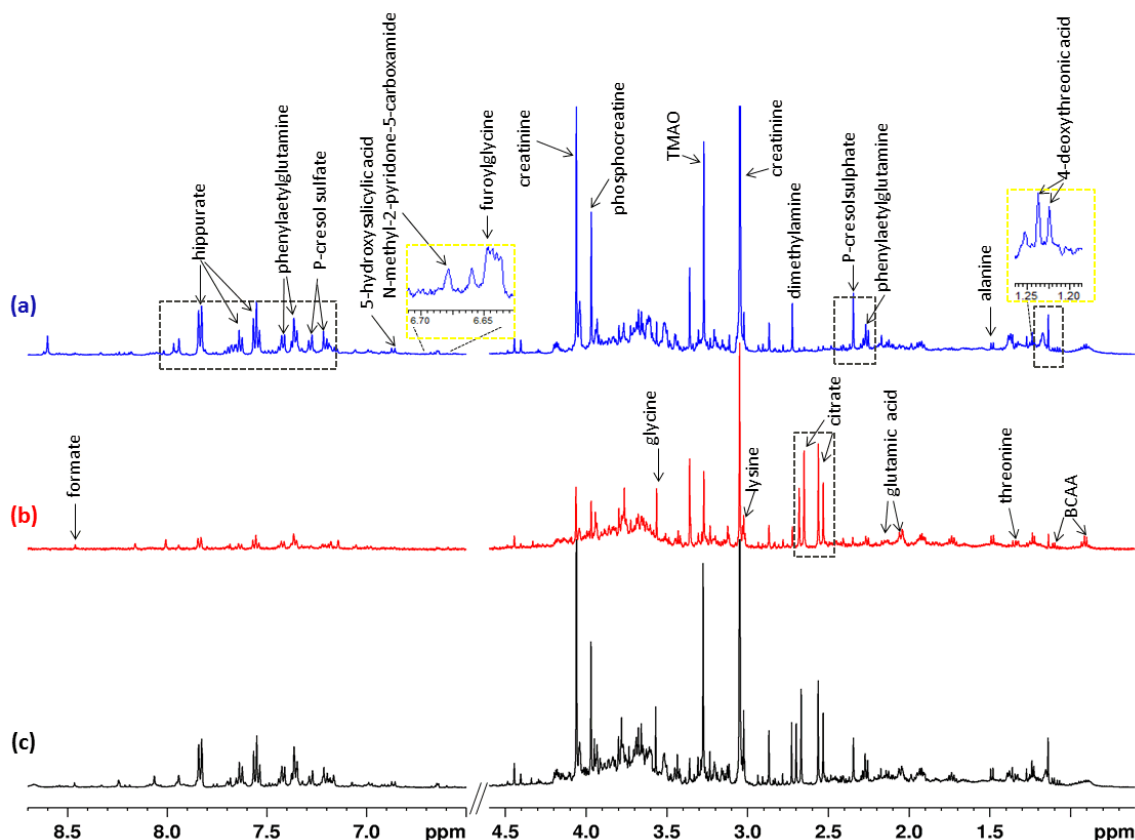
Figure 3.3 Representative 1D ^{1}H NMR spectra of native urine and extracts.
Native urine (black), polar extract (red) and non-polar extract (blue), resulting from the extraction using protocol 3 (MeOH/CHCl$_3$/H$_2$O volume ratio of 1: 2: 0.2). Spectral regions from 0.7 to 4.5 ppm and 6.5 to 8.7 ppm were shown. Black dotted regions highlight resonances that had higher abundance in either the polar or non-polar phase extract. Enlarged spectra at chemical shifts around 1.20-1.25 ppm and 6.65-6.70 ppm were shown in the yellow dotted squares.

By achieving metabolites separation using protocol 3 and measuring the 2D ^{1}H-^{1}H correlation spectrum of non-polar and polar extracts separately, the spectral complexity of urine can be greatly reduced. The simplified spectrum of each phase will aid in the assignment of 2D NMR spectrum of urine where hundreds of cross-peaks appear. For example, we have measured ^{1}H-^{1}H TOCSY of protocol 3 non-polar and polar parts separately and compared with TOCSY of native urine, in which great reduction in spectral complexity can be observed. In the overlay of 2D ^{1}H-^{1}H TOCSY spectra of the polar (blue spectrum) and non-polar extracts (red spectrum) (Figure 3.4), metabolites were

separated well based on the polarity. Non-polar metabolites such as isoleucine and alanine were present exclusively in the non-polar extract whereas polar metabolites such as lysine and glutamic acid were more abundant in the polar extract.
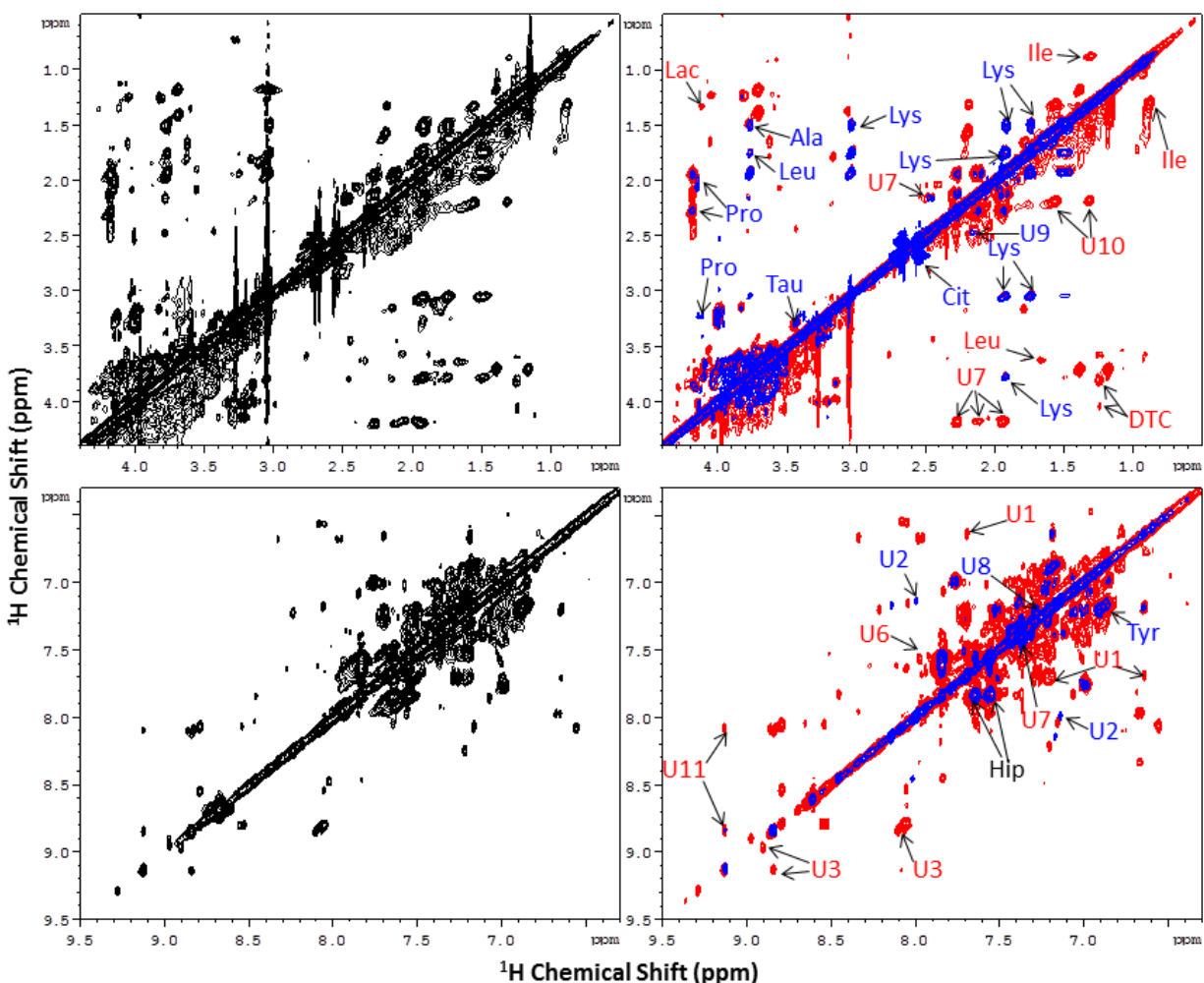
Figure 3.4 2D ^{1}H-^{1}H TOCSY spectra of neat urine and extracts.
Neat urine (black spectra on the left), overlaid spectra of polar (blue spectrum) and non-polar (red spectrum) extracts utilizing protocol 3 with MeOH/CHCl$_3$/H$_2$O (v/v/v) of 1: 2: 0.2. Expansion of the chemical shift regions (0.5-4.5 ppm and 6.3-9.5 ppm) are shown separately in order to highlight signals visible in one phase versus the other. Abbreviations: Ile, isoleucine; Lac, lactate; Gln, glutamine; Lys, lysine; Cit, citric acid; Leu, leucine; Ala, alanine; Pro, proline; Tyr, tyrosine; Tau, taurine; DTC, 4-deoxythreonic acid; U1, 2-furoylglycine; U2, histamine; U3, 1-methylnicotinic acid; U4, 5-hydroxysalicylic acid; U5, o-Cresol; U6, acetophenone; U7, phenylacetylglutamine; U8, p-cresol sulfate; U9, butanone; U10, suberate; U11, 1-methylnicotinic acid; Hip, hippurate.

To examine the influence of water volume on the extraction, we assessed the metabolite separation in the two-phase solvent system by quantifying several signals from protocol 1, 2 and 3. The mean ± standard deviation of integrals from four metabolite signals in both the non-polar and polar extracts are shown in Figure 3.5, clearly demonstrating an increase of the metabolite quantity in the non-polar phase as a function of decreased water content utilized in the solvent mixture. These results reveal that the addition of minimal water into MeOH:CHCl$_3$ solvent systems facilitates the partition of metabolites into the non-polar phase more readily, distinguishing these molecules from the metabolites that are consistently more soluble in the polar phase. Therefore, we propose 1:2:0.2 (MeOH:CHCl$_3$:H$_2$O v/v/v) two-phase solvent system for better separation of urinary metabolites.
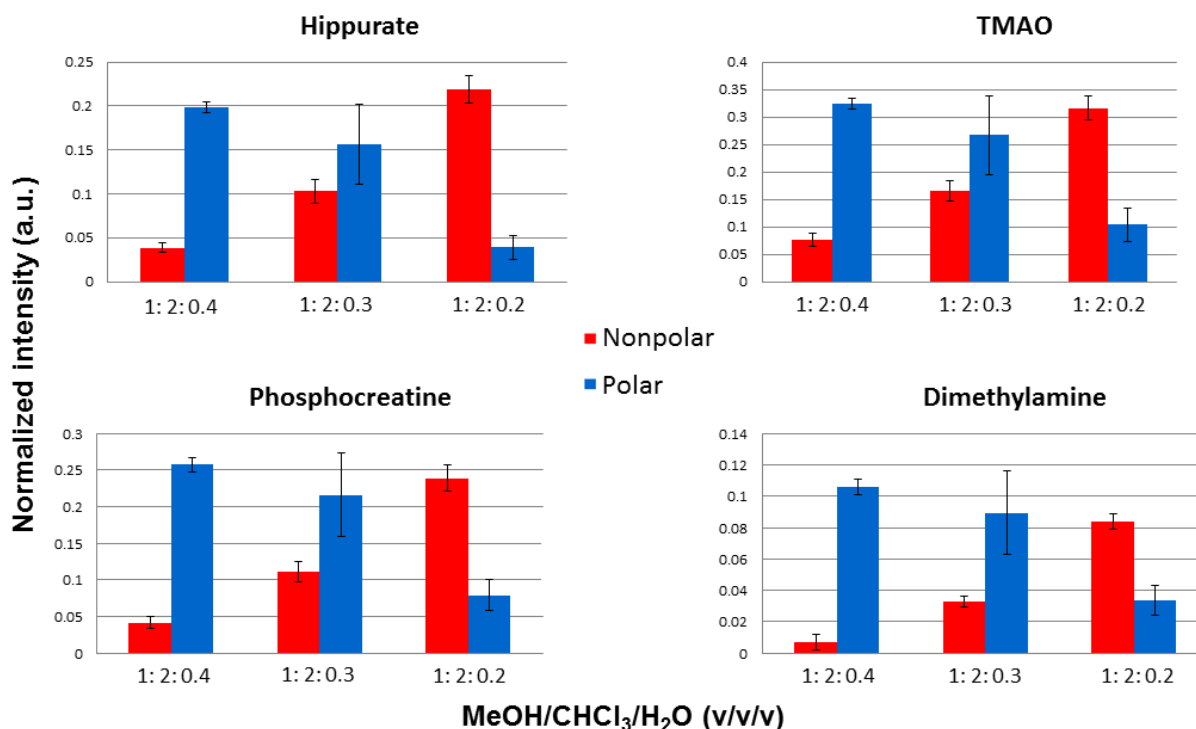
Figure 3.5 Bar plots of mean integrals calculated for metabolite signals in both the polar and non-polar phases with different water volumes in extraction.
Integrals were calculated after normalization to the TSP signal.

3.4.4 PCA analysis.

PCA was performed on the NMR data of both non-polar and polar extracts obtained with different protocols to assess the reproducibility of triplicate measurements as well as the effect of water content on metabolite separation in the two-phase solvent system. A clear separation between non-polar and polar extract NMR metabolic profiles was observed along principal component (PC) 2 (Figure 3.6 a).   Except for the polar extract of protocol 2, all the other triplicate measurements grouped together closely, revealing good reproducibility for each extraction approach. Largest group separation of non-polar and polar extracts along PC 1 was observed for samples obtained from protocol 3, revealing that protocol 3 is best suitable for metabolites separation. Loadings plot along PC 1 (Figure 3.6 b) represents the metabolites that contribute to the largest variance of the data (68.87%). Examination of the loadings along PC 2 (Figure 3.6 c) showed that creatinine, citrate and lysine were crucial for the segregation of the two phases, due to their strong hydrophilic character.
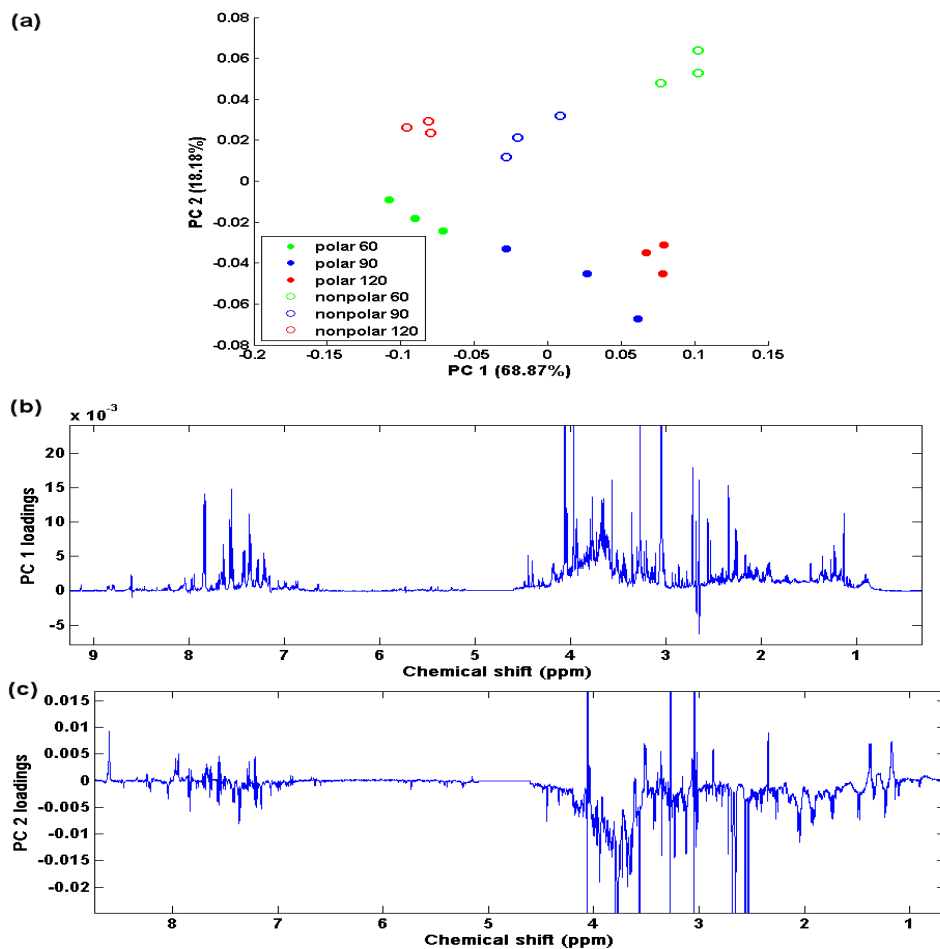
Figure 3.6 Principal component analysis of urinary metabolite extracts with varied water volume.
The scores plot **(a)** displays the relationship between samples with the proportion of variance explained by principal component (PC) 1 and PC 2. Each point represents a single spectrum. Both the polar extracts (filled circles) and non-polar extracts (empty circles) are separated from each other along PC 2, indicating difference in the polar and non-polar extracts NMR spectra. Loadings plots show the resonances that contribute to the variance of the specific principal component. **(b)** PC 1 loadings. **(c)** PC 2 loadings.

To assess the effect of water pH on the metabolites separation, PCA was also performed on NMR spectra of both the non-polar and polar phases obtained by 1:2:0.2 extraction in which the pH of water varied from 3-10. A separation of non-polar and polar phase profiles was observed in the scores plot in Figure 3.7, symbolizing a difference in metabolic profiles of the two phases. The clustering of the non-polar extract triplicates was not very tight, indicating poor reproducibility of the organic extracts. Thus, we focus

more on the water volume rather than the water pH for consistent extraction of urinary metabolites.
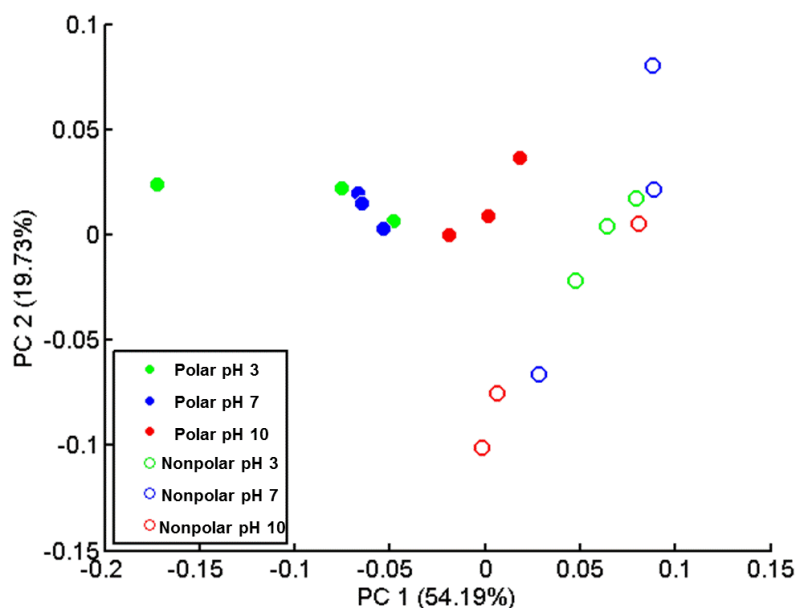
Figure 3.7 PCA scores plot of samples extracted with different water pH.

