

# Urine Metabolomics Analysis for Kidney Cancer Detection and Biomarker Discovery\*<sup>§</sup>

Kyoungmi Kim<sup>‡</sup>, Pavel Aronov<sup>§</sup>, Stanislav O. Zakharkin<sup>¶</sup>, Danielle Anderson<sup>§</sup>, Bertrand Perroud<sup>||</sup>, Ian M. Thompson<sup>\*\*</sup>, and Robert H. Weiss<sup>§†‡§§</sup>

Renal cell carcinoma (RCC) accounts for 11,000 deaths per year in the United States. When detected early, generally serendipitously by imaging conducted for other reasons, long term survival is generally excellent. When detected with symptoms, prognosis is poor. Under these circumstances, a screening biomarker has the potential for substantial public health benefit. The purpose of this study was to evaluate the utility of urine metabolomics analysis for metabolomic profiling, identification of biomarkers, and ultimately for devising a urine screening test for RCC. Fifty urine samples were obtained from RCC and control patients from two institutions, and in a separate study, urine samples were taken from 13 normal individuals. Hydrophilic interaction chromatography-mass spectrometry was performed to identify small molecule metabolites present in each sample. Cluster analysis, principal components analysis, linear discriminant analysis, differential analysis, and variance component analysis were used to analyze the data. Previous work is extended to confirm the effectiveness of urine metabolomics analysis using a larger and more diverse patient cohort. It is now shown that the utility of this technique is dependent on the site of urine collection and that there exist substantial sources of variation of the urinary metabolomic profile, although group variation is sufficient to yield viable biomarkers. Surprisingly there is a small degree of variation in the urinary metabolomic profile in normal patients due to time since the last meal, and there is little difference in the urinary metabolomic profile in a cohort of pre- and postnephrectomy (partial or radical) renal cell carcinoma patients, suggesting that metabolic changes associated with RCC persist after removal of the primary tumor. After further investigations relating to the discovery and identity of individual biomarkers and attenuation of residual sources of variation, our work shows that urine metabolomics analysis has potential to lead to a diagnostic assay for RCC. *Molecular & Cellular Proteomics* 8:558–570, 2009.

From the <sup>‡</sup>Division of Biostatistics, Department of Public Health Sciences, <sup>§</sup>Division of Nephrology, Department of Internal Medicine, and <sup>||</sup>Genome Center, University of California, Davis, California 95616, <sup>¶</sup>Department of Research and Development, Solae, St. Louis, Missouri 63188, <sup>\*\*</sup>Department of Urology, University of Texas Health Science Center, San Antonio, Texas 78229, and <sup>††</sup>Medical Service, Sacramento Veterans Affairs Medical Center, Sacramento, California 95655

Received, April 14, 2008, and in revised form, October 14, 2008  
Published, MCP Papers in Press, November 13, 2008, DOI 10.1074/mcp.M800165-MCP200

The study of all endogenously produced metabolites, known as metabolomics (or metabonomics), is the youngest of the omics sciences. It is becoming increasingly clear that, of all of the omics techniques, metabolomics has the greatest potential for biomarker discovery because this technique defines the signature of the actual processes that are occurring within the body rather than examining compounds (such as untranscribed DNA or pre- or post-translationally modified proteins) that may be superfluous to these processes (1). In addition, there is a relatively small number of metabolites to examine (with the notable exception of plants, which produce a plethora of secondary metabolites) as compared with genes, transcripts, and proteins in their respective omics fields, and therefore the data germane to metabolomics are more easily handled and analyzed. Proponents of metabolomics provide convincing justification that this technique offers more immediate translational benefit than the other omics fields (1, 2).

The use of metabolomics through examination of patient urine is in theory an ideal means to study diseases of the urinary tract given that low molecular weight compounds (such as small molecule metabolites) are freely filtered into the urine. In addition, obtaining this biofluid can be done quickly, easily, and in a non-invasive manner in the clinic. Thus, urine metabolomics has potential utility in metabolic profiling as well as for biomarker discovery for cancers of the urinary tract (3). Once urinary biomarkers are discovered and validated, they could conceivably be used for prognosis as well as to predict response to targeted therapies as obtaining urine is always more feasible than gaining access to tumor tissue.

There have been several studies looking at single compounds in the urine as markers of non-malignant renal disease. These compounds include *N*-acetyl- $\beta$ -D-glucosaminidase, neutrophil gelatinase-associated lipocalin, human kidney injury molecule-1, and interleukin-18 for kidney injury (4, 5); one of the same molecules, human kidney injury molecule-1, has also been proposed as a marker for RCC<sup>1</sup> (6). The metabolite glucose, when present in urine, has been

<sup>1</sup> The abbreviations used are: RCC, renal cell carcinoma; HILIC, hydrophilic interaction chromatography; PCA, principal component analysis; LDA, linear discriminant analysis; FDR, false discovery rate; PC, principal component; TX, University of Texas Health Science Center at San Antonio; CA, University of California, Davis; AU, approximately unbiased; BMI, body mass index.

utilized for centuries as a biomarker of diabetes. Studies in our laboratory have focused on examining the entire metabolome to determine whether a pattern of such metabolites in the urine (including known as well as unknown) can result in the diagnosis of RCC. Our pilot study, using relatively few samples, showed that these two groups can be clearly differentiated, demonstrating for the first time that this disease is amenable to such a technique (3).

