Published in final edited form as:

Urol Oncol. 2011 ; 29(5): 551–557. doi:10.1016/j.urolonc.2011.05.013.

Urine metabolomics for kidney cancer detection and biomarker discovery

Sheila Ganti^{1,2} and Robert H. Weiss^{1,2,3,4}

¹Division of Nephrology, Dept. of Internal Medicine, University of California, Davis, CA, USA

²MCIP Graduate Group, University of California, Davis, CA, USA

³Cancer Center, University of California, Davis, CA, USA

⁴Medical Service, Sacramento VA Medical Center, Sacramento, CA, USA

Abstract

Renal cell carcinoma (RCC) is one of the few human cancers whose incidence is increasing. The disease regularly progresses asymptotically and is frequently metastatic upon presentation, thereby necessitating the development of an early method of detection. A metabolomic approach for biomarker detection using urine as a biofluid is appropriate since the tumor is located in close proximity to the urinary space. By comparing the composition of urine from individuals with RCC to control individuals, differences in metabolite composition of this biofluid can be identified, and this data can be utilized to create a clinically applicable, and possibly bedside, assay. Recent studies have shown that sample handling and processing greatly influences the variability seen in the urinary metabolome of both cancer and control patients. Once a standard method of collection is developed, identifying metabolic derangements associated with RCC will also lead to the investigation of novel targets for therapeutic intervention. The objective of this review is to discuss existing methods for sample collection, processing, data analysis, and recent findings in this emerging field.

Keywords

Metabolomics; Renal Cell Carcinoma; Biomarker; Urine; Kidney

Introduction

Of the various “omicses”, metabolomics is a relatively new approach to analyze the composition of a biofluid (plasma or urine) or tissue whereby all small molecule metabolites less than 1500 daltons are examined [1]. When analyzed and interpreted, this collection of small molecules can yield a unique signature that can be exploited for diagnosis as well as to determine subtle and gross metabolic differences in a normal compared to a diseased state. Unlike the older technologies of genomics, transcriptomics, or proteomics, metabolomics utilizes a relatively small number of metabolites that can be identified, simplifying data

Please address editorial correspondence and reprint requests to: Dr. Robert H. Weiss, Division of Nephrology, Department of Internal Medicine, Genome and Biomedical Sciences Building, Room 6312, University of California, Davis, CA. 95616, rhweiss@ucdavis.edu, Tel: 530-752-4010; FAX: 530-752-3791.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

analysis. Furthermore, the modification of substrates seen with transcriptomics (pre-translational) and proteomics (post-translational) does not occur with metabolites, indicating a high potential for a metabolomics approach for biomarker discovery [2].

In addition to biomarker discovery, the metabolomics approach can identify novel druggable targets, because understanding the metabolic derangements and altered biochemical pathways that occur with disease progression can provide insight into possible therapies for that disease by identifying inhibitors of altered pathways among new as well as already existing drugs. Thus metabolomics lends itself to a two-pronged approach to the clinical problem by addressing both markers for disease as well as suggesting new therapeutic approaches.

The study of metabolomics relies heavily on the use of complex and expensive technology for separation and high-resolution identification of metabolites within a biofluid or tissue, such as gas chromatography (GC) or liquid chromatography (LC) coupled to mass spectrometry, or nuclear magnetic resonance (NMR). We and others have found that slight variations in sample handling or sample preparation cause slightly different data to be collected, thus standardization of collection and analysis of samples is of the utmost importance.

RCC is highly amenable to a metabolomics approach to biomarker discovery

Renal cell carcinoma (RCC) is a disease that is well suited to a metabolomic type of analysis, since its principal biofluid, urine, is obviously in close proximity to the disease of interest. RCC is one of the few cancers whose incidence is rising with 58,240 new cases projected in 2010 [3], and, since RCC regularly progresses asymptotically, diagnosis of this disease frequently occurs incidentally when patients are examined for diseases unrelated to cancer, such as during the work-up of acute renal failure. In one third of these diagnosed cases, the tumors have already metastasized and prognosis is dismal. Even with the advances being made in targeting cancer treatment, the progression free survival time for metastatic RCC is only 11 months [4]. These data emphasize the urgent need for development of a method for early detection such as an urinary metabolite biomarker.