3.5 Conclusions

Assignment of metabolites in NMR spectra is crucial for subsequent statistical and biological interpretation of data. Urine, a common biofluid utilized in metabolomic studies, poses significant challenges because of the highly congested NMR spectra, hindering the correct assignment of metabolites present in the biological system. In this report, we have introduced the methanol/chloroform/water extraction of urinary metabolites before NMR measurement with the aim of partitioning metabolites with different chemical properties into polar and non-polar solvent systems. The results indicate that using minimal water (MeOH:CHCl$_3$:H$_2$O, 1:2:0.2, v/v/v) for extraction allows more metabolites to be extracted into the non-polar phase and results in simplified urine spectra for each phase. More

extraction solvents can be used for better handling of the separation of the two phases in considering that the same extraction solvent ratio 1:2:0.2 was kept, as well as the proportional urine volume taken for the extraction. Moreover, the simplified metabolite patterns of the polar and non-polar extracts shown in the 2D TOCSY experiment is expected to further enhance metabolite assignment capabilities. We believe that this method of analyzing the effect of two-phase solvent extraction applied to urine with the specific intention of simplifying NMR spectra, providing a means for improved signal assignment. Further studies on the de-convolution of NMR spectral peaks is necessary for better signal identification in NMR-based metabolomics studies in order to reduce erroneous outcomes resulting from misinterpretation/misassignment of NMR spectra.

3.6 References

1.	Clarke, C. J.; Haselden, J. N., Metabolic profiling as a tool for understanding mechanisms of toxicity. *Toxicol Pathol* **2008,** 36, (1), 140-7.

2.	Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K., Global metabolic profiling procedures for urine using UPLC-MS. *Nat Protoc* **2010,** 5, (6), 1005-18.

3.	Barton, R. H.; Nicholson, J. K.; Elliott, P.; Holmes, E., High-throughput H-1 NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *International Journal of Epidemiology* **2008,** 37, 31-40.

4.	Fonville, J. M.; Maher, A. D.; Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Evaluation of Full-Resolution J-Resolved H-1 NMR Projections of Biofluids for Metabonomics Information Retrieval and Biomarker Identification. *Anal Chem* **2010,** 82, (5), 1811-1821.

5.	Schicho, R.; Shaykhutdinov, R.; Ngo, J.; Nazyrova, A.; Schneider, C.; Panaccione, R.; Kaplan, G. G.; Vogel, H. J.; Storr, M., Quantitative Metabolomic Profiling of Serum, Plasma, and Urine by (1)H NMR Spectroscopy Discriminates between Patients with Inflammatory Bowel Disease and Healthy Individuals. *J Proteome Res* **2012**.

6.	Chan, E. C. Y.; Pasikanti, K. K.; Nicholson, J. K., Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat Protoc* **2011,** 6, (10), 1483-1499.

7.   Want, E. J.; Masson, P.; Michopoulos, F.; Wilson, I. D.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Loftus, N.; Holmes, E.; Nicholson, J. K., Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc* **2012,** 8, (1), 17-32.

8.   Beckonert, O.; Coen, M.; Keun, H. C.; Wang, Y.; Ebbels, T. M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues. *Nat Protoc* **2010,** 5, (6), 1019-32.

9.   Cheng, L. L.; Burns, M. A.; Taylor, J. L.; He, W.; Halpern, E. F.; McDougal, W. S.; Wu, C. L., Metabolic characterization of human prostate cancer with tissue magnetic resonance spectroscopy. *Cancer Res* **2005,** 65, (8), 3030-4.

10.   Somashekar, B. S.; Amin, A. G.; Tripathi, P.; MacKinnon, N.; Rithner, C. D.; Shanley, C. A.; Basaraba, R.; Henao-Tamayo, M.; Kato-Maeda, M.; Ramamoorthy, A.; Orme, I. M.; Ordway, D. J.; Chatterjee, D., Metabolomic signatures in guinea pigs infected with epidemic-associated W-Beijing strains of Mycobacterium tuberculosis. *J Proteome Res* **2012,** 11, (10), 4873-84.

11.   Martineau, E.; Tea, I.; Loaec, G.; Giraudeau, P.; Akoka, S., Strategy for choosing extraction procedures for NMR-based metabolomic analysis of mammalian cells. *Anal Bioanal Chem* **2011,** 401, (7), 2133-2142.

12.   Sellick, C. A.; Hansen, R.; Stephens, G. M.; Goodacre, R.; Dickson, A. J., Metabolite extraction from suspension-cultured mammalian cells for global metabolite profiling. *Nat Protoc* **2011,** 6, (8), 1241-9.

13.   Ackerstaff, E.; Pflug, B. R.; Nelson, J. B.; Bhujwalla, Z. M., Detection of increased choline compounds with proton nuclear magnetic resonance spectroscopy subsequent to malignant transformation of human prostatic epithelial cells. *Cancer Res* **2001,** 61, (9), 3599-3603.

14.   Nicholson, J. K., Metabonomic and integrative systems biology investigations in experimental disease states. *Abstracts of Papers of the American Chemical Society* **2005,** 230, U502-U503.

15.   Gu, H.; Pan, Z.; Xi, B.; Hainline, B. E.; Shanaiah, N.; Asiago, V.; Gowda, G. A.; Raftery, D., 1H NMR metabolomics study of age profiling in children. *NMR Biomed* **2009,** 22, (8), 826-33.

16.   Assfalg, M.; Bertini, I.; Colangiuli, D.; Luchinat, C.; Schafer, H.; Schutz, B.; Spraul, M., Evidence of different metabolic phenotypes in humans. *Proc Natl Acad Sci U S A* **2008,** 105, (5), 1420-4.

17.   Lindon, J. C.; Holmes, E.; Nicholson, J. K., Metabonomics and its role in drug development and disease diagnosis. *Expert Review of Molecular Diagnostics* **2004,** 4, (2), 189-199.

18.   Martin, F.-P. J.; Sprenger, N.; Montoliu, I.; Rezzi, S.; Kochhar, S.; Nicholson, J. K., Dietary Modulation of Gut Functional Ecology Studied by Fecal Metabonomics. *J Proteome Res* **2010,** 9, (10), 5284-5295.

19.   Yap, I. K. S.; Li, J. V.; Saric, J.; Martin, F.-P.; Davies, H.; Wang, Y.; Wilson, I. D.; Nicholson, J. K.; Utzinger, J.; Marchesi, J. R.; Holmes, E., Metabonomic and

microbiological analysis of the dynamic effect of vancomycin-induced gut microbiota modification in the mouse. *J Proteome Res* **2008,** 7, (9), 3718-3728.

20.   Zhao, L.; Nicholson, J. K.; Lu, A.; Wang, Z.; Tang, H.; Holmes, E.; Shen, J.; Zhang, X.; Li, J. V.; Lindon, J. C., Targeting the Human Genome-Microbiome Axis for Drug Discovery: Inspirations from Global Systems Biology and Traditional Chinese Medicine. *J Proteome Res* **2012,** 11, (7), 3509-3519.

21.   Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Research* **2013,** 41, (D1), D801-D807.

22.   Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Wenger, R. K.; Yao, H.; Markley, J. L., BioMagResBank. *Nucleic Acids Research* **2008,** 36, D402-D408.

23.   Napoli, C.; Sperandio, N.; Lawlor, R. T.; Scarpa, A.; Molinari, H.; Assfalg, M., Urine metabolic signature of pancreatic ductal adenocarcinoma by (1)h nuclear magnetic resonance: identification, mapping, and evolution. *J Proteome Res* **2012,** 11, (2), 1274-83.

24.   Holmes, E.; Foxall, P. J. D.; Spraul, M.; Farrant, R. D.; Nicholson, J. K.; Lindon, J. C., 750 MHz H-1 NMR spectroscopy characterisation of the complex metabolic pattern of

urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease. *J Pharm Biomed Anal* **1997,** 15, (11), 1647-1659.

25.   Mackinnon, N.; Somashekar, B. S.; Tripathi, P.; Ge, W.; Rajendiran, T. M.; Chinnaiyan, A. M.; Ramamoorthy, A., MetaboID: A graphical user interface package for assignment of (1)H NMR spectra of bodyfluids and tissues. *J Magn Reson* **2013,** 226, 93-9.

26.   Tessem, M. B.; Swanson, M. G.; Keshari, K. R.; Albers, M. J.; Joun, D.; Tabatabai, Z. L.; Simko, J. P.; Shinohara, K.; Nelson, S. J.; Vigneron, D. B.; Gribbestad, I. S.; Kurhanewicz, J., Evaluation of lactate and alanine as metabolic biomarkers of prostate cancer using 1H HR-MAS spectroscopy of biopsy tissues. *Magn Reson Med* **2008,** 60, (3), 510-6.

27.   MacKinnon, N.; Khan, A. P.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., Androgen receptor activation results in metabolite signatures of an aggressive prostate cancer phenotype: an NMR-based metabonomics study. *Metabolomics* **2012,** 8, (6), 1026-1036.

28.   Tripathi, P.; Kamarajan, P.; Somashekar, B. S.; MacKinnon, N.; Chinnaiyan, A. M.; Kapila, Y. L.; Rajendiran, T. M.; Ramamoorthy, A., Delineating metabolic signatures of head and neck squamous cell carcinoma: phospholipase A2, a potential therapeutic target. *Int J Biochem Cell Biol* **2012,** 44, (11), 1852-61.

29.   Djukovic, D.; Appiah-Amponsah, E.; Shanaiah, N.; Gowda, G. A.; Henry, I.; Everly, M.; Tobias, B.; Raftery, D., Ibuprofen metabolite profiling using a combination of

SPE/column-trapping and HPLC-micro-coil NMR. *J Pharm Biomed Anal* **2008,** 47, (2), 328-34.

30.   Spraul, M.; Freund, A. S.; Nast, R. E.; Withers, R. S.; Maas, W. E.; Corcoran, O., Advancing NMR sensitivity for LC-NMR-MS using a cryoflow probe : application to the analysis of acetaminophen metabolites in urine. *Anal Chem* **2003,** 75, (6), 1546-1551.

31.   Beltran, A.; Suarez, M.; Rodriguez, M. A.; Vinaixa, M.; Samino, S.; Arola, L.; Correig, X.; Yanes, O., Assessment of compatibility between extraction methods for NMR- and LC/MS-based metabolomics. *Anal Chem* **2012,** 84, (14), 5838-44.

32.   MacKinnon, N.; Ge, W.; Khan, A. P.; Somashekar, B. S.; Tripathi, P.; Siddiqui, J.; Wei, J. T.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., Variable reference alignment: an improved peak alignment protocol for NMR spectral data with large intersample variation. *Anal Chem* **2012,** 84, (12), 5372-9.

33.   Coen, M.; Goldfain-Blanc, F.; Rolland-Valognes, G.; Walther, B.; Robertson, D. G.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Pharmacometabonomic investigation of dynamic metabolic phenotypes associated with variability in response to galactosamine hepatotoxicity. *J Proteome Res* **2012,** 11, (4), 2427-40.

34.   Diaz, S. O.; Pinto, J.; Graca, G.; Duarte, I. F.; Barros, A. S.; Galhano, E.; Pita, C.; Almeida Mdo, C.; Goodfellow, B. J.; Carreira, I. M.; Gil, A. M., Metabolic biomarkers of prenatal disorders: an exploratory NMR metabonomics study of second trimester maternal urine and blood plasma. *J Proteome Res* **2011,** 10, (8), 3732-42.

35.   Gronwald, W.; Klein, M. S.; Zeltner, R.; Schulze, B. D.; Reinhold, S. W.; Deutschmann, M.; Immervoll, A. K.; Boger, C. A.; Banas, B.; Eckardt, K. U.; Oefner, P. J., Detection of autosomal dominant polycystic kidney disease by NMR spectroscopic fingerprinting of urine. *Kidney Int* **2011,** 79, (11), 1244-53.

36.   Carrola, J.; Rocha, C. M.; Barros, A. S.; Gil, A. M.; Goodfellow, B. J.; Carreira, I. M.; Bernardo, J.; Gomes, A.; Sousa, V.; Carvalho, L.; Duarte, I. F., Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. *J Proteome Res* **2011,** 10, (1), 221-30.

37.   Pechlivanis, A.; Kostidis, S.; Saraslanidis, P.; Petridou, A.; Tsalis, G.; Mougios, V.; Gika, H. G.; Mikros, E.; Theodoridis, G. A., (1)H NMR-based metabonomic investigation of the effect of two different exercise sessions on the metabolic fingerprint of human urine. *J Proteome Res* **2010,** 9, (12), 6405-16.

38.   Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A.; Zhang, S.; Raftery, D., Chemoselective 15N tag for sensitive and high-resolution nuclear magnetic resonance profiling of the carboxyl-containing metabolome. *Anal Chem* **2009,** 81, (12), 4882-8.

39.   Ye, T.; Zhang, S. C.; Mo, H. P.; Tayyari, F.; Gowda, G. A. N.; Raftery, D., C-13-Formylation for Improved Nuclear Magnetic Resonance Profiling of Amino Metabolites in Biofluids. *Anal Chem* **2010,** 82, (6), 2303-2309.

40.   Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G., Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal Chem* **2006,** 78, (3), 743-7

CHAPTER 4

Prostate cancer diagnosis using NMR-based metabolic profiling of serum

4.1 Abstract

Prostate cancer (PCa) is the most common cancer of aging men and is usually diagnosed at an advanced stage by invasive biopsy tests. To alleviate patients' pain with the biopsy test, we propose the metabolic profiling of blood serum as a relatively non-invasive approach to facilitate the early detection of PCa. This is based on the hypothesis that the expressed metabolic profiles will be different between cancerous and benign states. We performed serum metabolic profiling of 25 benign and 25 PCa samples using ^{1}H nuclear magnetic resonance (NMR) spectroscopy. Both univariate wilcoxon rank sum tests and multivariate partial least squares discriminant analysis (PLS-DA) showed a consistent result that 2-hydroxyvalerate, glycine, fructose, choline and creatine can be used as potential biomarkers that distinguish between the benign and PCa state. This is a pioneer study of the metabolic alterations in serum samples that characterize PCa by NMR metabolic profiling, which can facilitate the development of clinically useful biomarkers for PCa detection.

4.2 Introduction

Prostate cancer (PCa) as the most severe form of prostate related disease, is among the common cancers detected in aging men, with an estimated 238,590 men diagnosed and 29,720 cancer-related deaths in the United States in 2013 according to the PCa statistics of National Cancer Institute.

To diagnose PCa, a screening test is performed which include digital rectal exam (DRE), prostate-specific antigen (PSA) / PSA Density (PSAD) test and transrectal ultrasound (TRUS). If the test results suggest cancer, a prostate biopsy needs to be performed, which is the only way to confirm the presence of PCa. However, the DRE result is subjective to the empirical knowledge of doctors and is only effective for detecting abnormalities close to the rectum, while omitting other areas of the prostate[1]. In addition, PSA has low specificity for PCa detection[2], and other prostate diseases (e.g. BPH) also demonstrate an elevated PSA level in the serum[3]. Even for patients with PSA levels less than the normal threshold value (3-4 ng/mL), the presence of PCa is not uncommon[4]. These controversial test results lead to a higher probability of unnecessary biopsy tests. Therefore, to alleviate the pain and avoid complication caused by the biopsy test, it is imperative that non-invasive biomarkers be identified that will reliably differentiate cancer and normal states of the prostate.

Metabolic profiling techniques have been increasingly applied to characterize the small molecular weight metabolites as the determinant of disease and have laid the framework for the new diagnostic approach of various types of cancer[5-9]. Different metabolite levels reflect the pathophysiological condition of the organism and metabolites that distinguish between cancer and control can be used to trace the cancer metabolic

pathways, providing insight in treatment methods to minimize the dysfunctional enzymatic processes. Metabolic profiling typically relies on gas chromatography (GC) and liquid chromatography (LC), in conjunction with mass spectrometry (MS) and nuclear magnetic resonance (NMR). In particular, MS enjoys high sensitivity (~fmol) and is capable of identifying hundreds of metabolites in a single sample. Various studies have reported the metabolomics approach in the diagnosis of prostate cancer on different sample types using MS. For example, Sreekumar et.al proposed several metabolites (e.g. sarcosine), as metabolite biomarkers for metastatic prostate cancer[10]. Besides, Zhou et.al identified plasma lipids (e.g. phosphatidylethanolamine, ether-linked phosphatidylethanolamine and ether-linked phosphatidylcholine)[11] that can differentiate PCa from control plasma samples by MS-based lipid profiling. Moreover, Zhang et.al reported urinary metabolites (e.g. 2-oxoglutarate and indolylacryloyglycine) as biomarkers for prostate cancer detection[12]. However, these MS-based analysis requires complex sample preparation, quantification cannot be easily applied to clinical tests since it requires the addition of the isotope labeled form of each specific metabolite and offers limited reproducibility. In addition, little is known about the metabolic profiles of blood serum samples that may differentiate PCa and benign states, given that the serum samples are easily accessible at clinics.

As another commonly used analytical techniques for metabolomics study, nuclear magnetic resonance (NMR) spectroscopy has been employed for its quantitative nature and broad coverage of metabolites, including those not amenable to mass spectrometry (e.g. non-ionizable metabolites). NMR measurements exhibit high reproducibility as compared with mass spectrometry, generating data with less variation that could be

attributed to the instrumental variation.   Thus in this study, we investigated NMR-based metabolic profiling of serum samples for the early detection of PCa.   The challenge remains the identification and quantification of a large number of metabolites in the proteinaceous serum samples and determining significant metabolic differentiation between pathophysiological states.   Therefore, serum deproteinization and delipidation were performed in order to obtain accurate quantification of serum metabolites in NMR-based metabolic profiling. Both univariate (e.g. Wilcoxon rank sum tests, ROC analysis) and multivariate statistical analysis (e.g. PCA, PLS-DA) were applied to determine the differential metabolites. This study highlights the effectiveness of NMR-based serum metabolic profiling for the early detection of PCa. The identification of five differential metabolites provides insight for future application in clinical diagnosis. Moreover, further understanding of the biochemical pathways of these metabolites will lead towards effective, personalized therapeutics.

4.3 Materials and Methods

4.3.1 Prostate serum sample collection

25 biopsy positive and 25 biopsy negative serum samples were collected according to the sample collection protocols approved by the University of Michigan Medical School Institutional Review Board for Human Subject Research.

4.3.2 Sample preparation for NMR metabolic profiling

To remove the interference of the broad NMR resonances of macromolecules on the metabolic profiles, serum proteins need to be removed. Therefore in our study, we added the TSP solution (50 μL, 0.38M) as an internal standard into 100 μL of serum and

mixing each serum sample with the cold methanol in 1:1 (v/v) for the protein precipitation. The mixture was left at -20 °C for 30 min. We then centrifuged the sample at 10,000 RPM for 10 min and obtained the supernatant for further metabolite extraction. 150 µL of cold chloroform was added into the supernatant for lipid removal. The sample was then vortexed and centrifuged in order to separate the aqueous and lipid extracts. We dried the aqueous extract using a vacuum concentrator and stored the dried sample at -20 °C before NMR metabolic profiling. Due to the limited sample availability which results in low concentration of metabolites, we used a solid state MAS-NMR instrument equipped with a nanoprobe as the measurement approach for improved S/N. It is more beneficial than a traditional solution state NMR since it employs a sample holder which requires less volume (~40 µL) than it is needed with a regular 5 mm NMR tube for a solution state NMR experiment (~500 µL).

4.3.3 ^{1}H MAS NMR spectroscopy

All metabolic profiling were conducted on a Varian Inova 600 MHz MAS NMR spectrometer equipped with the gHZ nanoprobe that enables the measurement of samples with small solvent volumes. The aqueous extract of each serum sample was reconstituted in 50 µL of phosphate buffer (0.1 M, prepared in D_2O) and 45 µL was used for NMR measurement at 298 K. 1D ^{1}H NMR experiments at 2500 Hz spinning rate and water saturation were performed in which 16,000 data points were collected in 128 scans. We processed the data using ACD/NMR Processor 12.01 (ACD/Labs, Toronto, Canada) in which 2x zero-filling and an exponential window function equivalent to 0.3 Hz line broadening were applied before Fourier transformation. We then phased the spectrum, performed baseline correction and set the reference signal TSP at 0 ppm. The processed

data was exported as ASCII files and subject to further data processing such as normalization and spectral alignment in Matlab (R2010a) as described below.