Although, based on this and our previous work, urine metabolomics appears to be a powerful technique for evaluation of RCC, there exist challenges associated with this technique that will need to be addressed prior to its general clinical applicability. In the current study, we identified some of these issues and confirmed our pilot study on the utility of urine metabolomics in RCC. Two sets of urine samples from patients with clear cell RCC and control patients were obtained from two separate institutions. A separate set of urines was obtained from a normal group of individuals to determine the influence of mealtimes on changes of metabolomic profiles. Statistical analysis of mass spectrometric data confirms that cancer and control patients can be segregated using a larger cohort of samples; however, the urinary metabolic profile is dependent on institutional site of sample collection. Additionally when sources of variation in the data sets were evaluated, it was found that only a small portion of measured urinary metabolites mainly contribute to the disease variability between cancer and control groups; these are the metabolites to target for future identification and biomarker discovery. Surprisingly urine samples collected from patients several months postnephrectomy were not separable from urine samples of pre-nephrectomy patients, suggesting that metabolic abnormalities of the disease were not altered in the short term by removal of the tumor. In addition, there was little variation in the urinary metabolome due to the time since the last meal, suggesting that fasting samples are not necessary for this type of analysis. These data confirm the utility of urine metabolomics analysis for RCC and inject a note of caution about issues that complicate interpretation of the data and that will need to be worked out before the technique is clinically useful.

## EXPERIMENTAL PROCEDURES

### *Sample Procurement*

After approval by the appropriate Institutional Review Boards, random (in the case of control) or fasting preoperative or 2–3-month postoperative (in the case of experimental) urine samples from clear cell RCC patients of various ages and genders and at various stages and grades were obtained from the urology clinics at the University of California, Davis (CA) Medical Center or the University of Texas Health Science Center at San Antonio (TX). Control patients from the University of California, Davis data set were patients seen in the urology clinic but who did not have known kidney cancer or renal insufficiency. At TX, two samples were obtained from RCC patients: samples prior to nephrectomy (treated as preoperative cases) and after nephrectomy in the same patients (treated as postoperative controls). All experimental urine samples were obtained prior to chemotherapy,

radiation, or (in the case of all University of California, Davis and preoperative TX samples) nephrectomy. In a separate study, urines were taken from a cohort of 13 normal female individuals at different times of the day and at variable times after meals. In all collection procedures, voided urine was collected in a urine specimen container and frozen at  $-80^{\circ}\text{C}$  within 2–6 h of collection. All urines were kept on ice until frozen.

### *Hydrophilic Interaction Chromatography (HILIC)-MS Analysis*

HILIC-MS analysis was performed as described previously (3). Briefly neat urine was mixed with an equal volume of acetonitrile at room temperature. All samples were spun for 5 min at 13,000 rpm prior to injection for particulate matter removal. Liquid chromatography was performed using acetonitrile (LC-MS grade, J. T. Baker Inc.) (A) and 13 mM ammonium acetate buffer (pH 9.1 for HILIC, adjusted by ammonium hydroxide) (B) as the mobile phase at flow rates of 0.5 ml/min at  $40^{\circ}\text{C}$ . Ammonium acetate (extra pure, EMD Biosciences), ammonium hydroxide solution, and acetic acid (glacial, J. T. Baker Inc.) were purchased from VWR. Water was used purified by a Milli-Q Gradient A 10 system (Millipore).

HILIC separations were performed with a Surveyor HPLC module (Thermo Fisher) and an Aphera NH<sub>2</sub> polymer ( $150 \times 2$  mm,  $5\text{-}\mu\text{m}$  particle size; Astec, Whippany, NJ). After a 5-min isocratic run at 20% B, a gradient to 35% B was concluded at 20 min, and then a gradient to 90% B was completed at 30 min. The injection volume was set to  $3\ \mu\text{l}$ . HPLC columns were connected to the electrospray interface of the Finnigan LTQ (Thermo Fisher) linear ion trap mass spectrometer without splitting. Nitrogen sheath gas pressure was set to 7 bars at a flow rate of 2–3 liters/min. Spray voltage was set to 5 kV. The temperature of the heated transfer capillary was maintained at  $350^{\circ}\text{C}$ . Full-scan mass spectra were acquired from 80 to 800 daltons and unit mass resolution in both modes, positive and negative.

### *Raw Data Processing and Deisotoping*

The peak finding in individual chromatograms and subsequent peak alignment across all chromatograms was performed using MarkerView 1.1 software (Applied Biosystems, Foster City, CA). Prior to data processing the files containing the Thermo LTQ chromatographic data were converted from the original Xcalibur (\*.raw) format into netCDF (\*.cdf) format using the XConverter program (Thermo Fisher) to ensure format compatibility with MarkerView 1.1. Peak finding options were set as follows: subtraction offset, 10 scans; subtraction multiplication factor, 1.3; noise threshold, 3; minimum spectral peak width, 0.5 amu, minimum retention time peak width, two scans; and maximum retention time width, 1000 scans. Peak alignment options were set as follows: retention time tolerance, 0.5 min; mass tolerance, 0.8 min; and maximum number of peaks, 5000. If peaks were found in fewer than five of the samples (10% of all samples), this feature was automatically discarded using a filter setting of MarkerView. Peak area integration was performed using raw data. No data normalization was implemented.

The data that constitute retention time, mass, and peak areas of detected and aligned peaks were exported from MarkerView into comma-separated variable (\*.csv) format. Overall the software detected 1929 aligned spectral features in a cancer *versus* control data set and 2593 aligned features in a control data set. First isotopic peaks (+1 amu), sodium (+22 amu), ammonium (+17 amu), and potassium (+38 amu) adducts were detected using a MatLab (MatWorks, Natick, MA) script that detects and marks mass differences listed above within a  $\pm 0.05$  min retention time window and a  $\pm 0.25$  amu mass window (supplemental Fig. 1). The marked ions were curated manually, and a higher mass counterpart corresponding to an isotopic peak or adduct was removed resulting in a reduction from

1929 to 1766 features in the cancer *versus* controls data set and from 2593 to 2333 features in the normal patient data set. The resulting data were directed for statistical analysis.