The cells that make up a cancerous tumor are highly metabolically active, producing a unique intracellular milieu that can in theory be readily distinguished from normal cells. The majority of RCC tumors occur within the tubule epithelium and result in aberrant secretion of metabolites into the tubular lumen (urinary space). This process can be exploited to separate the urinary metabolome of normal patients from patients with RCC. In this review, we will discuss both experimental setup and analytical and statistical methods in common use for urinary biomarker identification in RCC. We will conclude with current findings from our laboratory relevant to RCC biomarker discovery.

Sample Handling and Preparation

An outline of our approach using metabolomics for biomarker discovery is shown (Fig. 1). Urine samples, unlike tissue samples, must be collected and handled in a uniform manner to ensure consistency, mainly because there is as yet no consensus standard for urine collection and preservation in metabolomics. In the first study of its type, we performed HILIC-MS based metabolomic analysis and compared 26 urine samples collected at a California site (15 control and 11 RCC) with 24 samples collected from a site in Texas (12 control and 12 RCC). Even though all chemical analyses were conducted in the same laboratory, we found

that the site of collection radically increased the variability seen in the data set [5] underscoring the need for a standard operating procedure with respect to sample handling.

Urea is a highly abundant metabolite found in the urine that can obscure metabolite identification if metabolites of interest lie “under” this huge peak [6]. For this reason, some investigators add urease to their urine samples before analysis to remove urea. If the metabolite of interest elutes from the column at the same time as urea, the abundance of urea will mask the presence of this metabolite and could also “clog up” the analytical machinery. However, we have shown that urease treatment interferes with the detection of certain metabolites including some tricarboxylic acid cycle intermediates [7]. Instead of treating samples with urease, using a reverse phase column (in the case of LC) or optimizing other chromatography parameters will prevent urea from masking other metabolites [6].

The presence of bacteria in urine samples, a not uncommon event, may result in the production of metabolites from these exogenous organisms and thus unrelated to the disease being studied; this can lead to false positive biomarker identification. Thus, sodium azide has been added to urine samples to prevent bacterial growth [8]. Investigators who have used sodium azide claim that metabolite stability is enhanced in these samples. As far as urine storage conditions are concerned, there have been some studies on this issue; a recent investigation has shown that maximum stability is achieved when samples are stored at -80°C [8].

Analytical Techniques

Gas Chromatography (GC)

Gas chromatography separates analytes (metabolites) by exploiting their volatility. In addition to being easily volatilized, metabolites analyzed with this technique must be thermally stable and non-polar, properties which limit the number of metabolites that can be identified by this technique. However, a process called derivatization can increase the volatility of the analytes so that they are more amenable to GC analysis [6]. While GC-MS can identify fewer metabolites than liquid chromatography (see below) because of the volatility and polarity restrictions, many commercially available libraries exist for metabolite identification using the GC platform based on both retention time in the column and fragmentation pattern produced by the mass spectrometer, making GC-MS a highly attractive analytical system. In addition, given the presence of reliable libraries, GC-MS is frequently used for a targeted metabolite approach whereby a single metabolite or a subset of known metabolites are assayed [6].

Liquid Chromatography (LC)

LC offers a much broader range of analyte detection than GC depending on what type of column is used, and in this technique, separation of metabolites is achieved through exploiting differences in polarity, ionic charge and molecular size. Lipids, nucleic acids, and peptides can all easily be identified with this approach. Separation is also enhanced by varying the mobile phase, and a gradient elution system using a reverse phase C18 column is most frequently used. LC coupled to mass spectrometry offers a sensitive detection system, though a UV/Vis detector can also be used. Like GC, LC also is optimized with the use of internal standards which provide a “grid” to determine retention time for all metabolites. Even with column age, the retention time of the internal standards can shift. Unlike GC, however, the commercially available libraries for LC-MS are poorly developed making specific metabolite characterization using the latter technology more difficult; this is the main drawback of LC-MS as compared to GC-MS [6].

Nuclear Magnetic Resonance (NMR)

Metabolite identification and quantification relied heavily on NMR technologies prior to 2005. However, since then, MS-based techniques have become more common as that technology has improved [6]. Compared to MS, NMR sample preparation is easier, the results are highly reproducible, and structural information is superior, but the significant disadvantage in biomarker discovery is that more starting material is required, a non-trivial issue given the low abundance metabolites that can occur in biofluids. In addition, fewer metabolites can be identified by NMR as compared to MS in a single analysis, indicating suitability for targeted approaches over global screening of metabolites. A consideration when using NMR-based strategies relates to differences in pH between samples which may affect data output [9]. Since normal urine pH is highly variable, NMR may for this reason not be the ideal analytical tool to use for the metabolomics approach to urinary biomarker discovery.