Excluding TSP, methanol, water and chloroform resonances, total area normalization was performed on each sample to eliminate the systematic variation caused by different sample dilutions. Afterwards, all spectra were aligned based on the vr-icoShift method[13].

4.3.4 Multivariate modeling

To investigate the metabolites differentiating biopsy negative from biopsy positive serum, multivariate analysis was performed on the above processed data. First, the unsupervised principal component analysis (PCA) model was utilized to reduce the dimension of data as well as visualizing the data structure. We performed Monte Carlo cross validation with autoscaling to determine the optimal number of latent variables (LVs) to build the partial least squares discriminant analysis (PLS-DA) classification model. After building the PLS-DA model using the optimal LVs, variable importance in projection (VIP) score of each variable was calculated based on variable weights of LV1 and ranked in descending orders. Variables with VIP score>1 were selected as significant peaks that can differentiate biopsy negative from biopsy positive serum samples. All statistical analysis was performed in Matlab and SPSS software.

4.3.5 Univariate analysis

We performed a two-sided Wilcoxon rank sum test to compare the concentration of each metabolite in the benign and PCa groups to obtain the p-values to test the difference in these two groups. In order to identify altered metabolite signals, we also calculated the z-scores of both sample types by centering each metabolite by the mean

of benign samples and scaled by the standard deviation of benign for that particular metabolite. Moreover, metabolites identified as significant in differentiating the benign and PCa samples were quantified by integration and plotted in a receiver operating characteristic (ROC) curve to visualize the predictive ability of this metabolite at different sensitivity and specificity levels. The area under the curve (AUC) represents the discriminatory ability of this metabolite as the potential biomarker, with values close to 1 as a better classification.

4.4 Results and discussion

Diseases associated with the prostate, including benign prostatic hyperplasia (BPH), prostatitis, and cancer are among the leading diagnosed conditions among men. Current screening techniques for prostate cancer (PCa) are restricted to serum Prostate-Specific Antigen (PSA) levels and digital rectal examination (DRE). Unfortunately, high levels of serum PSA will not uniquely identify prostate cancer, and coupled with the high false-positive PSA testing rates, invasive biopsy procedures are required for disease confirmation[14]. Serum as a commonly used sample type in clinical practice, contains a wide variety of biochemical components which reveals the pathophysiological state of the organisms. Our study aims to generate a biomarker multiplex panel in relatively non-invasive serum samples that will indicate with greater sensitivity and specificity the development of PCa, thus alleviating the dependence on PSA and biopsy tests. We propose the utilization of ^{1}H NMR metabolic profiling approach due to its great potential in reporting on the current metabolic state, where changes at the genetic and protein expression levels on disease onset are amplified downstream in the metabolites. These noninvasive metabolite biomarkers will be useful for the early detection of PCa and

74

provide insight on the treatment methods based on the specific metabolic pathway analysis.

4.4.1 ^{1}H NMR metabolic profiles

In general, NMR-based metabolomics studies on serum samples were carried out using CPMG spectra, which gives information from small molecules by filtering out signals from macromolecules (proteins, lipoproteins, glycoproteins)[15]. However, CPMG spectra don't provide quantitative accuracy[16]. In order to get quantitative information of serum metabolites, in the present study, we extracted aqueous metabolites from serum samples and ^{1}H NMR metabolic profiling was carried out on twenty-five benign and twenty-five PCa serum samples. Two representative ^{1}H NMR spectra obtained from aqueous extracts of a benign and a PCa serum sample were shown in Figure 4.1. The chemical shift assignment of metabolites was conducted based on literature reports, metabolite data bases (e.g. HMDB, BMRB), and confirmed by resonance correlations in the acquired 2D [1]H-^{1}H TOCSY spectrum. More than twenty-two metabolites were identified both in benign and PCa serum samples (Figure 4.1) including amino acids, sugars, choline containing compounds and other organic acids. To detect overall changes in the ^{1}H signal intensities of benign and PCa samples, mean and standard deviation was calculated for benign group for each binned chemical shift variable. Furthermore, each benign and PCa sample was centered by the benign mean and scaled by the benign standard deviation. Z-scores plot of the ^{1}H signals deviation from benign state for both groups was shown in Figure 4.2, with several metabolite variables in the PCa samples showing altered metabolic profiles compared with benign samples.
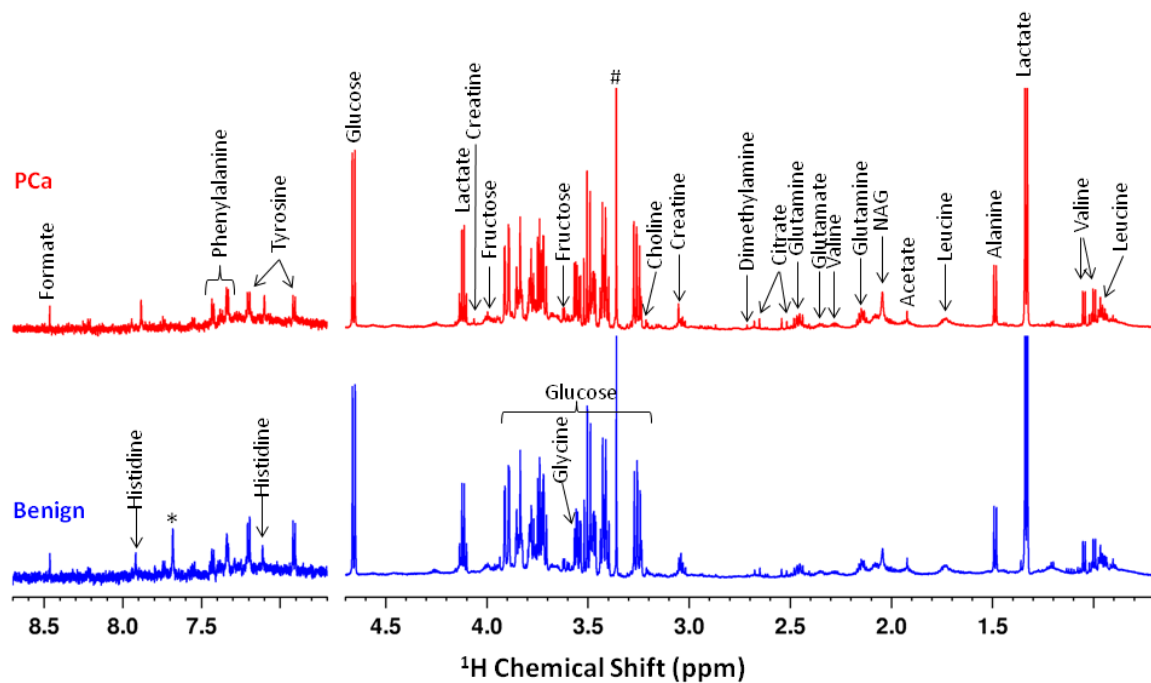
Figure 4.1 Representative 1D ^{1}H NMR spectra of benign and PCa serum metabolite extracts.
The pound sign represents methanol peak whereas the star represents chloroform peak.
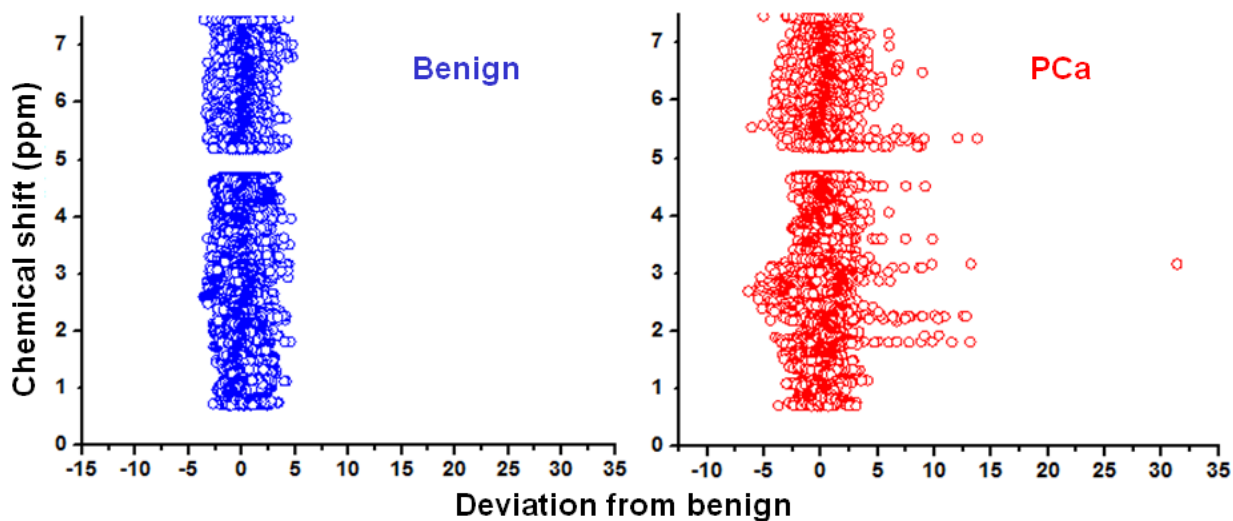
Figure 4.2 z-scores plot for each metabolite variables of the benign and PCa samples.
Metabolites of PCa samples showed the metabolic alteration from the benign metabolic profiles.

76

4.4.2 Statistical analysis

      To determine the differential metabolites between benign and PCa samples, unsupervised PCA was performed. Before this multivariate statistical analysis, unwanted chemical shift regions were removed from the ^{1}H spectra and the resultant data was normalized to the total intensity. Since the PCA score plot didn't show clustering of the two groups, a supervised PLS-DA was performed to maximize the covariance of metabolite signals with the sample class information (e.g. benign and PCa). The optimal number of components, namely the number of latent variables (LVs) to build the PLSDA model was selected based on the Monte Carlo cross validation (MCCV) approach. In contrast to the commonly used leave one out cross validation, MCCV leaves out a large part of data (e.g. leave out 30% of the data randomly in each round of MCCV analysis) as the test set to determine the prediction error rate. The number of LVs yielding the smallest error rate was retained as the optimal number of LVs for PLSDA model development. In our case, 500 MCCV was performed on the full data set, resulting in 2 LVs as the optimal parameter to build the PLSDA model.   The scores plot and LV 1 loadings plot of PLSDA was shown in Figure 4.3. Despite the overlap of a few benign with PCa samples, separation of the two groups was observed in the scores plot, indicating the difference of the metabolic profiles of these samples. The loadings plot of LV1 revealed increased level of branched chain amino acids, lactate, alanine, glutamine, glycine, choline and tyrosine, and decreased levels of glucose. Signals were ranked according to their contribution to the PLSDA model using variable importance to projection (VIP) scores as shown in Table 4.1.   And the top five VIP ranked metabolites were 2-hydroxyvalerate, glycine, fructose, choline and creatine.
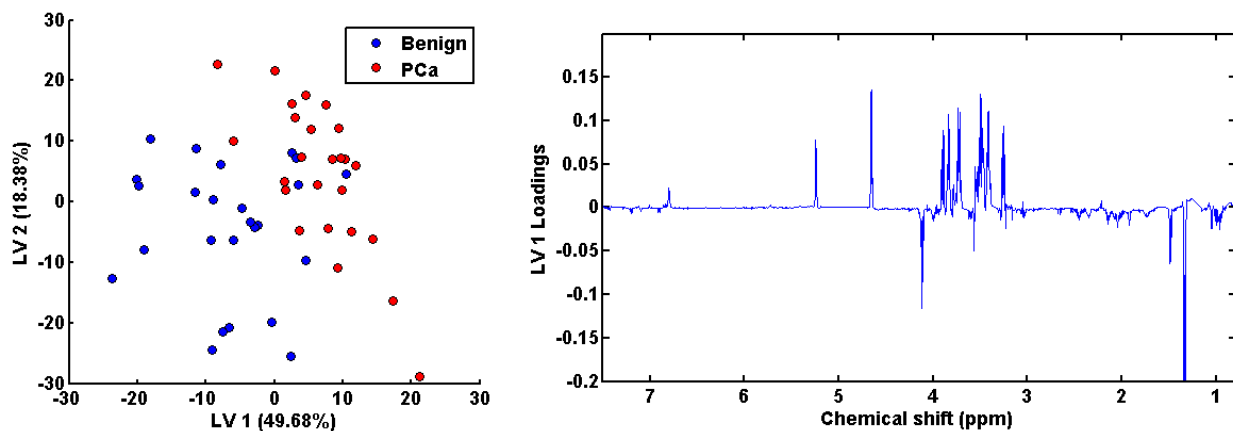
Figure 4.3 Scores and loadings plot based on PLSDA model performance.

In addition to the multivariate statistical analysis, we also performed univariate statistical analysis (Wilcoxon rank sum test) on binned NMR data (1732 bins; bin size 0.004 ppm), in which binned dataset with $P < 0.05$ was considered statistically significant. Not surprisingly, the observed statistically significant binned datasets is corresponded to the high VIP ranked metabolites (Table 4.1). Further, mean and standard deviation of peak intensities for 2-hydroxyvalerate, glycine, fructose, choline and creatine were calculated using the chemical shift range of 1.314 to 1.354 ppm, 3.564 to 3.573 ppm, 3.589 to 3.609 ppm, respectively. The levels of these five metabolites were shown in Figure 4.4. Among these five metabolites, only 2-hydroxyvalerate was significantly decreased and other four metabolites were significantly increased in PCa compared to benign serum. To address the diagnostic potential of these differential metabolites, ROC analysis was performed individually for each metabolite and the corresponding AUC values were included in Table 4.1. The diagnostic performances of these key differential metabolites are enhanced comparing with the PSA based test to discriminate PCa from no cancer with an AUC of 0.678[17].
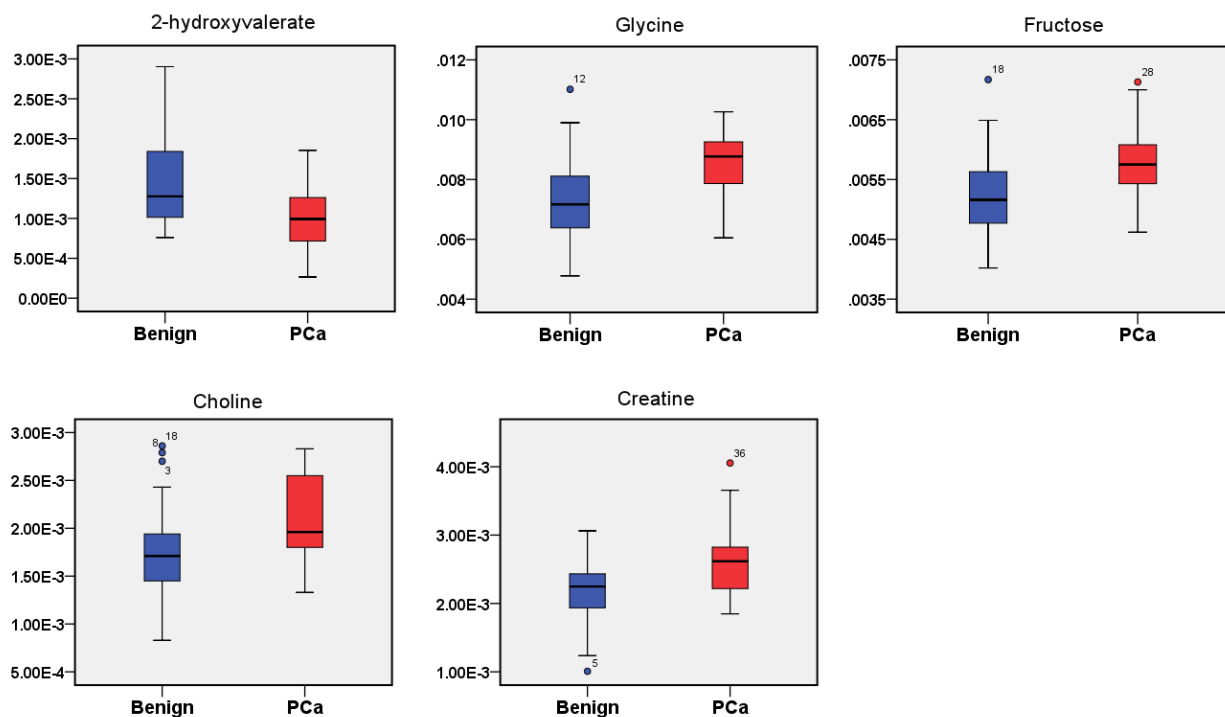
Figure 4.4 Metabolite levels of the potential biomarkers in distinguishing between benign and PCa serum samples.
Integrals were measured after total area normalization.

Table 4.1 Collection of potential metabolite biomarkers of PCa.
Signals ranked according to their VIP scores calculated from PLSDA model of LV1, p values resulting from Wilcoxon rank sum tests and AUC values of each metabolite.

Metabolite	VIP rank	p value	AUC
2-hydroxyvalerate	1	0.007	0.725
Glycine	2	0.002	0.762
Fructose	3	0.012	0.755
Choline	4	0.008	0.784
Creatine	5	0.02	0.694

4.4.3 Metabolic alterations and signal transduction pathways

Increased choline level is observed in PCa serum, which is associated with the breakdown of membrane phospholipid during malignant transformation of human prostatic epithelial cells[18]. The alteration of choline is related to increased choline kinase activity that catalyzes choline to form phosphorylcholine with the consumption of ATP as well as the increased phospholipase activity that breaks down phosphotidylcholine to reform choline[19]. These choline signaling pathways provide a rationale for a variety of applications. For example, the inhibition of phospholipase (PLCy1) showed a 50% reduction in phosphocholine levels and can be used as a therapeutic target to block the cancer cell invasion and metastasis[20]. Moreover, ^{11}C labeled choline has been applied as a radiopharmaceutical for positron emission tomography (PET) imaging to detect the local recurrence of prostate tumor[21].

Elevated levels of glycine was found in PCa serum, which was also reported in colorectal cancer[22], head and neck cancer studies by NMR[23]. The reliance on glycine for the proliferation of cancer cells was studied by means of gene expression of metabolic enzymes[24], which reveals the enhanced expression of glycine biosynthesis enzymes in cancer cells. Glycine is then incorporated into purine biosynthesis in cancer cell proliferation. Our observation of increased creatine level in PCa serum may further indicate the alteration in energy metabolism of PCa patients. 2-hydroxyvalerate is associated with lactic acidosis and is down-regulated in our PCa serum analysis.

4.5 Conclusion

We presented a novel metabolic study of serum using NMR-based profiling and identified several serum metabolites that can distinguish between benign and PCa patients. Further validation of these potential metabolite biomarkers will be assessed over a large group of independent serum bio-specimens that are linked to clinical/pathology data. The altered metabolite level of these potential biomarkers provides great insight into the diagnosis of PCa as well as understanding the mechanism of PCa development through the specific biochemical pathway analysis. This noninvasive serum metabolite biomarker panel may be potentially applied in screening of prostate cancer rather than limiting on the analysis of only the protein level (e.g. PSA) in serum.

4.6 References

1.   Philip, J.; Dutta Roy, S.; Ballal, M.; Foster, C. S.; Javle, P., Is a digital rectal examination necessary in the diagnosis and clinical staging of early prostate cancer? *BJU Int* **2005,** 95, (7), 969-71.