### Statistical Analysis

Statistical analysis was programmed with the R 2.6.2 language and environment (The R Foundation for Statistical Computing, Auckland University, Auckland, New Zealand) and SAS 9.1 (SAS Institute Inc., Cary, NC). We observed that the overall intensity of the metabolomic spectrum was slightly higher in the RCC samples than the control samples. Hence sample mean normalization was undertaken to facilitate comparison between the two disease groups before transformation and analysis. Statistical analysis methods that we had planned to apply in this study, such as dimension reduction, clustering, and differential analysis, are based on the assumption that variability of measurements does not depend on the measurement levels. However, as for other high throughput data, the variance in our data tended to rise with the intensity. Hence we applied log (base 2) transformation to the metabolite peak intensities prior to analysis in all models to stabilize the variance.

**Identifying the Groups of Like Samples**—Hierarchical cluster analysis was performed to join together or split off individual samples based upon a measure of their similarity/dissimilarity. The process starts with each sample in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step until there is just a single cluster. At each stage, distances between clusters were computed by Ward's linkage algorithm (7). This method uses an analysis of variance approach and minimizes the sum of squares of any two clusters that can be formed at each step. To assess the uncertainty in hierarchical cluster analysis, clustering was performed on 1000 bootstrap replicates, and consensus dendrograms were constructed using the R package pvclust (8). The measures of stability for each node were calculated using both normal bootstrap resampling (bootstrap probability  $p$  value) and multiscale bootstrap resampling (approximately unbiased (AU)  $p$  value). The dendrograms were examined for separation or clusters relating to the two groups of disease, RCC and control or pre- and postoperative nephrectomy, and variation related to site differences between CA and TX urine samples.

**Principal Component Analysis and Linear Discriminant Analysis**—Principal component analysis (PCA) was initially applied for dimension reduction. The goal of the PCA was to reduce 1766 intensity measurements to a small number of principal components that explain most of the variation in the data (9). Then linear discriminant analysis (LDA) (10) was performed on the first  $k$  principal components to derive a linear classifier. We estimated misclassification errors using Bayes' rule for each choice of  $k$ . The prior probability was set to be equal for all groups. The group membership (disease status) was determined based on the posterior probability of assigning a subject to a group. Error of classification was estimated by the proportion of subjects that were misclassified. The leave-one-out cross-validation method was used to evaluate the accuracy of the metabolic profile that predicts the group membership of a sample on the basis of the classifiers. The procedure is based on repeatedly withholding one patient at a time, and the complementary training set is used for the prediction error estimation. The validation method includes the construction of principal components based on all metabolites prior to the class separation step by LDA for every training set. The prediction error is calculated by the rate of misclassified samples when predicting for each sample using the training set. This procedure is repeated, leaving out each patient at a time until all patients have been classified and then averaging the prediction error rates over all the possible training sets.

**Differential Analysis**—Differential analysis was performed to identify metabolites whose expression differentiates the RCC case from nor-

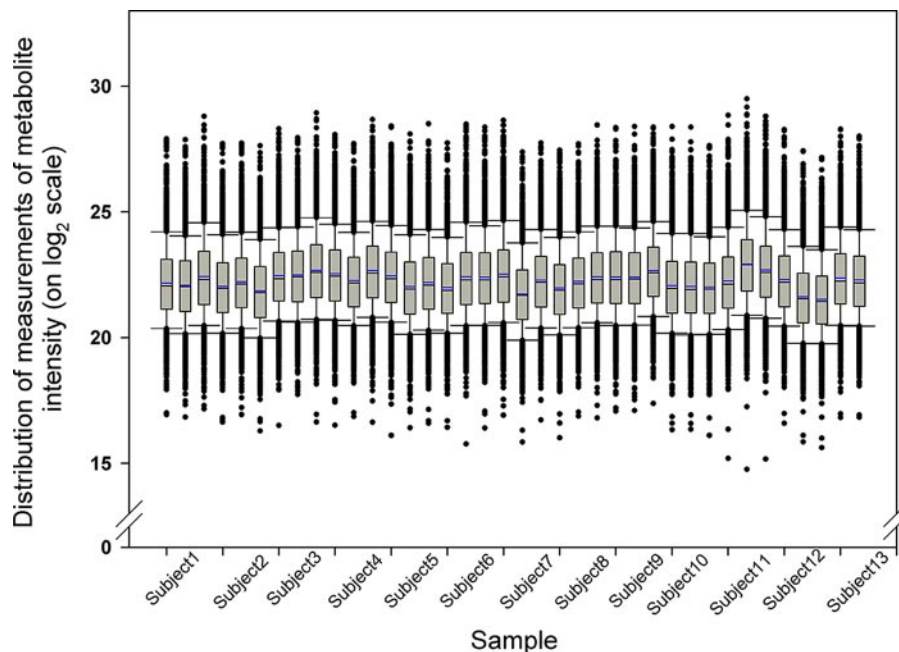
mal using a  $t$  test or paired  $t$  test. Multiple tests were controlled by the false discovery rate (FDR), the expected proportion of false positives among the tests declared significant (false plus true) (11). An FDR-adjusted  $p$  value  $<0.05$  was considered as significant. Based on the distribution of  $p$  values in  $t$  tests, a mixture model approach was used to estimate the posterior probability that a metabolite is a true positive by fitting the log likelihood function of a mixture model (12). Briefly a mixture model assumes that for each of  $k$  metabolites a null hypothesis of no difference in intensity level between cases,  $H_{0i}: \delta_i = 0, i = 1, \dots, k$  ( $\delta_i$  = difference in mean intensity level between two groups for the  $i$ th metabolite), is tested for all  $k$  metabolites with a valid test statistic, thus generating  $k$   $p$  values. Under the composite null hypothesis that no metabolite levels are different (*i.e.*  $H_{0i}: \delta_i = 0$  for all  $i = 1, \dots, k$ ), the  $k$   $p$  values are expected to follow a uniform distribution ranging from 0 to 1. For the  $m$  ( $m \leq k$ ) metabolites for which the alternative hypothesis that at least one metabolite has significantly different intensity levels is true, the  $k$   $p$  values are expected to follow a non-uniform distribution that tends to go higher near 0 than around 1. A mixture of uniform and non-uniform distributions models the distribution of  $p$  values. The parameter values in the mixture model were estimated using maximum likelihood techniques combined with a bootstrap procedure. The fitted mixture model was used to estimate the posterior probability that a metabolite is differentially expressed between the two groups. The posterior probability is the Bayesian probability that a metabolite with a given frequentist  $p$  value or smaller is truly different between the two groups being studied. In addition to the FDR-adjusted  $p$  values, the posterior probabilities were used to prioritize most promising metabolites in order of significance quantitatively.