Data Analysis

Urinary Concentration

Differences in urinary concentration, a universal occurrence when dealing with this biofluid, will skew the data and thus will result in erroneous conclusions if ignored. For example, if the same *volume* of urine is analyzed, the more concentrated urines will yield metabolites at higher concentrations as compared to a more dilute sample, yet this will not be a true representation of absolute (i.e. moles of metabolite) metabolite abundance. Several methods of correction can be employed to deal with this important issue. Most commonly, investigators utilize the fact that creatinine is secreted in the urine in a relatively constant mass amount (as a function of patient muscle mass), and thus normalization of all metabolite concentrations within a sample to urinary creatinine concentration in the same sample will yield a relative molar abundance of each metabolite [10]. This method makes several assumptions: (1) the individual is not experiencing acute renal failure, (2) the absolute mass amount of creatinine secreted is relatively constant in all individuals, and (3) urinary creatinine will in fact be identifiable by the analytical method chosen. Normalizing by creatinine is not quantitative, but relative metabolite abundance can be calculated using this method [10]. Alternatively, a “total ion count” method of normalization can be used. For this method, the peak intensities for all structurally identified metabolites within a sample are summed into a single total count. All metabolite peaks within that sample are divided by this total count in the same sample [11]. This method of sample normalization is preferred when samples originate from patients who have chronic kidney disease. The third method is to normalize by sample osmolarity such that all identified metabolites within a single sample are divided by that sample’s osmolarity [12]. This requires osmolarity measurements to be taken for each sample; we have found that this method results in similar values to the creatinine normalization method (unpublished data), and, due to the ease of determining sample osmolarity in the laboratory even with small samples volumes, this is the method we most commonly use. Of all the methods, creatinine and total count normalizations are most commonly utilized [6].

Statistical Analysis

Once samples are normalized to account for differences in urinary concentration, the data are subjected to statistical processing. Due to the fact that there exist a sizable number of low abundance metabolites which are not always measured by the analytical system used, it frequently happens that a dataset is incomplete. If a metabolite is present at a value that less than the signal to noise ratio set by the investigator, it will not be detected in the sample. Furthermore, not every metabolite will be identified in every sample resulting in missing concentration values (or values which equal zero), a phenomenon which precludes statistical

transformations. A non-zero value corresponding to a value half the lowest detected peak or the lowest detected peak can be imputed for these missing values. To account for spurious metabolites or those related to drug metabolism, a common technique utilized is to require any “true” analyzed metabolite (i.e. not a drug metabolite) to be detected in over half of the samples analyzed.

After data processing described above, a method for statistical analysis is chosen, generally in consultation with a statistician well versed in omics analysis. While a complete description of the statistical treatment of metabolomic data is beyond the scope of this review, suffice it to say that the type of test chosen depends on the study design and the variables involved. For example, a study examining the differences between two groups with matched patients should utilize an ANOVA or Wilcoxon to determine which metabolites contribute to differences seen between groups (Table). An untargeted or global metabolomics approach involves testing multiple hypotheses thereby necessitating a multiple testing statistical correction such as Benjamin-Hochberg’s false discovery rate. Utilization of such corrections reduces the false positive discovery rate [12].

Validation

Technical Validation

A rigorous metabolomics experiment requires analyzing a large number of samples, which may take days to weeks of machine time depending on the number of samples and the availability of equipment, which in many institutions is shared among several investigators. During the run time, the machine output can “drift”, causing a change in retention time or in peak baseline leading to possible peak misidentification and consequent inaccurate metabolite quantification. To account for these discrepancies, several different approaches can be taken. First, investigators can rerun a random subset of samples and compare the output to the original run [11, 12]. This method does not require true metabolite identification from the MS output, it only requires determination of peak intensity for comparisons between runs [12]. The results of this type of analysis will demonstrate technical (i.e. analytical equipment) reproducibility and is thus a critical step when identifying potential biomarkers of disease.