2.   Djavan, B.; Zlotta, A.; Kratzik, C.; Remzi, M.; Seitz, C.; Schulman, C. C.; Marberger, M., PSA, PSA density, PSA density of transition zone, free/total PSA ratio, and PSA velocity for early detection of prostate cancer in men with serum PSA 2.5 to 4.0 ng/mL. *Urology* **1999,** 54, (3), 517-22.

3.   Oesterling, J. E., PROSTATE-SPECIFIC ANTIGEN - ITS ROLE IN THE DIAGNOSIS AND STAGING OF PROSTATE-CANCER. *Cancer* **1995,** 75, (7), 1795-1804.

4.   Thompson, I. M.; Pauler, D. K.; Goodman, P. J.; Tangen, C. M.; Lucia, M. S.; Parnes, H. L.; Minasian, L. M.; Ford, L. G.; Lippman, S. M.; Crawford, E. D.; Crowley, J. J.; Coltman, C. A., Jr., Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* **2004,** 350, (22), 2239-46.

5.   Zhang, X.; Xu, L.; Shen, J.; Cao, B.; Cheng, T.; Zhao, T.; Liu, X.; Zhang, H., Metabolic signatures of esophageal cancer: NMR-based metabolomics and UHPLC-based focused metabolomics of blood serum. *Biochim Biophys Acta* **2013,** 1832, (8), 1207-16.

6.   Zhang, A.; Sun, H.; Yan, G.; Han, Y.; Ye, Y.; Wang, X., Urinary metabolic profiling identifies a key role for glycocholic acid in human liver cancer by ultra-performance liquid-chromatography coupled with high-definition mass spectrometry. *Clin Chim Acta* **2013,** 418, 86-90.

7.   Wang, H.; Wang, L.; Zhang, H.; Deng, P.; Chen, J.; Zhou, B.; Hu, J.; Zou, J.; Lu, W.; Xiang, P.; Wu, T.; Shao, X.; Li, Y.; Zhou, Z.; Zhao, Y. L., 1H NMR-based metabolic profiling of human rectal cancer tissue. *Mol Cancer* **2013,** 12, (1), 121.

8.   Tripathi, P.; Kamarajan, P.; Somashekar, B. S.; MacKinnon, N.; Chinnaiyan, A. M.; Kapila, Y. L.; Rajendiran, T. M.; Ramamoorthy, A., Delineating metabolic signatures of head and neck squamous cell carcinoma: phospholipase A2, a potential therapeutic target. *Int J Biochem Cell Biol* **2012,** 44, (11), 1852-61.

9.   Napoli, C.; Sperandio, N.; Lawlor, R. T.; Scarpa, A.; Molinari, H.; Assfalg, M., Urine metabolic signature of pancreatic ductal adenocarcinoma by (1)h nuclear magnetic

resonance: identification, mapping, and evolution. *J Proteome Res* **2012,** 11, (2), 1274-83.

10.   Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M., Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009,** 457, (7231), 910-4.

11.   Zhou, X.; Mao, J.; Ai, J.; Deng, Y.; Roth, M. R.; Pound, C.; Henegar, J.; Welti, R.; Bigler, S. A., Identification of plasma lipid biomarkers for prostate cancer by lipidomics and bioinformatics. *PLoS One* **2012,** 7, (11), e48889.

12.   Zhang, T.; Watson, D. G.; Wang, L.; Abbas, M.; Murdoch, L.; Bashford, L.; Ahmad, I.; Lam, N. Y.; Ng, A. C.; Leung, H. Y., Application of Holistic Liquid Chromatography-High Resolution Mass Spectrometry Based Urinary Metabolomics for Prostate Cancer Detection and Biomarker Discovery. *PLoS One* **2013,** 8, (6), e65880.

13.   MacKinnon, N.; Ge, W.; Khan, A. P.; Somashekar, B. S.; Tripathi, P.; Siddiqui, J.; Wei, J. T.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., Variable reference alignment: an improved peak alignment protocol for NMR spectral data with large intersample variation. *Anal Chem* **2012,** 84, (12), 5372-9.

14.   Smith, D. S.; Humphrey, P. A.; Catalona, W. J., The early detection of prostate carcinoma with prostate specific antigen: the Washington University experience. *Cancer* **1997,** 80, (9), 1852-6.

15.   Meiboom, S.; Gill, D., MODIFIED SPIN-ECHO METHOD FOR MEASURING NUCLEAR RELAXATION TIMES. *Review of Scientific Instruments* **1958,** 29, (8), 688-691.

16.   Song, Y. Q., Categories of coherence pathways for the CPMG sequence. *Journal of Magnetic Resonance* **2002,** 157, (1), 82-91.

17.   Thompson, I. M.; Ankerst, D. P.; Chi, C.; Lucia, M. S.; Goodman, P. J.; Crowley, J. J.; Parnes, H. L.; Coltman, C. A., Jr., Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. *JAMA* **2005,** 294, (1), 66-70.

18.   Ackerstaff, E.; Pflug, B. R.; Nelson, J. B.; Bhujwalla, Z. M., Detection of increased choline compounds with proton nuclear magnetic resonance spectroscopy subsequent to malignant transformation of human prostatic epithelial cells. *Cancer Res* **2001,** 61, (9), 3599-3603.

19.   Katz-Brull, R.; Seger, D.; Rivenson-Segal, D.; Rushkin, E.; Degani, H., Metabolic markers of breast cancer: enhanced choline metabolism and reduced choline-ether-phospholipid synthesis. *Cancer Res* **2002,** 62, (7), 1966-70.

20.   Beloueche-Babari, M.; Peak, J. C.; Jackson, L. E.; Tiet, M. Y.; Leach, M. O.; Eccles, S. A., Changes in choline metabolism as potential biomarkers of phospholipase C{gamma}1 inhibition in human prostate cancer cells. *Mol Cancer Ther* **2009,** 8, (5), 1305-11.

21.   Jadvar, H., Molecular imaging of prostate cancer with PET. *J Nucl Med* **2013,** 54, (10), 1685-8.

22.   Chan, E. C.; Koh, P. K.; Mal, M.; Cheah, P. Y.; Eu, K. W.; Backshall, A.; Cavill, R.; Nicholson, J. K.; Keun, H. C., Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J Proteome Res* **2009,** 8, (1), 352-61.

23.   Somashekar, B. S.; Kamarajan, P.; Danciu, T.; Kapila, Y. L.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., Magic angle spinning NMR-based metabolic profiling of head and neck squamous cell carcinoma tissues. *J Proteome Res* **2011,** 10, (11), 5232-41.

24.   Jain, M.; Nilsson, R.; Sharma, S.; Madhusudhan, N.; Kitami, T.; Souza, A. L.; Kafri, R.; Kirschner, M. W.; Clish, C. B.; Mootha, V. K., Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* **2012,** 336, (6084), 1040-4.

CHAPTER 5

NMR-based metabolomic profiling of urine nominates a signal biomarker panel for prostate cancer detection

5.1 Abstract

Detection of prostate cancer and distinguishing indolent versus aggressive forms of the disease is a critical clinical challenge. The current clinical test is circulating PSA levels, which faces particular challenges in cancer diagnosis in the range of 4 – 10 ng/mL. Thus, a concerted effort towards building a noninvasive biomarker panel has developed. In this report, the hypothesis that 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. ^{1}H NMR spectra of urine samples from 317 individuals (111 biopsy-negative, 206 biopsy-positive) were analyzed. A double cross-validation PLS-DA 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 AUC value of 0.94 after logistic regression analysis, while a 'master list' of 18 variables produced an ROCAUC of 0.80. As proof of principle, this study demonstrates the utility of NMR-based metabolomic profiling of urine biospecimens in the nomination of prostate cancer specific biomarker signals and suggests that further investigation is certainly warranted.

5.2 Introduction

Prostate cancer (PCa) is a disease that is estimated to be diagnosed in 238,590 American men in 2013. 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[1, 2, 3]. First is the low specificity of PCa, with multiple conditions associated with prostate swelling also resulting in elevated PSA levels. Second is the sensitivity, with a clinically challenging range of 4-10 ng/mL 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[4]. Thus, there is a critical requirement for clinically relevant screening tests 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[5, 6, 7, 8]. 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[9]. Further examples of candidate

protein markers include α-methylacyl-CoA racemase[10] and AGR2[11]. Several peptides have also been nominated in urine as potential biomarkers of PCa, although care must be taken in sample collection and preparation[12]. Several genes have been identified as PCa specific biomarkers, including PCA3 and the TMPRSS2:ERG gene fusion. These have been detected in both tissue[13] and urine[14, 15, 16], and are thus attractive from the standpoint of development of noninvasive PCa biomarker panels. In addition, DNA methylation[17] and single nucleotide polymorphisms[18] 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[19]. 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[20, 21, 22, 23].

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[24]. Mass spectrometry 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[25, 26, 27], and thus

this report addresses the hypothesis that urinary ^{1}H NMR profiles of biopsy-negative individuals will differ from biopsy-positive individuals, and the differential signals may be assembled into a biomarker panel for PCa detection.

5.3 Materials/Methods

5.3.1 Materials

3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (TSP), D_2O, and potassium phosphate (monobasic, dibasic) were obtained from Sigma/Aldrich (Milwaukee, USA) and used as received. All water was of MilliQ purity.

5.3.2 Urine specimens & sample treatment

Samples were collected post-digital rectal exam (DRE). Samples were centrifuged immediately upon receipt, and the supernatant and sediments were stored separately. All NMR experiments were performed on urine supernatants. Urine samples were stored at -80 °C until used. NMR sample preparation involved aliquoting 450 µL of neat urine with 50 µL of phosphate buffer (pH 7.2) containing TSP as a chemical shift and concentration reference ([phosphate] $_f$ = 100 mM, [TSP]$_f$ = 3.7 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.

5.3.3 NMR Spectroscopy

All NMR measurements were performed on a Bruker AVANCE™ 900 MHz (Bruker Biospin, Germany) with a TCI cryoprobe operating at a ^{1}H frequency of 889.79 MHz at 298 K. Each 1D ^{1}H spectrum was collected using a 30° flip angle, 128 scans containing

32K data points, with a spectral width of 16 ppm and recycle delay of 2 s. WATERGATE water suppression was utilized.

The raw FID data were imported in ACD/Labs NMR Processor, and were subjected to zero filling to 64K 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.

Spectra were normalized using the Probabilistic Quotient Normalization method, which has been shown to effectively reduce dilution variation[28]. The reference spectrum for normalization was selected according to a closeness index[29], 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[30]. 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: each variable had the mean subtracted and scaled to unit variance[31], and 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.

5.3.4 Statistical Analyses

5.3.4.1 Model development

A double cross validation (2CV) PLS-DA method[32] was used in order to develop models capable of classifying biopsy-negative versus biopsy-positive individuals. Briefly, the full dataset was first randomly divided into a calibration and test set. The calibration set is utilized to develop a PLS-DA model according to 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 Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) curve. Reference distributions for the null hypothesis of the three parameters were generated by performing 2CV on the full dataset after class label permutation. Significant values are expected only when using the correct 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 double cross validation procedure was then repeated 1000 times using the correct class labels, resulting in a collection of 1000 different, but related models. Both the distribution over all 1000 models and mean ± standard deviation of the three performance parameters are reported.

5.3.4.2 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 variable importance to projection (VIP) score of each variable for all models calculated[33]. 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 competitive adaptive reweighted sampling (CARS), described by Li and coworkers[34]. 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 the 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.

5.4 Results and Discussion

Prostate cancer diagnosis continues to be a difficult clinical challenge, and new diagnostic techniques based on biomarker detection in non-invasive samples are under active research. Urine is perhaps is the most attractive noninvasive biospecimen in terms of simplicity of collection, and several reports have discussed potential prostate cancer biomarkers detectable in urine supernatant or sediment[19, 35]. To date, metabolite biomarkers have been investigated by mass spectrometry, with a noticeable lack of NMR based metabolomic investigations for urinary prostate cancer detection. Representative ^{1}H NMR spectra for biopsy-negative and biopsy-positive individuals are given in Figure 5.1.
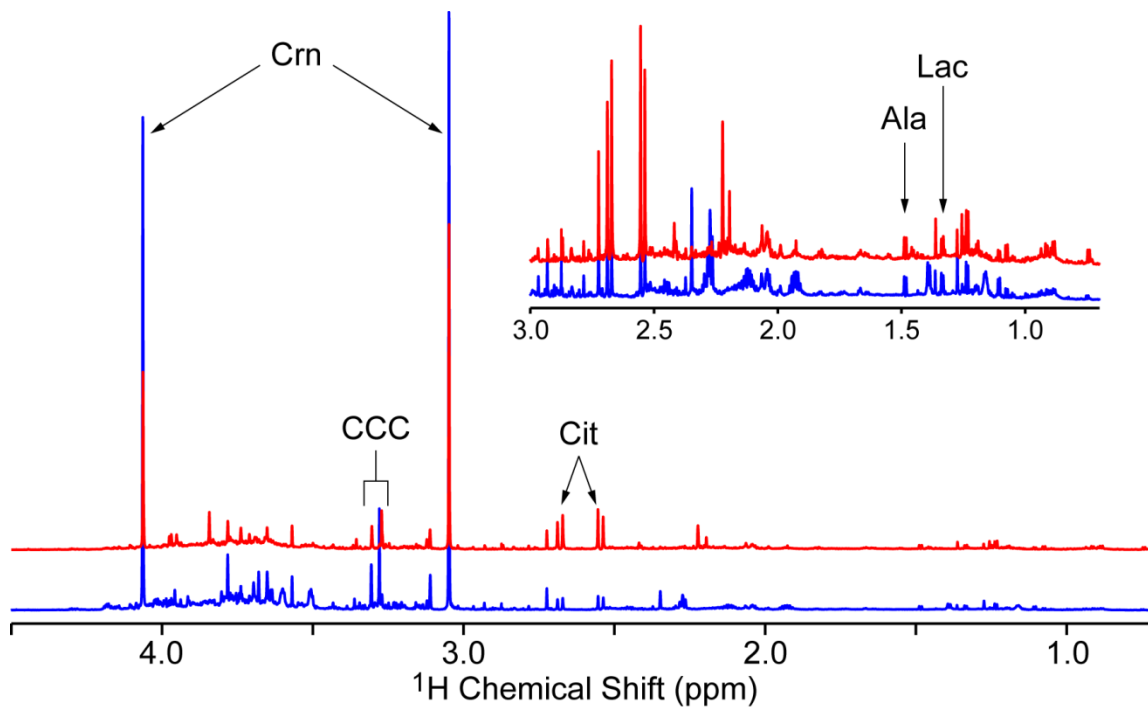
Figure 5.1 900 MHz 1H NMR spectra of a selected biopsy-negative (blue) and biopsy-positive (red) urine supernatant sample.
Spectra were referenced to TSP, PQN normalized, and aligned. Inset: expansion of chemical shift region from 0.7 to 3.0 ppm, with vertical scale increased by a factor of 10. Selected assignments of common PCa biomarkers are given: Crn – creatinine; CCC – choline containing compounds region; Cit – citrate; Ala – alanine; Lac – lactate.

Urine is a particularly challenging biospecimen to work with in terms of biomarker discovery since there is large compositional variability depending on diet, time of collection, gender, and medication taken by the patient[36, 37]. 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). Thus a PLS-DA approach was applied to the data.

5.4.1 Auto-scaled data

As argued by Westerhuis and coworkers[32], 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, AUC values, and Q^2 values after permutation of class labels are summarized in Figure 5.2 and Table 5.1. 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 ± 2 and 0.49 ± 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 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.
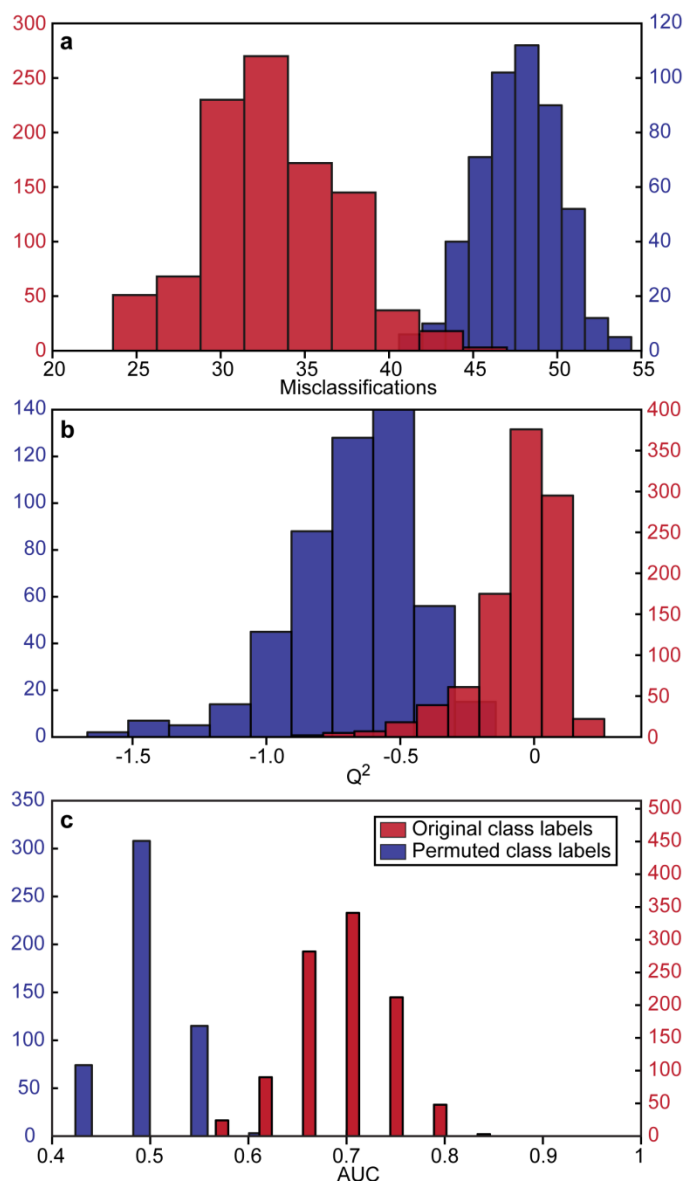
Figure 5.2 Distributions of the number of misclassifications, Q^2, and AUC after 2CV-PLS-DA modeling of the full NMR data set.
The data was auto-scaled prior to modeling. 1000 models were generated using the correct class labels (red), and 500 models using permuted class labels were generated (blue). 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 is summarized in Table 5.1.

Table 5.1 Summary of model performance parameters for 2CV-PLS-DA modeling based on NMR data that was auto-scaled.
Using the original labels, 1000 models were generated while 500 models were generated using the permuted labels. All values are reported as mean ± standard deviation. In the case of permuted data, mean parameter values were first calculated over the 20 models per label permutation followed by mean ± standard deviation calculation over all 500 permutations.