**Evaluation of Relative Magnitudes of Different Sources of Variation**—A metabolomics experiment has many different sources of variation that can be attributed to disease cause and other factors. Variation results from urine sample heterogeneity, sex, age, disease progress stage, and other factors. We performed variance component analysis to examine the relative contributions of various factors in a metabolomics experiment (13). The relative magnitudes of different sources of variation were estimated using a linear mixed model in the PROC MIXED procedure of SAS 9.1 using the REML option. The peak intensity levels of each metabolite,  $Y_i$ , were modeled as follows:  $Y_i = \mu_i + \text{Group}_i + \text{Sex}_i + \text{Age}_i + \text{Grade}(\text{Group})_i + \varepsilon_i$  where  $\text{Group} \sim N(0, \sigma_{\text{Group}}^2)$  is the effect of disease group variation among measurement units,  $\text{Sex} \sim N(0, \sigma_{\text{Sex}}^2)$  is the effect of sex variation among measurement units,  $\text{Age} \sim N(0, \sigma_{\text{Age}}^2)$  is the effect of age variation among measurement units,  $\text{Grade}(\text{Group}) \sim N(0, \sigma_{\text{Grade}(\text{Group})}^2)$  is the effect of tumor grade variation nested within disease group, and  $\varepsilon_i \sim N(0, \sigma_{\text{Residual}}^2)$  is the residual error, *i.e.* variation caused by factors other than the variables included in the model. Disease group effect could be confounded by variation caused by factors other than the sex, age, and tumor grade. For each metabolite, variance components were estimated. The total variance was assumed to be the sum of five components:  $\text{VAR}_{\text{Tot}} = \text{VAR}_{\text{Group}} + \text{VAR}_{\text{Sex}} + \text{VAR}_{\text{Age}} + \text{VAR}_{\text{Grade}(\text{Group})} + \text{VAR}_{\text{Residual}}$ . The relative proportion of each source of variation was calculated as a ratio of the variance estimate to the sum of all variance estimates. For example,  $p_{\text{Group}} = \text{VAR}_{\text{Group}}/\text{VAR}_{\text{Tot}}$  calculates the proportion of disease group variation, and  $p_{\text{Residual}} = \text{VAR}_{\text{Residual}}/\text{VAR}_{\text{Tot}}$  calculates the proportion of variation due to unaccounted variation (residual error).

## RESULTS

Recent advances in column technology such as HILIC coupled to electrospray mass spectrometry allow the detection of highly polar compounds that appear in urine (3, 14). All urines were analyzed by the HILIC-MS technique because of the

FIG. 1. Comparison of distribution of metabolite levels within patients and between patients. Box plots represent the distribution of metabolite peak intensity measurements (on  $\log_2$  scale) from inpatient urine samples (in order of a.m. (fasting), p.m., and random) across all subjects. The box is drawn from the 25th to 75th percentiles in the distribution of intensities. The median, or 50th percentile, is drawn as a black horizontal line inside the box. The mean is represented by the blue horizontal line inside the box. The whiskers (lines extending from the box) describe the spread of the data within the 10th and 90th percentiles. The dots display any points beyond the 10th and 90th percentiles.



three techniques examined previously (gas chromatography-TOF-MS, reverse phase LC-MS, and HILIC-MS; see Ref. 3) HILIC-MS yielded the best separation in this study. As in the previous study, we did not attempt to identify all of the detected peaks but rather focused on evaluation of the use of the mass spectrometric and peak processing techniques for the development of diagnostic tests for RCC. The hypothesis was that a large group of potential biomarkers was more likely to evolve patterns for disease recognition than any single compound; once discovered as a potential biomarker, such metabolites can be chemically identified in later studies.