Alternatively, instead of assaying a subset of samples and comparing retention time and peak intensity, all samples can be assayed again for a subset of metabolites. Trends between the first and second run can then be compared [13]. This represents a form of targeted metabolomics and can also serve as a means to test a hypothesis regarding this metabolite set. Targeted metabolomics is an appropriate approach when a group of known metabolites are being identified [1]. In this manner researchers can extend findings from global metabolomics assays while at the same time quantifying those metabolites.

A third way to validate findings is by using the test set/training set approach. This method necessitates a large sample size which is divided into a training set and a test set [14]. The training set is used to establish the metabolomic pattern, and with this data a statistical test for sample separation is employed to distinguish between cancer and control. The samples comprising the (blinded) test set are then used to confirm success of the separation criteria gleaned from the training set. Larger within-group variability will prevent proper separation of samples and will result in greater metabolite overlap between the cancer samples and the control.

A typical metabolomics experiment in which urine is used to identify potential biomarkers for RCC, i.e. metabolites which can separate RCC from other diseases of the urinary tract and from non-kidney cancers, is shown (Fig. 2). The goal of this global metabolomics

approach is to identify a single metabolite, or a group of metabolites, that best effect that separation and can thus be ultimately amenable to clinical assay. This cannot be done without validating the findings and demonstrating reproducibility by either assaying a subset of samples again, by a targeted approach, and/or by assessing the validity of the separation criteria.

Biological validation

The global metabolomics approach results in the identification of a subset of significantly different metabolites. While biomarker discovery is a goal of these experiments, as mentioned earlier this approach also has therapeutic ramifications. Thus, in order to determine if the identified diagnostic metabolites are causal for the disease or simply markers, *in vitro* and/or *in vivo* validation could be undertaken. These types of experiments may also provide information regarding specificity of the potential biomarkers by using a variety of cancer cell lines. *In vitro* experiments take advantage of cell culture models of RCC. Specific metabolites identified as described in previous sections are synthesized or purchased and added to cell culture medium, and cancer-relevant biological phenomena (e.g. proliferation, apoptosis, or cytotoxicity) are measured [12]. Metabolites that promote proliferation or protect cells from apoptotic or lytic death are thought to be biologically significant in oncology research, thus those metabolites that inhibit proliferation or promote cell death may represent pathogenic metabolites and thereby targets for future cancer intervention. Once it is determined that a metabolite has a biological role, the mechanism of action can be further examined more closely by looking at regulatory enzymes or signaling pathways by standard molecular techniques.

Animal models of disease can provide a more comprehensive study of cancer metabolomics, given that they frequently come from a similar genetic background and because a xenograft experiment introduces identical cancer cells. Examining serum and tissue combined with urine will supply the most complete information for analysis in this model, keeping in mind the artifactual urinary results that might arise if some of the mice develop renal failure. After the tumor grows to a sufficient size, the animals are sacrificed and all three biological samples are collected and subsequently analyzed using a global metabolomics approach. This type of analysis can distinguish between metabolites secreted by the tumor into the urine versus metabolites secreted systemically in response to the tumor. As a first assumption, metabolites present higher in the tumor tissue represent the cancer itself while metabolites present in higher concentration within the serum represent a systemic response.

Pathway Analysis

Once all the significant metabolites have been identified by global metabolomics, metabolomic pathways can be assembled based on those metabolites which are altered in the human or animal model. There exist several publically available pathway databases (KEGG, BioCyc, SMPDB) which group metabolites into the metabolic pathways based on extant published literature (Fig. 3). Some metabolites appear in multiple pathways, which can complicate their pathway analyses. In its ideal practice, which is still evolving in the case of metabolomics data, pathway analysis provides information about whole metabolic changes rather than on the single metabolite level [12] and can lead to novel markers or targets not found in the original dataset. Furthermore, placing significantly altered metabolites into a pathway may point to a single enzyme being dysregulated, when, for example, many metabolites downstream of the enzyme are down regulated and others upstream metabolites are up regulated.