	Data set	Misclassification	Q^2	AUC
Original Labels	Full	33 ± 4	-0.06 ± 0.16	0.69 ± 0.05
	CARS	20 ± 4	0.35 ± 0.08	0.86 ± 0.03
	Reg Coeff	28 ± 4	-0.01 ± 0.14	0.76 ± 0.04
	VIP	28 ± 4	0.01 ± 0.14	0.76 ± 0.04
Permuted Labels	Full	48 ± 2	-0.68 ± 0.24	0.49 ± 0.03
	CARS	48 ± 3	-1.6 ± 0.4	0.49 ± 0.04
	Reg Coeff	47 ± 3	-2.8 ± 0.9	0.49 ± 0.04
	VIP	47 ± 3	-2.9 ± 0.8	0.49 ± 0.01

Application of 2CV PLS-DA resulted in a collection of 1000 models, each built using a randomly selected subset of samples from the full data set. Thus, while each model is different, all models are related. 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[32]. 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[34, 38], and without a clear consensus on a particular strategy, three methods were explored. The VIP score is a

value based on PLS loadings and thus reports on the importance of each variable in contributing to the overall model[33]. Since the average VIP score is 1, variables whose value is greater than one 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 discriminant analysis. Finally, a method recently proposed by Li and coworkers termed CARS 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[34]. 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 5.2). 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. Selecting these variable subsets, logistic regression analysis was performed and ROC curves generated, resulting in AUC values of 0.83 and 0.94 using all regression

coefficient and CARS variables, respectively (Figure 5.3). Additionally, 2CV PLS-DA with permutation analysis was performed utilizing the regression coefficient and CARS variable subsets, with the results presented in Figure 5.4 and summarized in Table 5.1. Clearly, the predictive power of the models was enhanced when selecting variable subsets, with particular improvement using the CARS variable subset.

Table 5.2 The number of variables selected for each data treatment procedure and variable selection protocol.
Values in parentheses are the total number of unique variables identified over the 1000 models generated, while the second value indicates the number of variables that appeared in at least 50% of the models generated.

Data set	VIP	Regression Coefficient	CARS
Autoscaled	21 (723)	21 (692)	29 (1891)
Scaled wrt biopsy-negative	34 (169)	31 (342)	9 (1773)
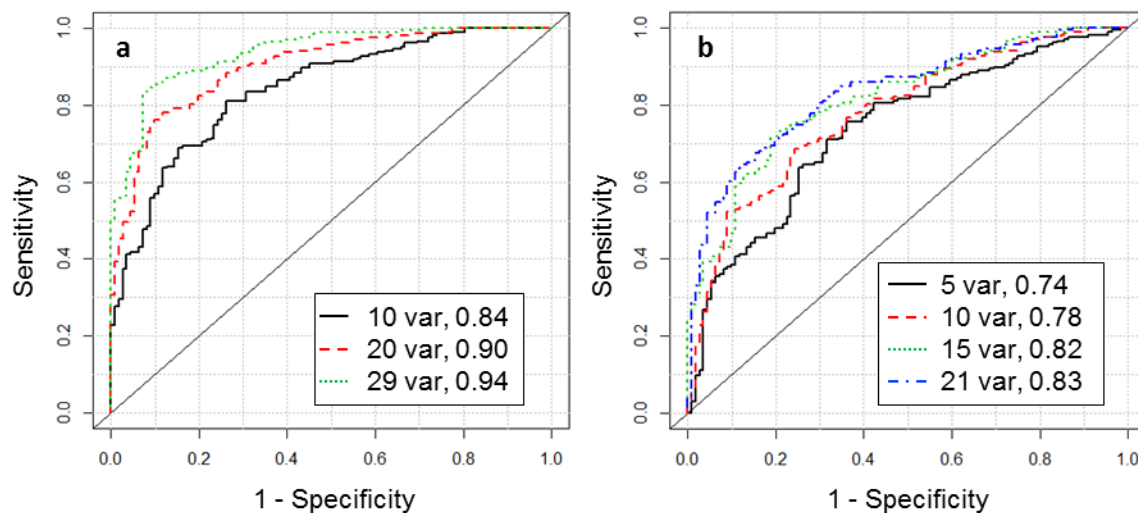
Figure 5.3 Logistic regression analysis performed on the auto-scaled NMR data.
Taking(a) CARS variables and (b) regression coefficient ranked variable subsets identified from 2CV-PLS-DA analysis. ROC curves were generated using the first 10, 20, and 29 variables of the CARS subset, and the first 5, 10, 15, and 21 variables of the regression coefficient ranked subset. The corresponding AUC values for each ROC curve are given in the legend.
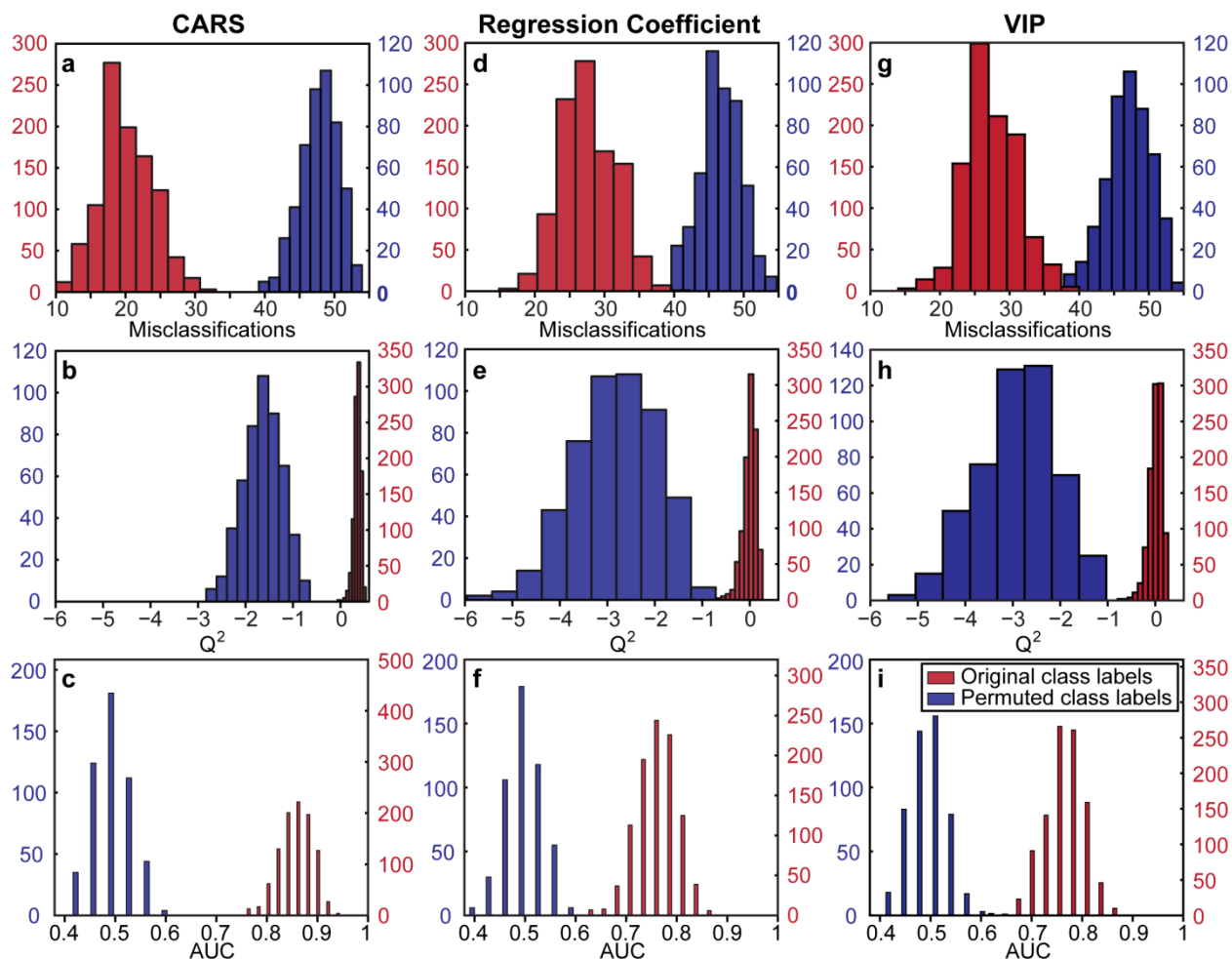
Figure 5.4 Distributions of the number of misclassifications, Q^2, and AUC after 2CV-PLS-DA modeling using variable subsets.
The NMR data was auto-scaled prior to modeling. The subsets are the CARS variable subset (**a-c**), regression coefficient ranked subset (**d-f**), and the VIP ranked subset (**g-i**). 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 is summarized in Table 5.3, and the number of variables included in each model is given in Table 5.2.

5.4.2 Variable changes relative to biopsy-negative data

The effect of centering and scaling all variable with respect to the mean and standard deviation of the normal samples (i.e. biopsy-negative) is a common approach to gain insight into 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 5.3. Utilizing the correct class labels, the

101

predictive performance was observed to slightly increase when compared to permutation test. 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 5.2). 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 5.3). 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 5.4). This is particularly true for the VIP variable subset which had the lowest modeling performance.

Table 5.3 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.
Using the original labels, 1000 models were generated while 500 models were generated using the permuted labels. All values are reported as mean ± standard deviation. In the case of permuted data, mean parameter values were first calculated over the 20 models per label permutation followed by mean ± standard deviation calculation over all 500 permutations.

	Data set	Misclassification	Q^2	AUC
Original Labels	Full	32 ± 4	0.01 ± 0.11	0.74 ± 0.05
	CARS	27 ± 4	-0.03 ± 0.13	0.74 ± 0.04
	Reg Coeff	27 ± 4	0.06 ± 0.11	0.80 ± 0.04
	VIP	41 ± 5	-0.4 ± 0.4	0.71 ± 0.04
Permuted Labels	Full	47 ± 5	-4 ± 6	0.49 ± 0.04
	CARS	48 ± 3	-5 ± 3	0.49 ± 0.03
	Reg Coeff	49 ± 3	-3 ± 1	0.49 ± 0.03
	VIP	46 ± 6	-6 ± 3	0.49 ± 0.03

Table 5.4 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 set	Biopsy-Negative (%)	Biopsy-Positive (%)	Overall (%)
Autoscaled	Full	69 ± 34	63 ± 39	65 ± 37
	CARS	86 ± 27	75 ± 37	79 ± 34
	Reg Coeff	69 ± 36	72 ± 40	71 ± 38
	VIP	70 ± 37	72 ± 40	71 ± 39
Biopsy-negative centered & scaled	Full	80 ± 32	59 ± 41	66 ± 39
	CARS	74 ± 41	71 ± 41	72 ± 41
	Reg Coeff	88 ± 28	64 ± 45	72 ± 41
	VIP	87 ± 27	41 ± 43	57 ± 44

5.4.3 Variable subset comparisons – univariate statistics & 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 negative groups. Wilcoxon rank-sum tests were done for each variable and the appropriate data scaling procedure. The number of variables deemed significant (i.e. $p < 0.05$) for each model are summarized in Figure 5.5. 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.
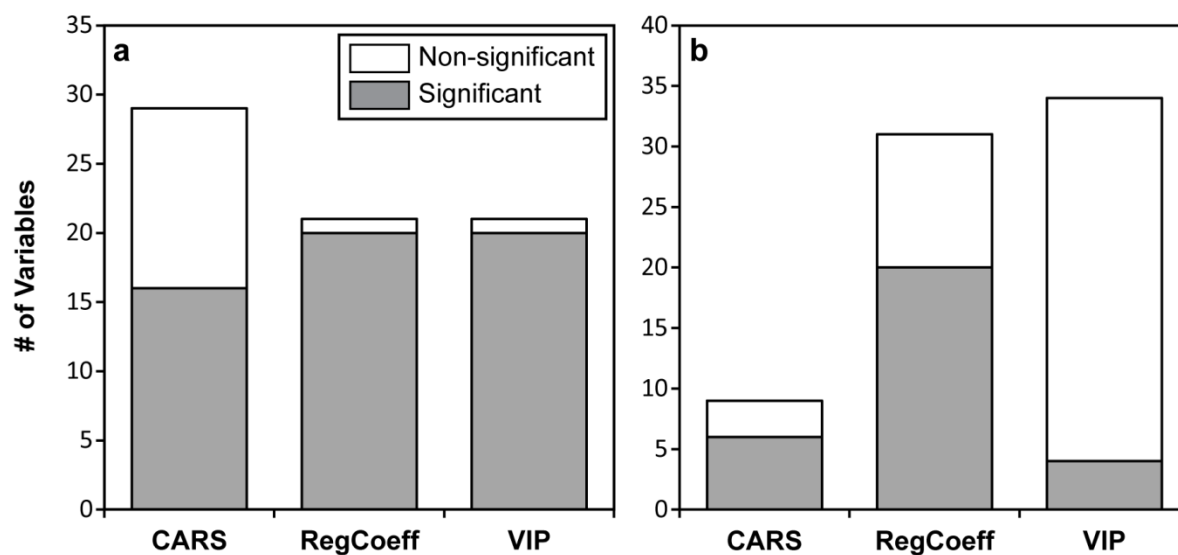
Figure 5.5 Bar plots representing the number of significant and non-significant variables identified in each variable subset using univariate statistics (Wilcoxon rank-sum test).
(a) NMR data was auto-scaled, (b) NMR data was centered and scaled relative to the biopsy-negative samples. All *p*-values 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.

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 datasets were subjected to a STOCSY analysis and the correlation between each signal within each subset of variables was examined[39]. 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[40] (Figure 5.6). However, the VIP data set subjected to normalization and scaling relative to biopsy-negative samples strongly deviates from this trend. A set of strong correlations was observed, with a high percentage possessing r > 0.9. The correlated variables primarily involved signals in the range of 3.2 – 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.
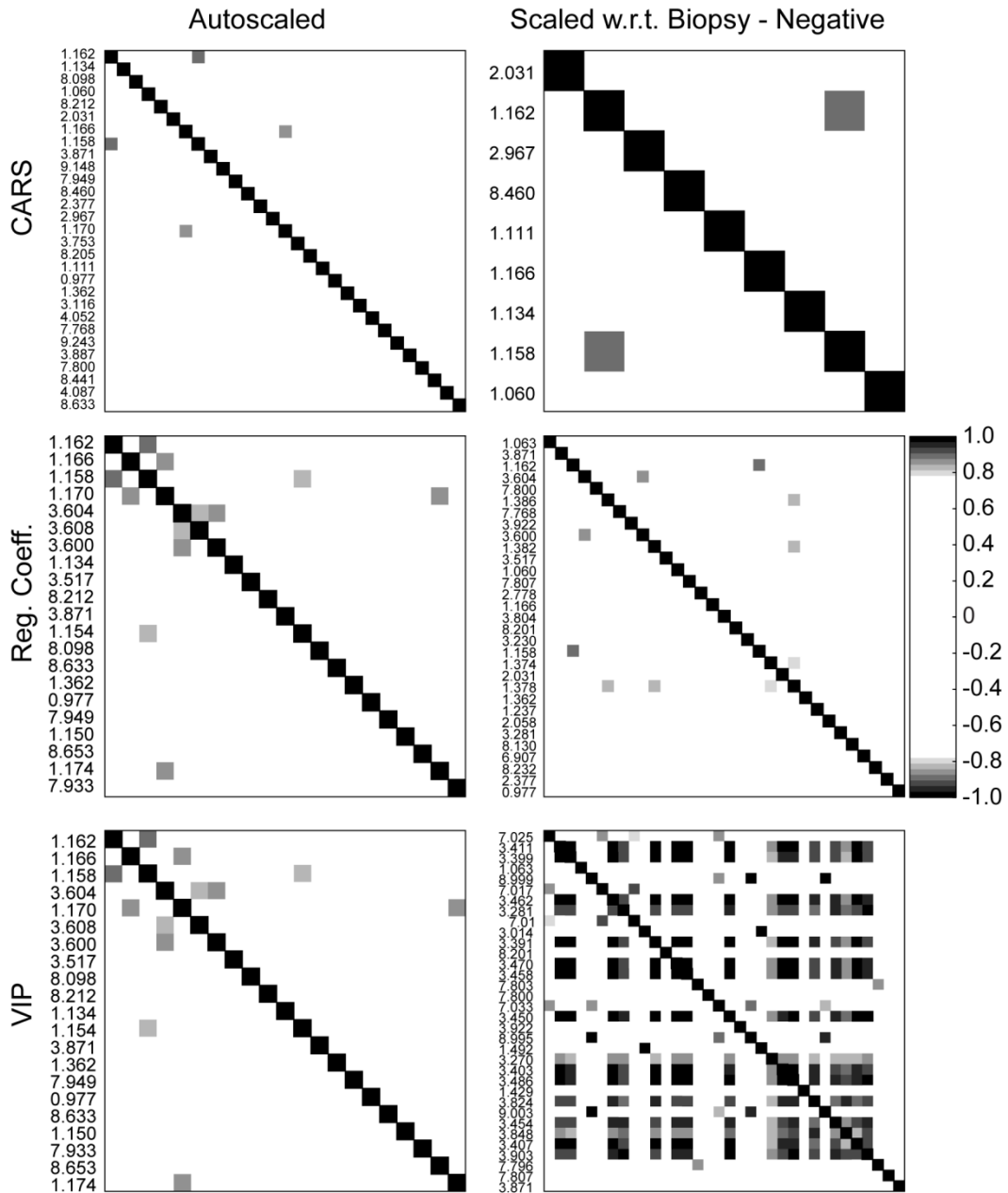
Figure 5.6 Correlations between variables within each subset for each data normalization/scaling procedure.

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 the 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.

107

5.4.4 An optimal biomarker panel for PCa?

In total, six 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 six data sets. Boxplots of a selection of these variables are given in Figure 5.7, where biopsy-positive cases are seen to be both decreased and increased 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 potential biomarkers (Table 5.5). The modeling procedure was repeated with this master list and the performance characteristics are summarized in Table 5.6. 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.
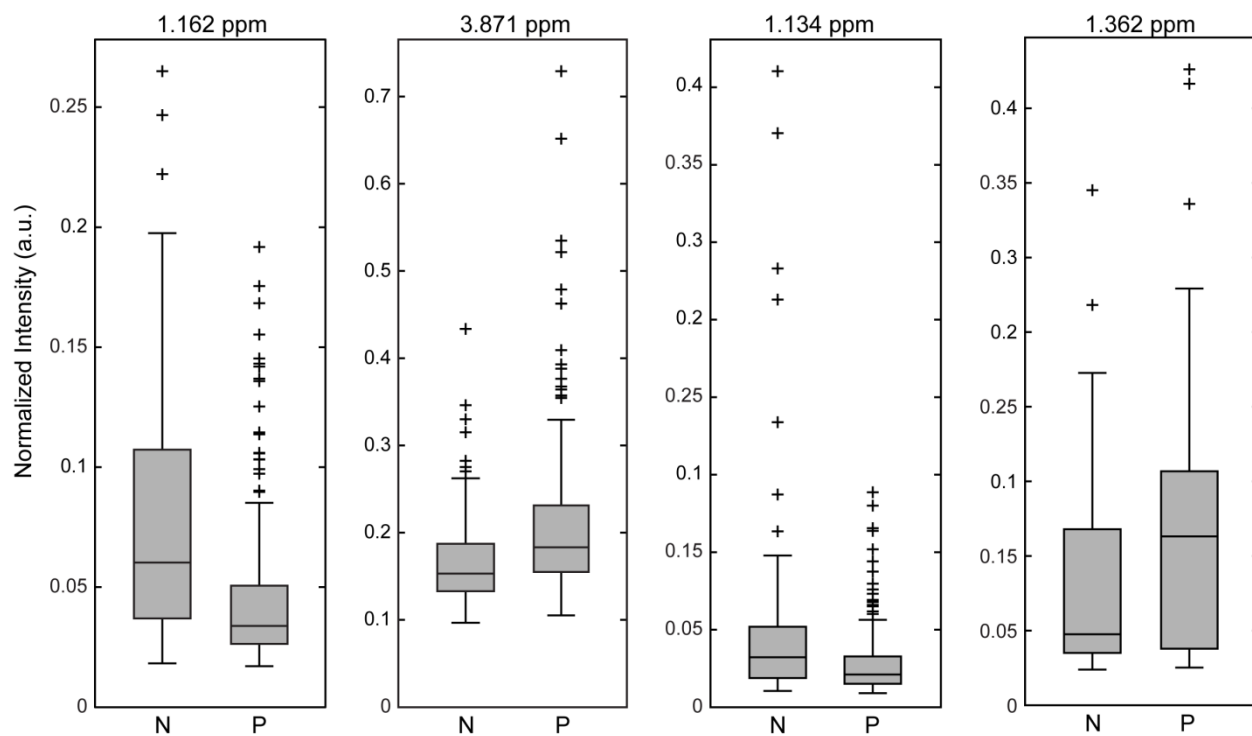
Figure 5.7 Representative signal boxplots of selected variables identified in the 'master list'.
The intensity was integrated at the signal indicated for the biopsy-negative (N) and biopsy-positive (P) spectra after all NMR data were normalized according the PQN procedure. The Wilcoxon p-values were $p < 0.0001$, $p < 0.0001$, $p = 0.0002$, 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.