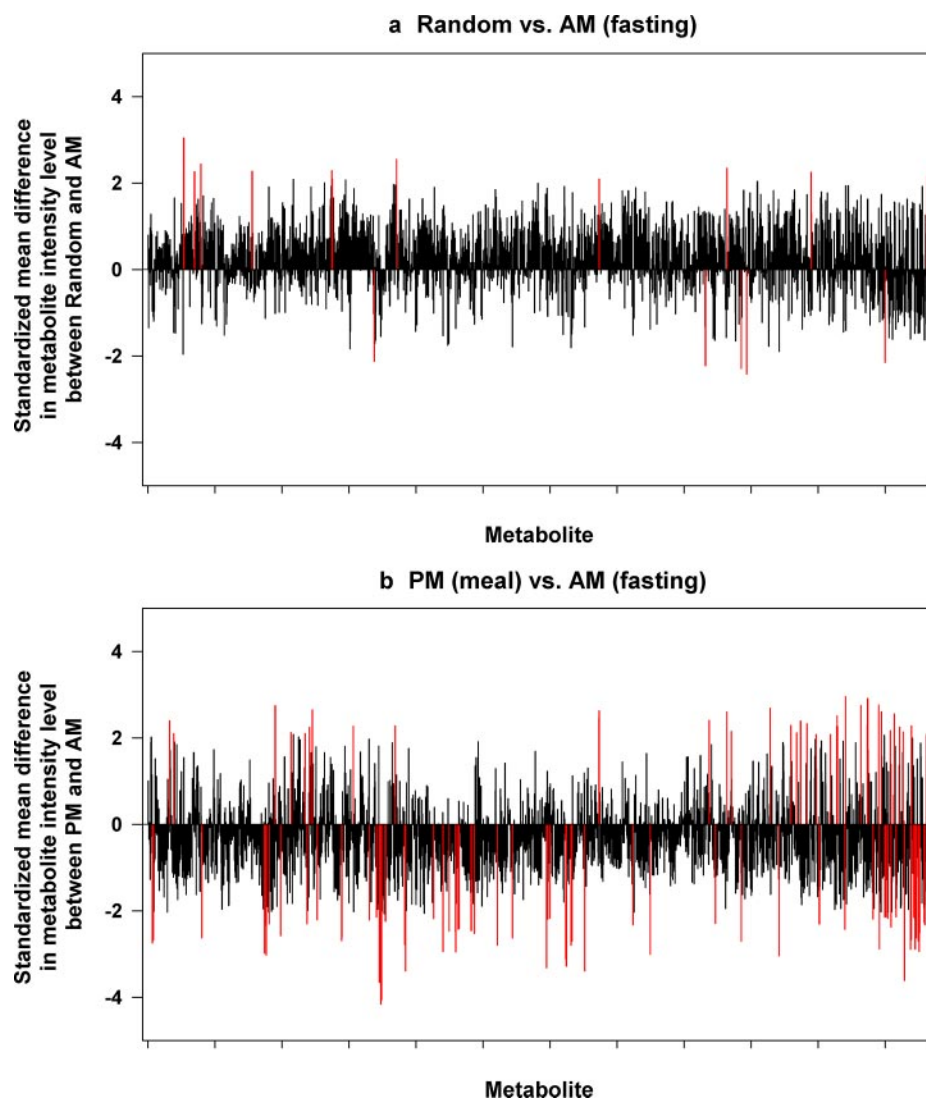
**Variability in the Measurements of Metabolite Peak Intensity Levels in Normal Subjects**—To determine the sources of variability in metabolomics analysis in normal subjects to answer the question as to whether fasting, age, body mass index (BMI), medication, or menopausal state affects the urinary metabolome, we recruited 13 healthy female subjects with varied characteristics. The participants were aged 23–76 years ( $46.31 \pm 12.76$  years). BMI was  $26.52 \pm 7.13$  kg/m<sup>2</sup> and individual BMI varied from 17.6 to 38.6 kg/m<sup>2</sup>. Of the 13 participants who were included in the study, six (46%) were premenopausal, four (31%) were perimenopausal (*i.e.* had early symptoms of menopause), and three (23%) were postmenopausal. Subjects were more likely to take medication(s) in the morning hours after the fasting urine sample was obtained. Eleven participants had taken medication in the mornings, and four had taken medication in the afternoons. We utilized several urine samples from each subject: a.m. (fasting), p.m. (30 min to 1 h after a meal), both in the same day, and a random sample (10 min to >8 h after a meal) on a separate day. When separated by HILIC-MS and the significant features were analyzed as described under “Experimen-

tal Procedures,” 2333 metabolites were discovered. Fig. 1 shows the distributions of magnitudes of peak intensity levels of all measured metabolites in each sample from each subject.

To analyze the overall levels of all urinary metabolites within and between these normal subjects, box plot analysis of metabolite levels of identified metabolites (*i.e.* chromatographic peak areas) was performed. It can be seen from this analysis that all samples showed a similar range of metabolite levels (Fig. 1). However, metabolite levels for some subjects were more tightly grouped than for others. For example, the distribution of metabolite expression values for each urine sample of subject 10 tended to be similar, whereas those for subject 7 were more divergent. For subject 10, the values from the a.m. urine sample correlated with the values from the p.m. urine sample ( $r = 0.987$ ) approximately as well as with the values from the random urine sample ( $r = 0.979$ ) as judged by the Pearson correlation coefficient. For subject 7, the correlation between the a.m. and p.m. urine samples ( $r = 0.912$ ) was smaller than that between the a.m. and random urine sample ( $r = 0.929$ ). The intrasubject urine sample correlations varied from  $r = 0.910$  to  $r = 0.987$ , whereas the intersubject urine sample correlations ranged from  $r = 0.785$  to  $r = 0.977$ , suggesting that intrasubject variability is smaller than intersubject variability.

**The Effect of Mealtimes on Metabolomic Variance in Normal Subjects**—To determine the influence of meals and mealtimes on changes of urinary metabolite profiles, we compared the intensity levels for individual metabolites between the a.m. samples and random samples and p.m. samples using the paired *t* test. Among 2333 metabolites, 16 (0.68%) metabolites showed significant differences between the a.m. and

FIG. 2. **Comparisons of metabolomic spectra between varied times of day of urine collection.** A first morning fasting sample (a.m.), a 30-min to 1-h postmeal sample (p.m.) on the same day, and a random sample taken irrespective of meals taken on a subsequent day (random) were obtained from normal subjects, and metabolites were analyzed by HILIC-MS. The standardized mean difference in intensity level for each of the metabolites ( $x$  axis) is shown on  $y$  axis. The *red line* represents the metabolites whose levels are significantly different between the two urine collection times.