Metabolomics and Renal Cell Carcinoma

Due to its commonly asymptomatic presentation, RCC is frequently diagnosed incidentally when an affected patient is seen in the general medicine or renal clinic for unrelated complaints. One-third of these tumors are already metastatic at presentation, greatly limiting the likelihood of successful treatment [4]. The current focus of metabolomics in this field is to identify novel urinary biomarkers for early diagnosis and to determine possible perturbations in metabolic pathways with the hopes of identifying novel targets for therapeutic intervention. Using a global metabolomics approach, we have identified several metabolites that may separate cancer patients from matched control patients; these metabolites are in the process of being validated. These metabolites generally function within the amino acid metabolism pathways and energy utilization pathways. Three of the most significant metabolites were assayed for cell proliferative effects *in vitro* using three cancer cell lines and one normal cell line. Two metabolites, α -ketoglutarate and quinolinate, were shown to increase proliferation in two of the three cancer cell lines. Both of these metabolites were also elevated in cancer urine compared to control suggesting increased secretion by the tumor [12].

A proteomic study examining tissues from human patients with RCC from our laboratory found a similar pattern of metabolic pathways affected by RCC [15]. In both studies, enzymes regulating energy production and utilization such as amino acid metabolism and glycolysis are disrupted in cancer [15]. Others have shown that peroxiredoxin is elevated in the tissue of individuals with clear cell RCC, conferring an increased ability to handle oxidative stress [16]. This is also one potential role that elevated quinolinate may serve since it is an intermediate in NAD⁺ or NADP⁺ synthesis [12].

When assaying changes in urinary metabolite composition, it is important determine the source of the differences. When metabolites are found in higher concentrations in cancer compared to control, this may indicate increased tumor secretion. An alternative hypothesis is that increased urinary metabolite concentration is a systemic response to the tumor itself. Development of xenograft models in which serum, urine, and tissue are all analyzed can address these questions.

Conclusion

Urine is a abundant biofluid that can be readily obtained by non-invasive means. These two attributes make urine a practical choice for developing a method of diagnosis for renal cell carcinoma. Since the cells making up the tumor usually occur within the tubule epithelium which does not block the secretion of small molecular weight metabolites, these compounds arising from the tumor can easily cross from the cells into the urinary space making this fluid ideal for metabolomic discovery of RCC biomarkers.

Human urine samples are highly heterogeneous, a property that increases intra-group variation and masks inter-group variation. As a result, a large sample size is required to clearly separate cancer from control. Previous work has demonstrated that sample handling prior to analysis greatly influences variation emphasizing the need for a standard method of operation [5]. Future work in this field will focus on identifying metabolites that distinguish between RCC and non RCC urine, identifying pathways perturbed in this disease, and understanding disease progression as determined either by grade or staging information.

Acknowledgments

This work was supported by NIH grants 5U01CA86402 (Early Detection Research Network), 1R01CA135401-01A1, and 1R01DK082690-01A1, and the Medical Service of the US Department of Veterans' Affairs (all to R.H.W.).