Table 5.5 'Master list' of differential variables.
These variables appeared in at least half of the 6 variable lists generated.

Signal (ppm)
1.158
1.162
1.166
3.871
0.977
1.134
1.362
1.060
1.170
2.031
3.517
3.600
3.604
7.800
7.949
8.098
8.212
8.633

Table 5.6 Summary of model performance parameters for 2CV-PLS-DA modeling based on the 'master list' variable subset.
Using the original labels, 1000 models were generated while 500 models were generated using the permuted labels. All values are reported as mean ± standard deviation. In the case of permuted data, mean parameter values were first calculated over the 20 models per label permutation followed by mean ± standard deviation calculation over all 500 permutations.

	Data set	Misclassification	Q^2	AUC
Original Labels	Auto-scaled	26 ± 4	0.15 ± 0.11	0.80 ± 0.04
	BNCS*	27 ± 4	0.10 ± 0.11	0.79 ± 0.04
Permuted Labels	Auto-scaled	48 ± 3	-2.8 ± 0.9	0.49 ± 0.04
	BNCS*	48 ± 3	-5 ± 2	0.49 ± 0.03

NMR detection of metabolite markers of PCa has been primarily performed on prostate tissue (*e.x. vivo* or *in vitro*), 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. 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 mass spectrometry has identified several metabolites correlated with PCa, however these studies were performed on urine sediment. The sediment from urine collected post-DRE is expected to contain prostate cells which may be the source for these signals. Thus, urine supernatants 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 non-invasive biospecimen, may be further explored for PCa diagnostic potential. In combination with genetic[17] and protein urinary biomarker diagnostics under active investigation, a

completely non-invasive clinical test may be on the horizon capable of outperforming PSA levels for PCa identification.

5.5 Conclusion

It is encouraging that taken what 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, 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 resulting in increased difficulty in obtaining suitable multi-dimensional 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_2 moieties. Nevertheless, this study demonstrates the utility of NMR-based metabolomics of urine for the nomination of a PCa metabolite biomarker panel.

5.6 References

1.   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 (R Coll Radiol)* **2012,** 24, (8), 545-55.

2.   Lee, Y. J.; Park, J. E.; Jeon, B. R.; Lee, S. M.; Kim, S. Y.; Lee, Y. K., Is prostate-specific antigen effective for population screening of prostate cancer? A systematic review. *Ann Lab Med* **2013,** 33, (4), 233-41.

3.   Djulbegovic, M.; Beyth, R. J.; Neuberger, M. M.; Stoffs, T. L.; Vieweg, J.; Djulbegovic, B.; Dahm, P., Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. *Bmj-British Medical Journal* **2010,** 341.

4.   Walsh, P. C., Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *J Urol* **2003,** 170, (1), 313-4.

5.   Armstrong, A. J.; Eisenberger, M. A.; Halabi, S.; Oudard, S.; Nanus, D. M.; Petrylak, D. P.; Sartor, A. O.; Scher, H. I., Biomarkers in the management and treatment of men with metastatic castration-resistant prostate cancer. *Eur Urol* **2012,** 61, (3), 549-59.

6.   Roobol, M. J.; Haese, A.; Bjartell, A., Tumour markers in prostate cancer III: biomarkers in urine. *Acta Oncol* **2011,** 50 Suppl 1, 85-9.

7.   Prensner, J. R.; Rubin, M. A.; Wei, J. T.; Chinnaiyan, A. M., Beyond PSA: the next generation of prostate cancer biomarkers. *Sci Transl Med* **2012,** 4, (127), 127rv3.

8.   Roberts, M. J.; Schirra, H. J.; Lavin, M. F.; Gardiner, R. A., Metabolomics: a novel approach to early and noninvasive prostate cancer detection. *Korean J Urol* **2011,** 52, (2), 79-89.

9.   Catalona, W. J.; Smith, D. S.; Wolfert, R. L.; Wang, T. J.; Rittenhouse, H. G.; Ratliff, T. L.; Nadler, R. B., Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. *JAMA* **1995,** 274, (15), 1214-20.

10.   Sreekumar, A.; Laxman, B.; Rhodes, D. R.; Bhagavathula, S.; Harwood, J.; Giacherio, D.; Ghosh, D.; Sanda, M. G.; Rubin, M. A.; Chinnaiyan, A. M., Humoral immune response to alpha-methylacyl-CoA racemase and prostate cancer. *J Natl Cancer Inst* **2004,** 96, (11), 834-43.

11.   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.

12.   Theodorescu, D.; Fliser, D.; Wittke, S.; Mischak, H.; Krebs, R.; Walden, M.; Ross, M.; Eltze, E.; Bettendorf, O.; Wulfing, C.; Semjonow, A., Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis* **2005,** 26, (14), 2797-808.

13.   Tomlins, S. A.; Rhodes, D. R.; Perner, S.; Dhanasekaran, S. M.; Mehra, R.; Sun, X. W.; Varambally, S.; Cao, X.; Tchinda, J.; Kuefer, R.; Lee, C.; Montie, J. E.; Shah, R. B.; Pienta, K. J.; Rubin, M. A.; Chinnaiyan, A. M., Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **2005,** 310, (5748), 644-8.

14.   de Kok, J. B.; Verhaegh, G. W.; Roelofs, R. W.; Hessels, D.; Kiemeney, L. A.; Aalders, T. W.; Swinkels, D. W.; Schalken, J. A., DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *Cancer Res* **2002,** 62, (9), 2695-8.

15.   Laxman, B.; Tomlins, S. A.; Mehra, R.; Morris, D. S.; Wang, L.; Helgeson, B. E.; Shah, R. B.; Rubin, M. A.; Wei, J. T.; Chinnaiyan, A. M., Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia* **2006,** 8, (10), 885-8.

16.   Tomlins, S. A.; Aubin, S. M.; Siddiqui, J.; Lonigro, R. J.; Sefton-Miller, L.; Miick, S.; Williamsen, S.; Hodge, P.; Meinke, J.; Blase, A.; Penabella, Y.; Day, J. R.; Varambally, R.; Han, B.; Wood, D.; Wang, L.; Sanda, M. G.; Rubin, M. A.; Rhodes, D. R.; Hollenbeck, B.; Sakamoto, K.; Silberstein, J. L.; Fradet, Y.; Amberson, J. B.; Meyers, S.; Palanisamy, N.; Rittenhouse, H.; Wei, J. T.; Groskopf, J.; Chinnaiyan, A. M., Urine TMPRSS2:ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. *Sci Transl Med* **2011,** 3, (94), 94ra72.

17.   Payne, S. R.; Serth, J.; Schostak, M.; Kamradt, J.; Strauss, A.; Thelen, P.; Model, F.; Day, J. K.; Liebenberg, V.; Morotti, A.; Yamamura, S.; Lograsso, J.; Sledziewski, A.; Semjonow, A., 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), 1257-69.

18.   Whitaker, H. C.; Kote-Jarai, Z.; Ross-Adams, H.; Warren, A. Y.; Burge, J.; George, A.; Bancroft, E.; Jhavar, S.; Leongamornlert, D.; Tymrakiewicz, M.; Saunders, E.; Page, E.; Mitra, A.; Mitchell, G.; Lindeman, G. J.; Evans, D. G.; Blanco, I.; Mercer, C.; Rubinstein,

W. S.; Clowes, V.; Douglas, F.; Hodgson, S.; Walker, L.; Donaldson, A.; Izatt, L.; Dorkins, H.; Male, A.; Tucker, K.; Stapleton, A.; Lam, J.; Kirk, J.; Lilja, H.; Easton, D.; Cooper, C.; Eeles, R.; Neal, D. E., The rs10993994 risk allele for prostate cancer results in clinically relevant changes in microseminoprotein-beta expression in tissue and urine. *PLoS One* **2010,** 5, (10), e13363.

19.   Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M., Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009,** 457, (7231), 910-4.

20.   Jentzmik, F.; Stephan, C.; Miller, K.; Schrader, M.; Erbersdobler, A.; Kristiansen, G.; Lein, M.; Jung, K., 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), 12-8; discussion 20-1.

21.   Lucarelli, G.; Fanelli, M.; Larocca, A. M. V.; Germinario, C. A.; Rutigliano, M.; Vavallo, A.; Selvaggi, F. P.; Bettocchi, C.; Battaglia, M.; Ditonno, P., 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.

22.   Wu, H.; Liu, T.; Ma, C.; Xue, R.; Deng, C.; Zeng, H.; Shen, X., 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), 635-46.

23.   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), 166-72.

24.   Pan, Z.; Raftery, D., Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem* **2007,** 387, (2), 525-7.

25.   Zhang, J.; Wei, S.; Liu, L.; Nagana Gowda, G. A.; Bonney, P.; Stewart, J.; Knapp, D. W.; Raftery, D., NMR-based metabolomics study of canine bladder cancer. *Biochim Biophys Acta* **2012,** 1822, (11), 1807-14.

26.   Lodi, A.; Tiziani, S.; Khanim, F. L.; Gunther, U. L.; Viant, M. R.; Morgan, G. J.; Bunce, C. M.; Drayson, M. T., Proton NMR-based metabolite analyses of archived serial paired serum and urine samples from myeloma patients at different stages of disease activity identifies acetylcarnitine as a novel marker of active disease. *PLoS One* **2013,** 8, (2), e56422.

27.   Leo, G. C.; Darrow, A. L., NMR-based metabolomics of urine for the atherosclerotic mouse model using apolipoprotein-E deficient mice. *Magn Reson Chem* **2009,** 47 Suppl 1, S20-5.

28.   Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H., Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem* **2006,** 78, (13), 4281-90.

29.   Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K., Recursive segment-wise peak alignment of biological (1)h NMR spectra for improved metabolic biomarker recovery. *Anal Chem* **2009,** 81, (1), 56-66.

30.   MacKinnon, N.; Ge, W.; Khan, A. P.; Somashekar, B. S.; Tripathi, P.; Siddiqui, J.; Wei, J. T.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., Variable reference alignment: an improved peak alignment protocol for NMR spectral data with large intersample variation. *Anal Chem* **2012,** 84, (12), 5372-9.

31.   van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006,** 7, 142.

32.   Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A., Assessment of PLSDA cross validation. *Metabolomics* **2008,** 4, (1), 81-89.

33.   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.

34.   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.

35.   Wayner, E. A.; Quek, S. I.; Ahmad, R.; Ho, M. E.; Loprieno, M. A.; Zhou, Y.; Ellis, W. J.; True, L. D.; Liu, A. Y., Development of an ELISA to detect the secreted prostate cancer biomarker AGR2 in voided urine. *Prostate* **2012,** 72, (9), 1023-34.

36.   Edmands, W. M.; Beckonert, O. P.; Stella, C.; Campbell, A.; Lake, B. G.; Lindon, J. C.; Holmes, E.; Gooderham, N. J., Identification of human urinary biomarkers of cruciferous vegetable consumption by metabonomic profiling. *J Proteome Res* **2011,** 10, (10), 4513-21.

37.   Lenz, E. M.; Bright, J.; Wilson, I. D.; Hughes, A.; Morrisson, J.; Lindberg, H.; Lockton, A., 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), 841-9.

38.   Brereton, R. G., 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.

39.   Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J., Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal Chem* **2005,** 77, (5), 1282-9.

40.   Coen, M.; Hong, Y. S.; Cloarec, O.; Rhode, C. M.; Reily, M. D.; Robertson, D. G.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Heteronuclear 1H-31P statistical total correlation NMR spectroscopy of intact liver for metabolic biomarker assignment: application to galactosamine-induced hepatotoxicity. *Anal Chem* **2007,** 79, (23), 8956-66.

CHAPTER 6

NMR Spectroscopy Based Metabolic Profiling of Urine for the Non-invasive Detection of Human Bladder Cancer

6.1 Abstract

Diagnosis and surveillance of bladder cancer (BCa) with the combination of cystoscopy and urine cytology, currently the gold standard, suffers from poor sensitivity, particularly for low-grade tumors leading to invasive follow-up procedures. To explore alternate approaches for non-invasive detection of BCa, we performed ^{1}H NMR-based global metabolic profiling of urine specimens collected from BCa (n=50) and healthy controls (n=35), and identified more than thirty-two differentially expressed metabolites. Unbiased, principal component analysis performed on NMR dataset (creatinine-normalized) showed no clear segregation between control and BCa group. Univariate statistical analysis was performed on binned NMR dataset to investigate significantly altered ^{1}H chemical shift regions ($P < 0.05$). Receiver operating characteristic curve was generated on these significant chemical regions and only regions with area under the curve (AUC) > 0.6 were chosen for inclusion in a panel of metabolic biomarkers. In total seven metabolites, 1,2 propanediol (P <0.0001), trimethylamine N-oxide (P <0.007), taurine (P <0.01), p-cresol sulfate (P <0.0001), phenylacetylglutamine (P <0.001) and 3-aminobutyric acid (P <0.006) were significantly elevated, whereas 2-furoylglycine (P <0.001) was significantly decreased in BCa group compared to normal samples. All seven

elevated metabolites showed AUC values ranging from 0.67 to 0.81. Majority of these altered metabolites were gut microbial co-metabolites suggesting intimate association of gut microbiome dynamics with bladder cancer progression. In summary, our results suggest that NMR-based metabolic profiling of bladder cancer urine can be a quick, cost effective and promising tool for non-invasive detection of BCa.

6.2 Introduction

Bladder Cancer (BCa) is the ninth most frequent genito-urinary malignancy diagnosed worldwide and fifth leading new cancer diagnosed in the United States[1]. Of newly diagnosed BCa patients having non-muscle invasive papillary disease (i. e. restricted to urothelium or lamina propia), a subgroup of high-risk patients will progress to lethal, metastatic disease resulting in significant morbidity. Current clinical diagnosis and surveillance of BCa rely on urine cytology and cystoscopy followed by histopathology of tissue specimens obtained from tumor biopsies, the current gold standard. Cystoscopy has only modest sensitivity particularly for low-grade (Ta-T1 stage) tumors[2] but adds significantly to patients' discomfort, as well as economic burden. To overcome these challenges, alternative strategies are warranted specifically non-invasive diagnostic biomarkers that could replace cystoscopy, especially for low-risk patients that can accurately predict disease aggressiveness, recurrence and progression. Due to the functional, biologically significant relationship between bladder and urine, urine samples are ideal and relevant medium for identifying urinary biomarkers for the non-invasive detection of BCa.

Recently transcriptomic and proteomic approaches identified multiple potential urinary biomarkers for non-invasive detection of BCa[3, 4, 5]. While genomic and transcriptomic approaches discover aberrations at the DNA, RNA and protein levels, metabolomic approaches uncover global changes in low molecular weight metabolites in biological samples that occur in response to various environmental and cellular signals. Metabolomics studies produce plethora of information that are complementary to those obtained from genomics and proteomics and collective information has been used for the global assessment of metabolic network within a living system[6]. Similar to genomics and transcriptomics, metabolomics approach has also been used for developing diagnostic/prognostic biomarkers for specific diseases[7]. Generally metabolomic profiling has been performed using analytical platforms such as Nuclear Magnetic Resonance (NMR) and Mass spectrometry (MS) techniques[8] individually or in combination. Both these techniques have inherent advantages and disadvantages but can provide complementary and supplementary metabolic data that can be used to construct a panel of disease-specific biomarkers. Metabolic signatures have been identified in many cancer types such as prostate[7], esophageal[9], pancreatic ductal[10, 11], colon[12] among others.

Recently we performed metabolic profiling of bladder tissues using high-resolution magic angle spinning (HR-MAS) NMR to distinguish BCa tissue metabolome from that of benign tissues[8]. Only a limited number of metabolomics studies exploring urinary metabolic signatures for human bladder cancer have been performed till date using GC/TOFMS (gas chromatography/time-of-flight MS) and LC-MS (liquid chromatography-MS) [13, 14, 15]. One NMR study reportedly identified elevated taurine levels in BCa urine samples[16]. Here, towards the goal of developing a panel of urinary metabolic signatures

for noninvasive detection of BCa, we performed NMR-based global metabolic profiling of urine samples collected from patients and healthy volunteers. We identified dysregulation in several metabolites in BCa urine samples that can potentially be developed into a multiplex metabolic panel to detect BCa.

6.3 Materials and Methods

6.3.1 Urine samples collection

Urine samples were collected in sterile containers either during a clinic visit or prior to transurethral resection of bladder tumor (TURBT) or cystectomy from BCa patients enrolled in the bladder cancer program in the Department of Urology at the University of Michigan. All urine samples were collected with prior approval from Institutional Review Board and informed consent from patients. The tumor stage and grade were confirmed by histopathological evaluation of tumor biopsies obtained through cystoscopy, which was established according to World Health Organization and International Society of Urologic Pathology Standards[17]. Normal urine samples were collected from healthy volunteers who do not have symptoms of bladder cancer or other urological disorders. Immediately after urine collection, the samples were stored at -80°C until sample preparation for metabolomic analysis. Detailed information of the patients' clinico-pathologic parameters is included in Table 6.1.

Table 6.1 Clinical Characteristics of the BCa Patients.

Characteristics	BCa patients	Normal
No. of Patients	50	35
Gender (Male/Female)	44/6	25/10
Age (Mean ± SD)	68.1 ± 10.8 years	55.2 ± 17.6 years
Pathological stage/grade		
Tis	10	
Ta – Low grade/High grade	5/15	
T1	10	
T2	10	

6.3.2 Chemicals

K_2HPO_4, KH_2PO_4, 3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (TSP), Deuterium oxide (D_2O), 3-aminoisobutyric acid, creatinine, Dimethylformamide and N-methyl-N-(tert-butylmethylsilyltrifluoroacetamide) (MtBSTFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

6.3.3 Sample preparation for NMR Spectroscopy

Urine samples stored at -80°C were obtained and allowed to thaw at room temperature. 600 µL of each sample was centrifuged (10,000 rpm, 10 min at 4°C) and 450 µL of clear supernatant liquid was transferred into a 5 mm NMR tube (Wilmad, NJ, USA). 50 µL D_2O containing a mixture of 1 M phosphate buffer (pH 7.2) and 100 µg TSP was thoroughly mixed with 450 µL urine in NMR tube and analyzed by NMR.

6.3.4 NMR Spectroscopy

All the NMR experiments were carried out on a Bruker Avance 900 MHz NMR spectrometer (Bruker BioSpin, Karlsruhe, Germany) equipped with a 5 mm HCN TCI cryo-probe at 298 K temperature and NMR spectra were processed using Bruker TOPSPIN 2.1 software. One-dimensional (1D) ^1H spectra were acquired using a one-pulse experiment with water presaturation during relaxation delay. Typical parameters used were: pulse length- 45°; spectral width- 14492.75 Hz (16.1091 ppm); data points-32078; relaxation delay- 2s; number of transients- 256. The resultant free induction decays were zero filled with 32,768 points and multiplied by an exponential window function with a line broadening factor of 0.3 Hz prior to Fourier transformation, and the frequency domain spectra were manually phased and baseline corrected.