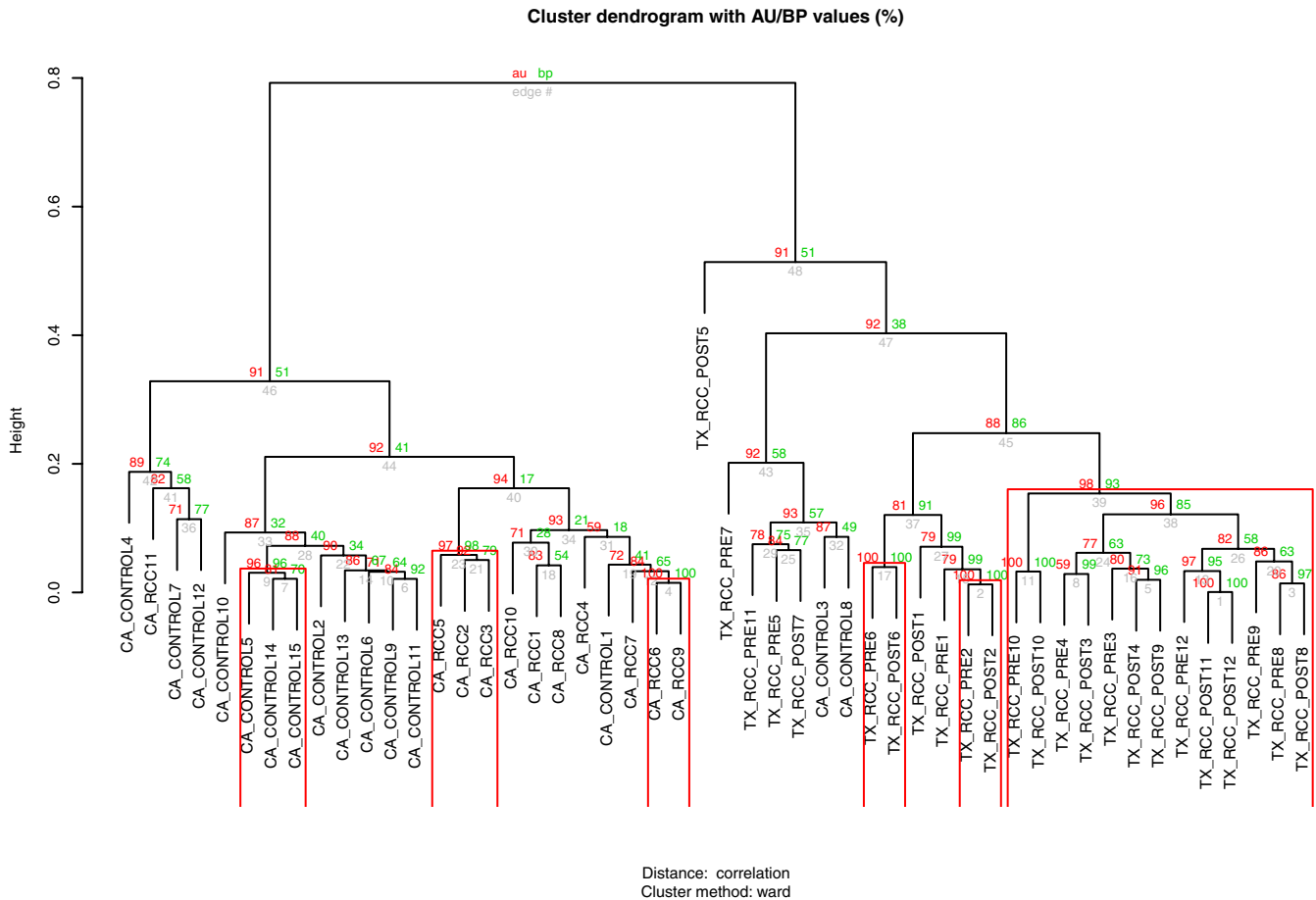


random samples (see Fig. 2a, *red lines*). There were 104 (4.45%) metabolites whose expression was differentially expressed between the a.m. samples and p.m. samples (Fig. 2b). However, after adjustment for FDR, none of those significant metabolites were declared as significant at FDR-adjusted  $p$  value  $<0.05$ . It appears that more variation of metabolites was seen in the p.m. (obtained postmeal) urines as compared with the random urines taken irrespective of meals taken on a subsequent day, suggesting that changes in urine metabolite analysis are more pronounced in urines after a meal as compared with urines randomly collected and fasting urines. For the majority of the metabolites, the effects of meals on metabolomic variance were minimal. For these reasons, we did not control for fasting or time after meal status in subsequent analyses.

*Investigation of RCC Group- and Study Site-related Variations by Cluster Analysis*—In the next set of experiments, urine samples from 11 patients with clear cell RCC (various stages and grades) and 15 non-RCC control patients were

obtained from the urology clinic at CA, and 12 pre-nephrectomy clear cell RCC and 12 post-nephrectomy (seven partial and five radical) were obtained from urology clinics at TX. Pre- and postoperative urines were from the same patients obtained at clinic visits from 1 to 5 months after surgery, and fasting status was not controlled due to the observation described above under “The Effect of Mealtimes on Metabolomic Variance in Normal Subjects” using normal subjects. Un-supervised hierarchical cluster analysis was first performed to investigate whether the urine samples from RCC and control patients separated according to disease status and/or study site. The dendrograms showed separation between RCCs and controls and samples obtained from the CA and TX collection sites (Fig. 3).

There was distinct separation of the CA control urine sample cluster from both the CA RCC and the TX RCC urine clusters (both preoperative and postoperative). Separation of the preoperative TX RCC and postoperative TX RCC urine sample groups appeared to be less distinct and formed a



**FIG. 3. Urine metabolomic profiles segregate by cancer status and collection site but not operative status.** Values on the *edges* of the clustering are *p* values (%). *Red* values are AU *p* values, which are computed by multiscale bootstrap resampling, and *green* values are bootstrap probability values, which are computed by normal bootstrap resampling. Clusters with AU larger than 95% are highlighted by *rectangles* and are strongly supported by data. Urines are labeled as follows: the first term refers to site of collection; the second is RCC versus control; the third is pre- (PRE) or postoperative (POST) (where indicated); and the fourth is patient number.

tighter cluster although the (preoperative) CA RCC cluster clearly separated from the preoperative TX RCC cluster. Thus, urine sample metabolomic features segregate in part due to locale of collection, and 1–5-month postoperative samples demonstrate features similar to those of the preoperative samples collected from the same collection site.