References

1. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner H-P. Targeted Metabolomics for Biomarker Discovery. *Angew Chem Int Ed*. 2010; 49:5426–5445.
2. German JB, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics*. 2005; 1:3–9. [PubMed: 16680201]
3. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin*. 2010; 60:277–300. [PubMed: 20610543]
4. Cella D, Cappelleri JC, Bushmakina A, Charbonneau C, Li JZ, Kim ST, Isan C, Michaelson MD, Motzer RJ. Quality of life predicts progression-free survival in patients with metastatic renal cell carcinoma treated with sunitinib versus interferon alfa. *J Oncology Prac*. 2009; 5:66–70.
5. Kim K, Aronov P, Zakharkin SO, Anderson D, Perroud B, Thompson IM, Weiss RH. Urine Metabolomics Analysis for Kidney Cancer Detection and Biomarker Discovery. *Mol Cell Proteomics*. 2009; 8:558–570. [PubMed: 19008263]
6. Ryana D, Robards K, Prenzler PD, Kendall M. Recent and potential developments in the analysis of urine: A review. *Analytica Chimica Acta*. 2011; 684:17–29.
7. Kind T, Tolstikov V, Fiehn O, Weiss RH. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal Biochem*. 2007; 363:185–95. [PubMed: 17316536]
8. Saude EJ, Sykes BD. Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics*. 2007; 3:19–27.
9. Schripsema J. Application of NMR in Plant Metabolomics: Techniques, Problems and Prospects. *Phytochem Anal*. 2009; 21:14–21. [PubMed: 19904731]
10. Zuppi C, Messana I, Forni F, Rossi C, Pennacchietti L, Ferrari F, Giardina B. ¹H NMR spectra of normal urines” Reference ranges of the major metabolites. *Clin Chim Acta*. 1997; 265:85–97. [PubMed: 9352132]
11. Taylor SL, Ganti S, Bukanov NO, Chapman A, Fiehn O, Osier M, Kim K, Weiss RH. A metabolomics approach using juvenile cystic mice to identify urinary biomarkers and altered pathways in polycystic kidney disease. *Am J Physiol Renal Physiol*. 2010; 298:F909–F922. [PubMed: 20130118]
12. Kim K, Taylor SL, Ganti S, Guo L, Osier M, Weiss RH. Urine metabolomic analysis identifies potential biomarkers and pathogenic pathways in kidney cancer. *OMICS*. 2011 epub ahead of print.
13. Kim KB, Yang JY, Kwack SJ, Park KL, Kim HS, Ryu do H, Kim YJ, Hwang GS, Lee BM. Toxicometabolomics of urinary biomarkers for human gastric cancer in a mouse model. *J Toxicol Environ Health A*. 2010; 73:1420–1430. [PubMed: 20954069]
14. Schaefer ML, Wongravee K, Holmboe ME, Heinrich NM, Dixon SJ, Zeskind JE, Kulaga HM, Brereton RG, Reed RR, Trevejo JM. Mouse urinary biomarkers provide signatures of maturation, diet, stress level, and diurnal rhythm. *Chem Senses*. 2010; 35:459–471. [PubMed: 20418335]
15. Perroud B, Ishimaru T, Borowsky AD, Weiss RH. Grade-dependent proteomics characterization of kidney cancer. *Mol Cell Proteomics*. 2009; 8:971–985. [PubMed: 19164279]
16. Sun CY, Zang YC, San YX, Sun W, Zhang L. Proteomic analysis of clear cell renal cell carcinoma. *Saudi Med J*. 2010; 31:525–532. [PubMed: 20464042]

Figure 1. A general approach to biomarker discovery using urine metabolomics

Urine samples are taken from patients, preferably by the same clinical coordinator, and handled uniformly. Subsequently, samples are analyzed by one or several of a variety of analytical techniques, and output data is subjected to statistical analysis to find groupings of metabolite sets as a function of disease or non-diseased patients. Finally, an assay, preferably applicable to a bedside test, is developed to be able to identify one or several metabolites which embody this disease signature.

Figure 2. A typical metabolomics experimental setup used in our laboratory for biomarker discovery

Once urine samples are obtained, samples undergo LC-MS and GC-MS (as well as sometimes HILIC-MS) analysis. After true peak identification by available software, metabolite structural elucidation is performed prior to biological and test-set validation.

Figure 3. Pathway analysis utilizing the KEGG dataset

Once significantly altered metabolites are identified, they can be placed on one of several available biochemical pathway diagrams such as, in this case, nicotinate and nicotinamide metabolism. In this manner, other metabolites within these pathways can be targeted for further testing, and, should they be available, inhibitors of the pathway can be considered for clinical trials in the disease of interest. (reproduced from Urine Metabolomic Analysis Identifies Potential Biomarkers and Pathogenic Pathways in Kidney Cancer. Kim K, Taylor SL, Ganti S, Guo L, Osier MV, Weiss RH. OMICS. 2011 Feb 24.)

Table 1

Key components of statistical analysis of metabolomic data

Component	Objective	Statistical method
Experimental Approach	Determine the hypothesis to test. This will either utilize a global or targeted approach	Randomization of sample injection
Data pre-processing	Remove systematic variation caused by urine osmolality.	Normalization
	Rescale data to satisfy underlying assumptions of statistical tests.	Transformation
Inference test	Identify metabolites present at different amounts between groups	Hypothesis-testing methods such as t-test, ANOVA, Wilcoxon, regression etc.
Classification	Classify patients to the pre-defined classes (known disease status/genotype).	Supervised methods such as PLS, PLS-DA, and machine learning methods coupled with feature selection techniques
	Discover new clinical sub-classes.	Unsupervised methods such as PCA or cluster analysis
Validation	Ascertain reproducibility of the results obtained. Determine utility of metabolites as biomarkers for disease. This must be done using a training set/test set approach	Cross-validation or independent validation