For chemical assignment of urine metabolites, two-dimensional (2D) experiments such as ^{1}H-^{1}H total correlation spectroscopy (TOCSY) and ^{1}H-^{13}C heteronuclear single quantum coherence spectroscopy (HSQC) were performed on urine samples. 2D TOCSY spectra were acquired in the phase-sensitive time proportional phase incrementation (TPPI) mode with MLEV17 spin lock mixing time of 70 ms. Parameters used for ^{1}H-^{1}H TOCSY: spectral width in both dimension- 14492.75 Hz; relaxation delay- 1.5s; data points- 2048; number of transients- 64; number of t_1 increments- 256. ^{1}H-^{13}C HSQC spectra were acquired using phase-sensitive echo-antiecho mode: spectral width in ^{1}H and ^{13}C dimension- 14492.75 Hz and 40722.58 Hz, respectively; data points- 2048; number of transients- 128; number of t_1 increments- 256; relaxation delay- 1.5 s. Both TOCSY and HSQC data were zero filled to 2048 × 1024 matrix and Fourier transformed along both dimensions after multiplying the data by a squared sine-bell window function shifted by $\pi/2$.

6.3.5 Statistical analysis

Before performing any statistical analyses, the chemical shift regions of 0.5 – 10.0 ppm of all the ^{1}H NMR spectra of urine samples were converted to ASCII-files and the spectral regions of residual water and urea (4.5 – 6.0 ppm) were excluded in further analyses. All the ^{1}H spectra were normalized to creatinine signal at 3.05 ppm and were aligned using variable reference-interval correlated shifting (Vr-icoShift)[18]. Principal component analysis (PCA) was performed with mean-centered data pretreatment and the results were depicted as scores and loadings plot (The Unscrambler V10.1, CAMO, Oslo, Norway). For univariate statistical analysis (SPSS v 21, SPSS Inc., Chicago, IL, USA), the creatinine-normalized spectra were reduced with a bin size of 0.004 ppm. Each

variable was compared between normal and BCa samples using Mann-Whitney U test and $P < 0.05$ was considered statistically significant. Further statistically significant chemical shift variables were selected and the corresponding intensities of metabolites were calculated separately for normal and BCa urine samples and subjected to receiver operating characteristic (ROC) curve analysis. The metabolites with AUC > 0.60 were considered differentially altered metabolite.

6.3.6 Z-scores

To visualize overall changes in the urinary metabolic pattern, z-scores were calculated using creatinine-normalized binned NMR data. The mean and standard deviation of the normal samples were calculated for each binned chemical shift variable. Further, each normal and BCa sample was centered by the normal mean and scaled by the normal standard deviation.

6.3.7 GC-MS Targeted Analysis

Urine samples were homogenized in methanol after spiking labeled internal standards (d^3-aminoisobutyric acid and d^1 creatinine) and placed on a shaker overnight at 4°C. The extraction was carried out using 1:1 molar ratio of water/chloroform at room temperature. The aqueous methanolic layer was collected and dried completely under nitrogen. The dried methanolic amino acid residue was azeotroped twice by adding 100 μL dimethylformamide (DMF), vortexing, and then drying in a speed vac for 30 minutes. 100 μL of DMF and N-methyl-N-(tert-butylmethylsilyltrifluoroacetamide (MtBSTFA) + 1% t-butyl-dimethylchlorosilane was added to the dried sample, capped and incubated at 60°C for 1 hr and then resuspended in ethyl acetate before injection into a GC-MS. Selective Ion Monitoring (SIM) was used for quantification of 3-aminoisobutyric acid and

creatinine. The amount of metabolite in the sample was calculated by measuring the peak area of the native metabolite normalized to the area of the peak for the isotope-labeled internal standard. Creatinine was used to normalize 3-aminoisobutyric acid levels.

6.4 Results

6.4.1 Clinical characteristics of BCa patients and urine samples

BCa patient urine samples were procured from the Bladder Cancer Program at Department of urology, University of Michigan. Fifty BCa patients at various stages of BCa were included in the study: ten patients with non-invasive flat carcinoma (Tis), twenty patients with non-invasive papillary carcinoma (Ta) with five low- and fifteen high grade, ten were staged at T1 and ten at T2. For comparison, urine samples were collected from thirty-five healthy volunteers who had no symptoms of urological disease. Detailed clinical parameters of patients are shown in Table 6.1.

6.4.2 ^{1}H NMR profiling of BCa and normal urine samples

^{1}H NMR spectra were obtained for fifty BCa and thirty-five normal urine samples. The overall workflow of ^{1}H NMR metabolic profiling of urine samples is shown in Figure 6.1. The chemical shift assignment of various metabolites was carried out using a combination of 2D ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC, Human Metabolome Database (HMDB)[19], Biological Magnetic Resonance Data Bank (BMRB)[20], MetaboID[21] and literature chemical shift database[22, 23]. Figure 6.2 depicts a representative ^{1}H NMR spectra (normalized to creatinine) of two each from normal (in black) and BCa patients (in red) with chemical assignment of various metabolites. A wide range of metabolites were detected including amino acids and their derivatives, TCA cycle intermediates,

nicotinamide metabolites, ketone bodies, short-chain fatty acids, metabolites related to osmoregulation, gut microbial metabolites and several other endogenous metabolites.
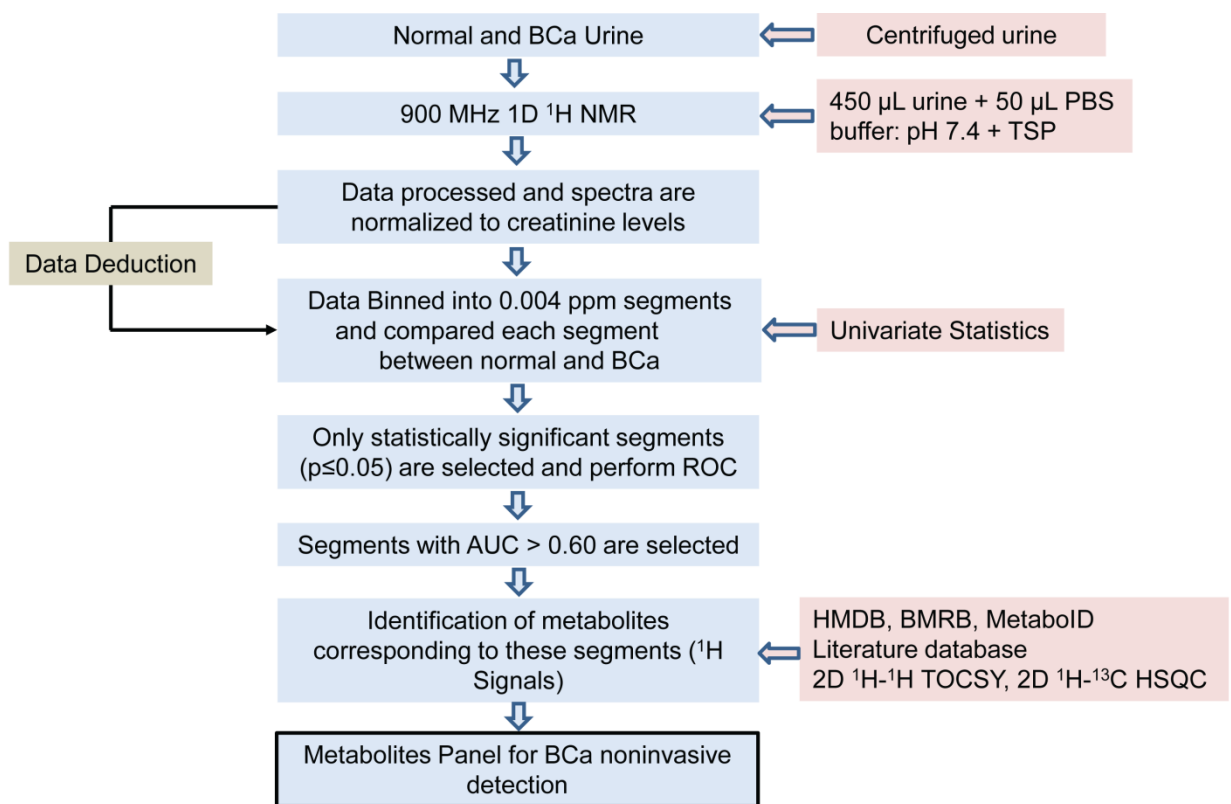
Figure 6.1 Flow chart showing an overview of metabolomics workflow in NMR-based metabolic profiling of bladder cancer and normal urine.
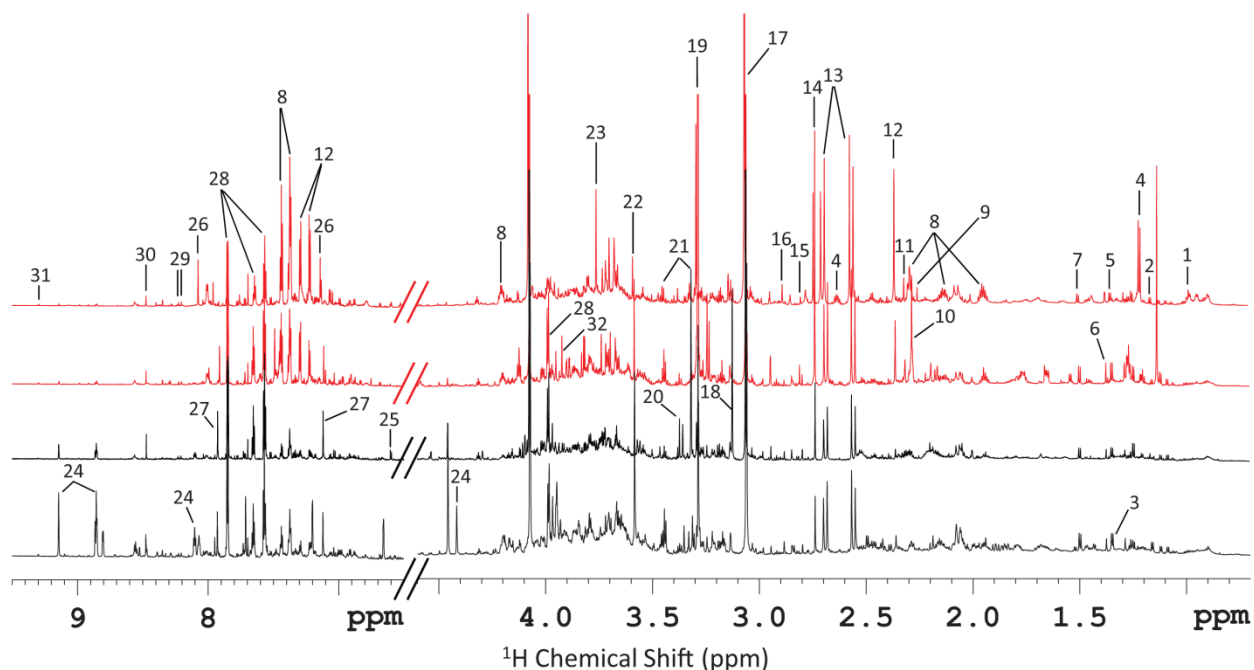
Figure 6.2 Representative 900 MHz 1H NMR spectra of urine samples from healthy controls (bottom two in black) and bladder cancer (top two in red) with chemical shift assignment of various metabolites.

The spectral regions of residual water and urea (4.6 – 6.5 ppm) were not shown and all the spectra were normalized to ^{1}H peak of creatinine at 3.05 ppm. Chemical shift assignments were: 1) Leucine, 2) 1,2 propanediol, 3) Threonine, 4) 3-aminoisobutyric acid, 5) Lactate, 6) 2-hydroxyvalerate, 7) Alanine, 8) Phenylacetylglutamine, 9) Acetoacetate, 10) Pyruvate, 11) Oxaloacetate, 12) p-Cresol sulfate, 13) Citrate, 14) Dimethylamine, 15) Dimethylglycine, 16) Trimethylamine, 17) Creatinine, 18) Malonate, 19) Trimethylamine N-oxide, 20) *Scyllo*-Inositol, 21) Taurine, 22) Glycine, 23) Guanidoacetate, 24) Trigonelline, 25) 2-furoylglycine, 26) 3-methylhistidine, 27) 1-methylhistidine, 28) Hippurate, 29) Hypoxanthine, 30) Formate, 31) N-methylnicotinamide, 32) Betaine.

6.4.3 Statistical Analysis of urine ^{1}H NMR data

In order to visualize the differential metabolic profiles of normal and BCa group, unsupervised PCA was performed on the creatinine-normalized NMR data; however, PCA score plot did not show clear segregation between normal and BCa samples. We then performed univariate statistical analysis on NMR dataset to identify differentially expressed metabolites in BCa samples relative to control samples. After removing unwanted signals from creatinine-normalized NMR data, the dataset was reduced to

0.004 ppm bins resulting in 1932 variables. Each of these variables was compared between normal and BCa groups using U test and only statistically significant variables ($P < 0.05$) were selected for ROC curve analysis. Further, the variables having area under the curve (AUC) > 0.6 were selected for inclusion in a multiplex panel of markers for BCa.

6.4.4 Urinary Metabolic Signatures of Bladder Cancer

Z-score plots of 1932 bins of normal (blue) and BCa groups (red) normalized to the normal mean and scaled to normal standard deviation are shown in Figure 6.3 where overall changes in the ^{1}H signal intensities between the two groups can be visualized. Furthermore, variables having AUC > 0.6 were used to identify the metabolites from these chemical shift values. These chemical shift values corresponded to seven metabolites: 1,2 propanediol, trimethylamine N-oxide (TMAO), taurine, 2-furoylglycine, p-cresol sulfate, phenylacetylglutamine (PAGN) and 3-aminoisobutyric acid (ABA). Relative intensities of these significantly elevated metabolites (with respect to creatinine levels) were calculated separately for normal and BCa groups using full resolution creatinine-normalized NMR data set. A selective non-overlapping ^{1}H signal from each metabolite was used to calculate the ratio of metabolite to creatinine. ^{1}H signal used for quantification were: 1,2 propanediol (1.158 ppm), taurine (3.44 ppm), furoylglycine (6.656 ppm), p-cresol sulfate (7.295 ppm) and phenylacetylglutamine (7.438 ppm). Since the singlet ^{1}H signal of TMAO (3.28 ppm) was overlapping with that of S-CH$_2$ triplet signal of taurine, the intensity of N-CH$_2$ triplet of taurine (3.44 ppm) was subtracted from the total intensity of TMAO (N-(CH$_3$) + taurine (S-CH$_2$) to obtain the correct intensity of TMAO. Similarly, ^{1}H signals of 3-aminoisobutyric acid also overlapped with other ^{1}H signals of many other metabolites, making it difficult to accurately measure their levels using NMR dataset. Therefore, we

131

used targeted GC-MS analysis to measure the levels of 3-aminoisobutyric acid with respect to creatinine.
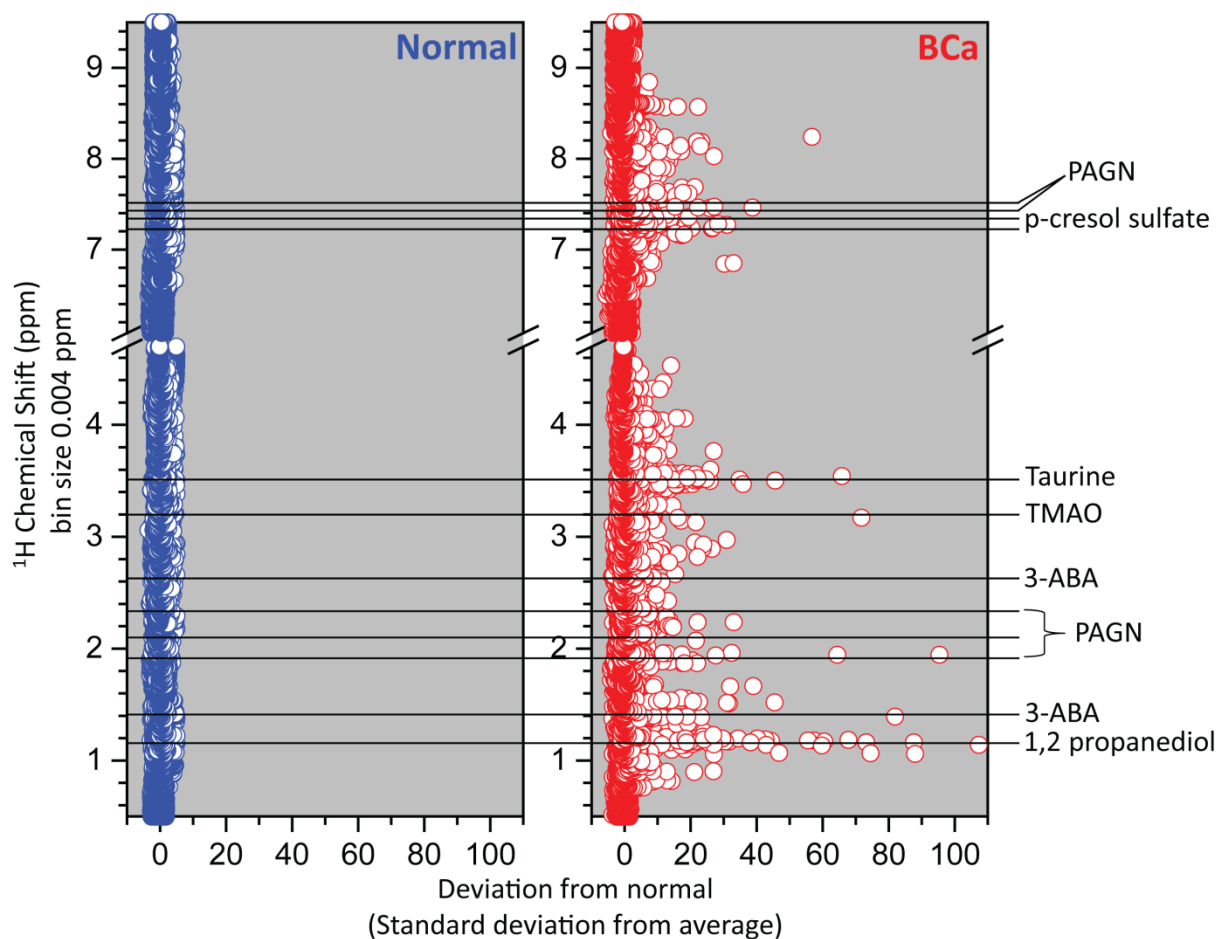
Figure 6.3 z-score plots of ^{1}H NMR data (1932 bins) of normal and bladder cancer urine samples. Both sample groups were normalized to the mean and scaled to standard deviation of normal samples. Significant variation in the z-scores of bladder cancer urine can be observed and altered ^{1}H signal intensities are highlighted with corresponding metabolites.