**Identifying Differential Metabolites Associated with Disease Status and Surgery Status**—To identify metabolites that influence the differences between disease group and surgical status the most, we next compared the intensity levels for individual metabolites between the control group and RCC case group in the CA data set. Among 1766 metabolites, 455 metabolites showed statistically significant differences between the two disease groups (supplemental Table 1). However, after adjustment for FDR, 212 metabolites were significantly differentially expressed between the two groups at FDR-adjusted *p* value <0.05. For the TX data set, there were only five metabolites whose intensity level altered after surgery. Among these five metabolites, none was declared as significant at FDR-adjusted *p* value <0.05. To prioritize the

most promising metabolites as metabolomic profile biomarkers, we applied a mixture model approach to estimate the posterior probability of a metabolite being a true positive by fitting the log likelihood function of a mixture model on the distribution of resulting *t* test *p* values. The mixture models estimated that 212 metabolites had posterior probabilities greater than 92% that a metabolite is truly expressed differentially between the control and RCC groups in the CA data set. These metabolites will be chemically identified in future studies.

**Principal Component Analysis of HILIC-MS Urinary Metabolomic Profiles**—To discern the presence of inherent similarities in spectral profiles, we initially performed PCA of all HILIC-MS spectra obtained from each sample. Because the cluster analysis showed that the urinary metabolite profile is dependent on institutional site of sample collection, we kept CA samples separate from TX samples for the remainder of the analyses. A representative spectrum of all the urine samples from CA is mapped in the space spanned by the first three principal components PC1 versus PC2 versus PC3 (Fig.

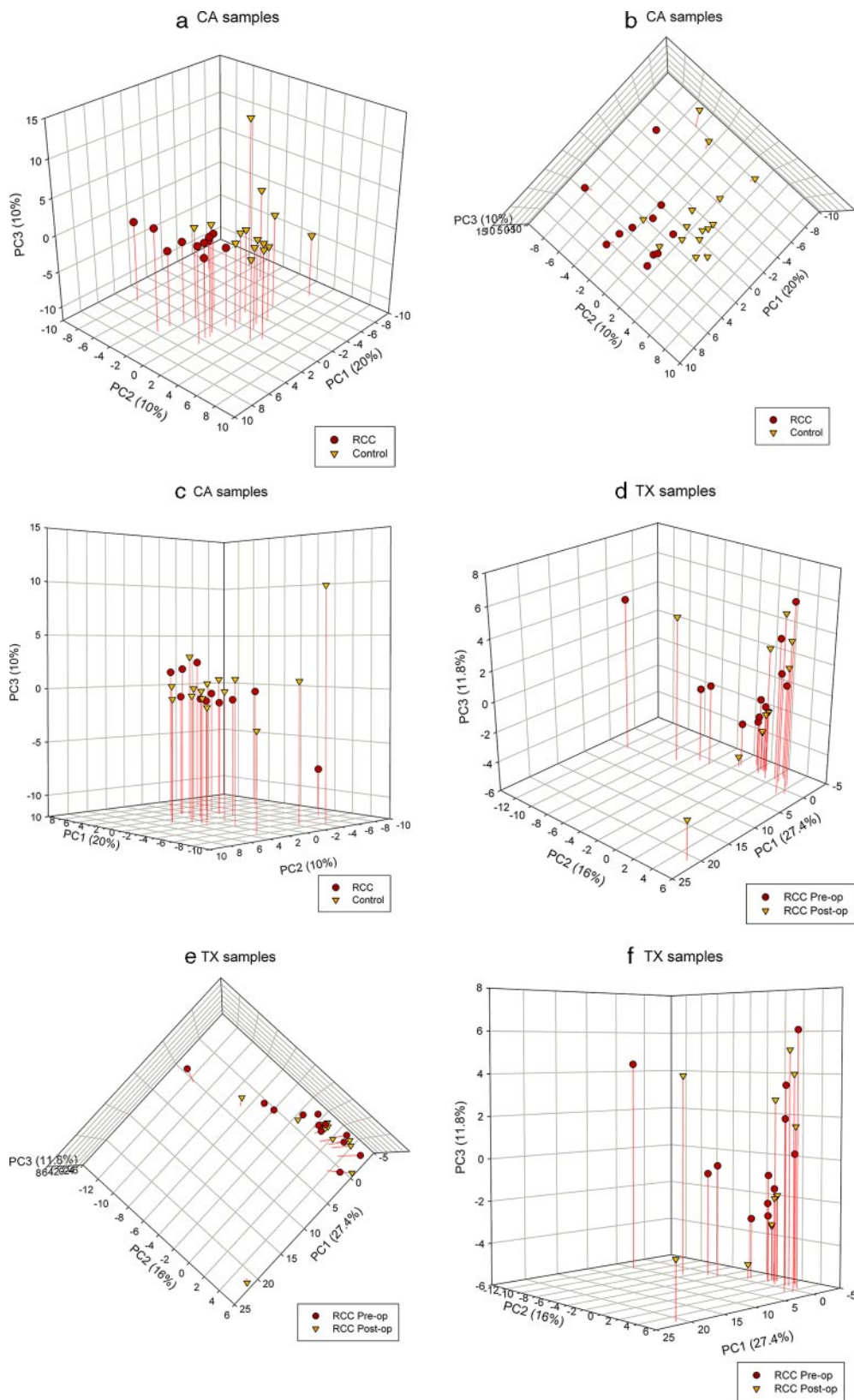


FIG. 4. Three-dimensional PC scores plots derived from the LC-HILIC spectra of urine samples show differences between cancer and control urine metabolomic profiles. PCA plots are derived from spectral data with percentage of variance captured by each PC for CA samples (a-c) and TX samples (d-f). *op*, operative.

TABLE I

Summary of prediction of group membership using LDA for each choice of  $k$ , the number of PC scores

$k$ , the number of the first principal components used in linear discriminant analysis. Percentage of total variation, the percentage of total variance in data that is explained by the first  $k$  principal components.