Figure 6.4 shows bar plots ratio of seven differentially expressed metabolites with respect to creatinine levels. Six metabolites, 1,2 propanediol ($P < 0.0001$), TMAO ($P < 0.007$), taurine ($P < 0.01$), p-cresol sulfate ($P < 0.0001$), phenylacetylglutamine ($P < 0.001$) and 3-aminoisobutyric acid ($P < 0.006$) were significantly elevated and furoylglycine ($P < 0.0001$) was significantly depleted in BCa samples compared to control samples. ROC

curves of these differential metabolites are shown in Figure 6.5 and corresponding AUC values varied from 0.67 to 0.81.
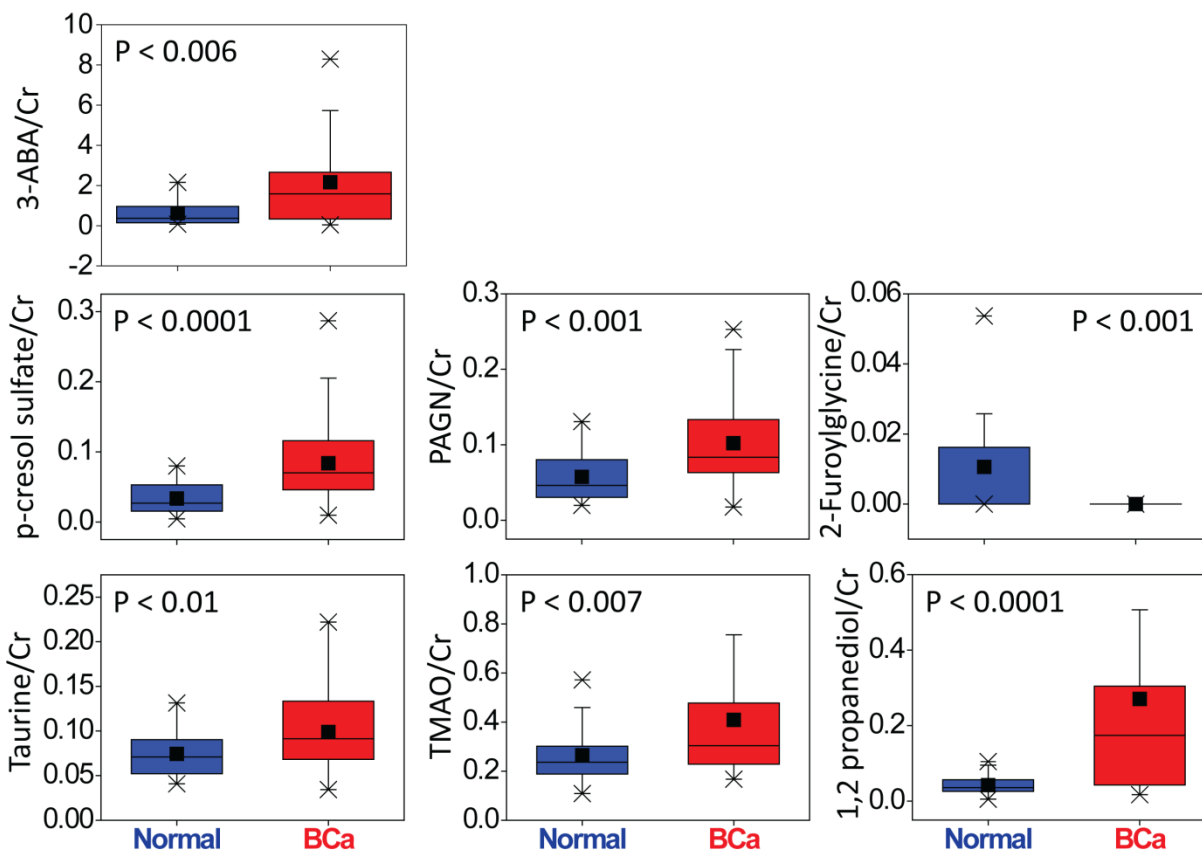
Figure 6.4 Box plot showing the levels of seven urine metabolites (normalized to creatinine) that were measured in healthy and bladder cancer urine.
The level of 3-aminoisobutyric acid was measured using GC-MS and rest of the six metabolites was measured by NMR. In each case, the ratio of metabolite/creatinine was compared between normal and bladder cancer group using Mann-Whitney U test and six metabolites were significantly elevated and 2-furoylglycine is significantly reduced in BCa sample compared to normal.
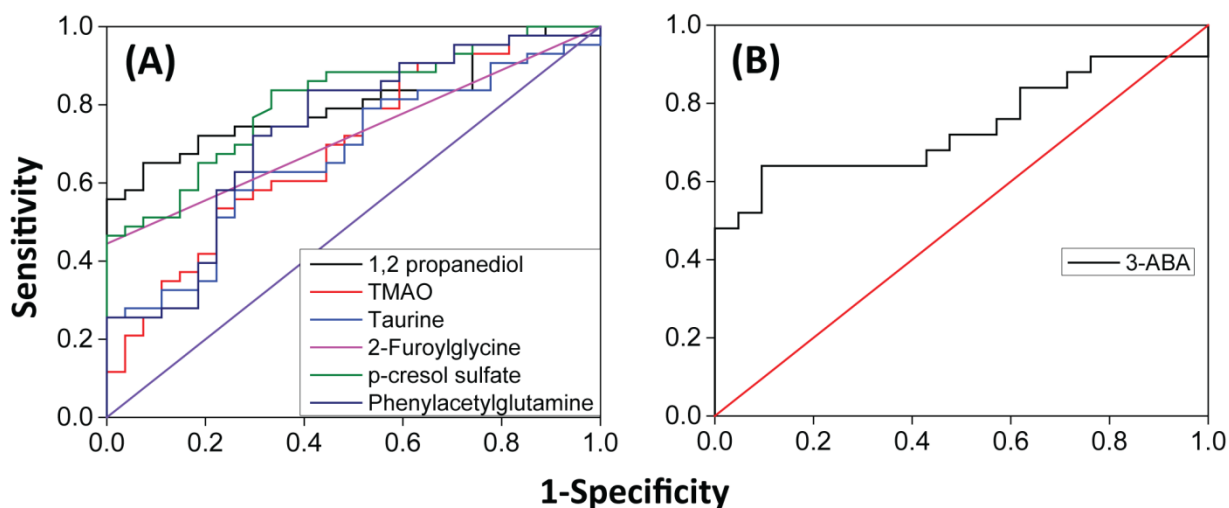
Figure 6.5 ROC curves generated from metabolite/creatinine ratio of six metabolites measured by NMR (A) and (B) 3-aminoisobutyric acid measured from GC-MS.
AUC values for 1,2 propanediol, TMAO, taurine, 2-furoylglycine, p-cresol sulfate, phenylacetylglutamine and 3-aminoisobutyric acid are 0.80, 0.69, 0.67, 0.72, 0.81, 0.73 and 0.74, respectively.

6.5 Discussion

Current diagnosis of bladder cancer based on cystoscopy and urine cytology suffer from low sensitivity especially towards low-grade tumors; moreover, cystoscopy is one of the most expensive procedures[3]. Therefore the discovery of non-invasive biomarkers that are efficient and cost-effective would be of significant value. Utilize metabolomics approach, we recently performed HR-MAS NMR metabolic profiling of BCa and benign tissues, and identified several dysregulated metabolites specifically in BCa that could potentially be employed to distinguish tumor from benign tissues[8]. Extending upon those studies here, we conducted a NMR-based metabolomics study focused on unbiased metabolic profiling of BCa and normal urine in an effort to develop a non-invasive diagnostic tool. Due to the functional significance of bladder, urine is an ideally suited specimen for exploring urinary biomarkers for the non-invasive detection of BCa. We performed ^{1}H NMR metabolic profiling of BCa and normal urine samples in which wide

134

range of metabolites were identified. In the present study we adopted creatinine normalization based on the fact that creatinine clearance in bladder cancer urine is constant and comparable to normal urine[24]. Unsupervised PCA performed on creatinine-normalized NMR dataset did not segregate normal and BCa urine samples. If PCA fails to segregate different classes, generally supervised statistical analysis such as partial least square discriminant analysis (PLS-DA) and OPLS-DA are performed. However we did not pursue any of these supervised analyses to avoid potential over-fitting in segregating normal and BCa groups. Instead we performed U test ($P < 0.05$) followed by ROC analysis (AUC > 0.6) which revealed differential expression of seven metabolites, 1,2 propanediol, TMAO, taurine, p-cresol sulfate, phenylacetylglutamine, 3-aminoisobutyric acid and 2-furoylglycine.

Elevated taurine in BCa urine observed in the present study is consistent with recent NMR report on bladder cancer[16]. Notably, taurine is the only common metabolite elevated in both BCa urine and tissues[8]. Taurine is associated with multiple biological processes such as osmoprotection, volume-activated organic flux pathways[25], antineoplastic and antioxidant properties[26], and gut microbial metabolism[27]. Although it is not certain how taurine is involved in BCa development or progression, bladder cancer tissues experience oxidative stress[8, 28] and elevated taurine may be associated with antioxidant pathway. Increased excretion of TMAO in BCa urine may be correlated with osmotic stress mechanism in bladder cancer where TMAO acts as an omsoprotectant. The metabolite 2-furoylglycine, a conjugation product of glycine and 2-furoic acid was not detected in any BCa samples but detected only in normal urine samples. Elevated excretion of furoylglycine in urine is strongly associated with dietary consumption of

chocolates and heated fruit juice[29] but its correlation with bladder cancer remains unknown. Similarly the specific mechanism associated with increased excretion of 1,2 propanediol in bladder cancer urine is also unclear.

The gut microbial co-metabolites such as phenylacetylglutamine, p-cresol sulfate and 3-aminoisobutyric acid were significantly elevated in BCa urine. The correlation of these metabolites with human gut microbiome were described[27] wherein phenylacetylglutamine and 4-cresol sulfate are linked with *Subdoligranulum variable BI-114* and *Bifidobacterium pseudocatenulatum,* and 3-aminoisobutyric acid *is* linked with three gut microbes, *Bacteroidetes thetaiotaomicron, Bacteroidetes uniformis and Ruminococcus bromii*. Dysregulation of human microbiota has been reported in various diseases and associated with drug metabolism[30]. The association of gut microbes with colon[31] and liver cancer[32] has also been reported. Our present study clearly indicates a correlation between gut microbial metabolites and bladder cancer and warrants further investigation. Many of the significantly altered metabolites observed in our NMR study have not been previously reported in mass spectrometry-based metabolomics studies of bladder cancer[13, 14, 15].

In the present study, we explored the possibility of NMR-based metabolomics approach for the non-invasive detection of bladder cancer in urine samples. Due to limited sample size we did not analyze metabolite levels across various stages and grades of bladder cancer. Ideally a large sample cohort involving multi-institutions would be required for a validation study. We are acquiring additional urine samples from bladder cancer patients and healthy volunteers to conduct a validation study of seven metabolites reported here. Since we presented our results as metabolite/creatinine ratio for

significantly altered metabolites, it will be a straightforward for other research groups to validate these metabolites in urine samples of bladder cancer patients. The panel of seven urinary metabolite biomarkers reported here from NMR study complements the study by Putluri et al[15] and Pasikanti et al[13, 14].

We believe that NMR-based metabolic profiling of bladder cancer urine is a promising approach that warrants further independent validation and eventual development as an efficient and cost effective diagnostic tool. Importantly, our results uncovered the association of gut microbiome with bladder cancer progression.

6.6 References

1.   Siegel, R.; Naishadham, D.; Jemal, A., Cancer statistics, 2012. *CA Cancer J Clin* **2012,** 62, (1), 10-29.

2.   Avritscher, E. B.; Cooksley, C. D.; Grossman, H. B.; Sabichi, A. L.; Hamblin, L.; Dinney, C. P.; Elting, L. S., Clinical model of lifetime cost of treating bladder cancer and associated complications. *Urology* **2006,** 68, (3), 549-53.

3.   Frantzi, M.; Makridakis, M.; Vlahou, A., Biomarkers for bladder cancer aggressiveness. *Curr Opin Urol* **2012,** 22, (5), 390-6.

4.   Chen, C. L.; Lin, T. S.; Tsai, C. H.; Wu, C. C.; Chung, T.; Chien, K. Y.; Wu, M.; Chang, Y. S.; Yu, J. S.; Chen, Y. T., Identification of potential bladder cancer markers in urine by abundant-protein depletion coupled with quantitative proteomics. *J Proteomics* **2013,** 85, 28-43.

5.   Shimwell, N. J.; Bryan, R. T.; Wei, W.; James, N. D.; Cheng, K. K.; Zeegers, M. P.; Johnson, P. J.; Martin, A.; Ward, D. G., Combined proteome and transcriptome analyses for the discovery of urinary biomarkers for urothelial carcinoma. *Br J Cancer* **2013,** 108, (9), 1854-61.

6.   Tomita, M.; Kami, K., Cancer. Systems biology, metabolomics, and cancer metabolism. *Science* **2012,** 336, (6084), 990-1.

7.   Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.;

Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M., Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009,** 457, (7231), 910-4.

8.   Tripathi, P.; Somashekar, B. S.; Ponnusamy, M.; Gursky, A.; Dailey, S.; Kunju, P.; Lee, C. T.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., HR-MAS NMR tissue metabolomic signatures cross-validated by mass spectrometry distinguish bladder cancer from benign disease. *J Proteome Res* **2013,** 12, (7), 3519-28.

9.   Davis, V. W.; Schiller, D. E.; Eurich, D.; Sawyer, M. B., Urinary metabolomic signature of esophageal cancer and Barrett's esophagus. *World J Surg Oncol* **2012,** 10, 271.

10.   Davis, V. W.; Schiller, D. E.; Eurich, D.; Bathe, O. F.; Sawyer, M. B., Pancreatic Ductal Adenocarcinoma is Associated with a Distinct Urinary Metabolomic Signature. *Ann Surg Oncol* **2012**.

11.   Napoli, C.; Sperandio, N.; Lawlor, R. T.; Scarpa, A.; Molinari, H.; Assfalg, M., Urine metabolic signature of pancreatic ductal adenocarcinoma by (1)h nuclear magnetic resonance: identification, mapping, and evolution. *J Proteome Res* **2012,** 11, (2), 1274-83.

12.   Qiu, Y.; Cai, G.; Su, M.; Chen, T.; Liu, Y.; Xu, Y.; Ni, Y.; Zhao, A.; Cai, S.; Xu, L. X.; Jia, W., Urinary metabonomic study on colorectal cancer. *J Proteome Res* **2010,** 9, (3), 1627-34.

13.   Pasikanti, K. K.; Esuvaranathan, K.; Ho, P. C.; Mahendran, R.; Kamaraj, R.; Wu, Q. H.; Chiong, E.; Chan, E. C., Noninvasive urinary metabonomic diagnosis of human bladder cancer. *J Proteome Res* **2010,** 9, (6), 2988-95.

14.   Pasikanti, K. K.; Esuvaranathan, K.; Hong, Y.; Ho, P. C.; Mahendran, R.; Raman Nee Mani, L.; Chiong, E.; Chan, E. C., Urinary metabotyping of bladder cancer using two-dimensional gas chromatography time-of-flight mass spectrometry. *J Proteome Res* **2013,** 12, (9), 3865-73.

15.   Putluri, N.; Shojaie, A.; Vasu, V. T.; Vareed, S. K.; Nalluri, S.; Putluri, V.; Thangjam, G. S.; Panzitt, K.; Tallman, C. T.; Butler, C.; Sana, T. R.; Fischer, S. M.; Sica, G.; Brat, D. J.; Shi, H.; Palapattu, G. S.; Lotan, Y.; Weizer, A. Z.; Terris, M. K.; Shariat, S. F.; Michailidis, G.; Sreekumar, A., Metabolomic profiling reveals potential markers and bioprocesses altered in bladder cancer progression. *Cancer Res* **2011,** 71, (24), 7376-86.

16.   Srivastava, S.; Roy, R.; Singh, S.; Kumar, P.; Dalela, D.; Sankhwar, S. N.; Goel, A.; Sonkar, A. A., Taurine - a possible fingerprint biomarker in non-muscle invasive bladder cancer: A pilot study by 1H NMR spectroscopy. *Cancer Biomark* **2010,** 6, (1), 11-20.

17.   Epstein, J. I.; Amin, M. B.; Reuter, V. R.; Mostofi, F. K., The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of the urinary bladder. Bladder Consensus Conference Committee. *Am J Surg Pathol* **1998,** 22, (12), 1435-48.

18.   MacKinnon, N.; Ge, W.; Khan, A. P.; Somashekar, B. S.; Tripathi, P.; Siddiqui, J.; Wei, J. T.; Chinnaiyan, A. M.; Rajendiran, T. M.; Ramamoorthy, A., Variable reference

alignment: an improved peak alignment protocol for NMR spectral data with large intersample variation. *Anal Chem* **2012,** 84, (12), 5372-9.

19.   Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* **2013,** 41, (D1), D801-7.

20.   Markley, J. L.; Ulrich, E. L.; Berman, H. M.; Henrick, K.; Nakamura, H.; Akutsu, H., BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. *J Biomol NMR* **2008,** 40, (3), 153-5.

21.   Mackinnon, N.; Somashekar, B. S.; Tripathi, P.; Ge, W.; Rajendiran, T. M.; Chinnaiyan, A. M.; Ramamoorthy, A., MetaboID: A graphical user interface package for assignment of (1)H NMR spectra of bodyfluids and tissues. *J Magn Reson* **2013,** 226, 93-9.

22.   Holmes, E.; Foxall, P. J.; Spraul, M.; Farrant, R. D.; Nicholson, J. K.; Lindon, J. C., 750 MHz 1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease. *J Pharm Biomed Anal* **1997,** 15, (11), 1647-59.

23.   Diaz, S. O.; Barros, A. S.; Goodfellow, B. J.; Duarte, I. F.; Carreira, I. M.; Galhano, E.; Pita, C.; Almeida Mdo, C.; Gil, A. M., Following healthy pregnancy by nuclear magnetic

resonance (NMR) metabolic profiling of human urine. *J Proteome Res* **2013,** 12, (2), 969-79.

24.   Abogunrin, F.; O'Kane, H. F.; Ruddock, M. W.; Stevenson, M.; Reid, C. N.; O'Sullivan, J. M.; Anderson, N. H.; O'Rourke, D.; Duggan, B.; Lamont, J. V.; Boyd, R. E.; Hamilton, P.; Nambirajan, T.; Williamson, K. E., The impact of biomarkers in multivariate algorithms for bladder cancer diagnosis in patients with hematuria. *Cancer* **2012,** 118, (10), 2641-50.

25.   Lambert, I. H., Regulation of the cellular content of the organic osmolyte taurine in mammalian cells. *Neurochem Res* **2004,** 29, (1), 27-63.

26.   El Agouza, I. M.; Eissa, S. S.; El Houseini, M. M.; El-Nashar, D. E.; Abd El Hameed, O. M., Taurine: a novel tumor marker for enhanced detection of breast cancer among female patients. *Angiogenesis* **2011,** 14, (3), 321-30.

27.   Li, M.; Wang, B.; Zhang, M.; Rantalainen, M.; Wang, S.; Zhou, H.; Zhang, Y.; Shen, J.; Pang, X.; Wei, H.; Chen, Y.; Lu, H.; Zuo, J.; Su, M.; Qiu, Y.; Jia, W.; Xiao, C.; Smith, L. M.; Yang, S.; Holmes, E.; Tang, H.; Zhao, G.; Nicholson, J. K.; Li, L.; Zhao, L., Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci U S A* **2008,** 105, (6), 2117-22.

28.   Pendyala, L.; Velagapudi, S.; Toth, K.; Zdanowicz, J.; Glaves, D.; Slocum, H.; Perez, R.; Huben, R.; Creaven, P. J.; Raghavan, D., Translational studies of glutathione in bladder cancer cell lines and human specimens. *Clin Cancer Res* **1997,** 3, (5), 793-8.

29.	Kumps, A.; Duez, P.; Mardens, Y., Metabolic, nutritional, iatrogenic, and artifactual sources of urinary organic acids: a comprehensive table. *Clin Chem* **2002,** 48, (5), 708-17.

30.	Nicholson, J. K.; Holmes, E.; Wilson, I. D., Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol* **2005,** 3, (5), 431-8.

31.	Zhu, Q.; Gao, R.; Wu, W.; Qin, H., The role of gut microbiota in the pathogenesis of colorectal cancer. *Tumour Biol* **2013,** 34, (3), 1285-300.

32.	Le Bot, N., Obesity-associated gut microbiota induce liver cancer. *Nat Cell Biol* **2013,** 15, (8), 894.