$k$	Percentage of total variation	Overall percentage of correct predicted classifications	Percentage of correctly classified RCCs <sup>a</sup>	Percentage of misclassified RCCs <sup>b</sup>
CA, $n = 26$ (prediction of RCC)				
2	30	88.46	90.91	13.33
3	40	88.46	81.82	6.67
4	50	88.46	90.91	13.33
5	57	88.46	90.91	13.33
6	63	100	100	0
TX, $n = 24$ (prediction of preop)				
2	43.4	54.17	66.67	58.33
3	55.2	66.67	58.33	25
4	62.9	66.67	50	16.67
5	68.85	58.33	58.33	41.67
10	88.2	58.33	58.33	41.67
15	96.08	79.17	75	16.67
20	99.46	95.83	91.67	0
21	99.75	100	100	0

<sup>a</sup> Percentage of correctly classified RCCs is the percentage of RCC (preoperative (preop)) samples correctly classified as RCC (preoperative).

<sup>b</sup> Percentage of misclassified RCCs is the percentage of controls (postoperative) misclassified as RCC (preoperative).

4a). 40% of the total variance in the data was captured by the first three components. The score plot of PC1 versus PC2 shows separation between RCC (red circle) and control (orange triangle) samples as expected based on the cluster analysis (Fig. 4b). The score plot of PC3 against PC1 and PC2 illustrates that PC3 did not help in separating samples according to disease status (Fig. 4c). It is clearly shown that PC1 and PC2 together (30% of variance) captured most of the variations between the controls and RCC cases. This observation is consistent with cluster analysis shown above and using samples distinct from our pilot study (3) and confirms the utility of urine metabolomics analysis in segregating RCC from control patients. The PCA score plots of the PCs showed no separation between the preoperative and postoperative TX samples (Fig. 4, d–f) consistent with the cluster analysis shown above where the preoperative and postoperative samples formed a large single cluster rather than two smaller clusters. Thus, urine metabolite profiles (i) can separate RCC from urology clinic control patients and (ii) are not separable in urine samples taken before and soon after removal of the primary tumor.

To determine which regions of the spectra are causing separation between the two groups, we examined the loadings of PC1 and PC2 and found that highly differentiated peaks between the two groups ( $t$  test  $p$  value  $<0.05$  and posterior probability  $>0.85$ ) appeared to have heavy loadings, suggesting that those peaks were the most influential for the disease group separation in the CA samples (data not shown). Thus, the peaks with heavy loadings play a major role in classification of RCC versus control. The investigation of the higher order PCs revealed no additional separation between

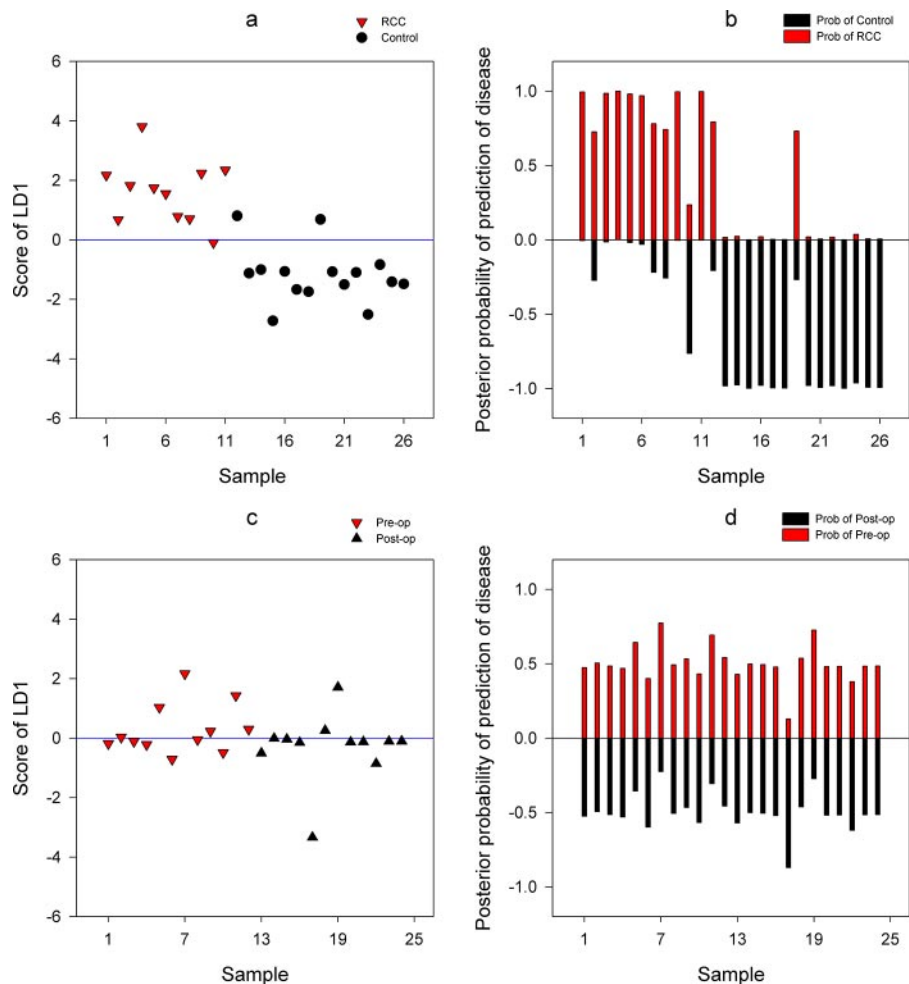
the groups; thus, the first two to four components (30–50% of variation) appeared to yield sufficient discriminatory power for the present study. For the TX samples, the first 20 components (99.46% of total variation) were needed for good separation between the pre- and postoperative samples.

**Prediction of Group Membership by Linear Discriminant Analysis**—To determine the classification rule and derive the best classifier that can separate urine samples into a number of groups, we utilized the supervised discriminant method, LDA. The LDA was carried out to derive a linear classifier using the first  $k$  principal component scores as new features that describe samples. Table I summarizes the prediction results for each choice of  $k$ , the number of PC scores to LDA. The predictive performance varied by the choice of  $k$ . The score plot of the first discriminant and the posterior probabilities of correctly assigning samples to the true group ( $k = 4$ ) are shown graphically according to cancer status and site of collection (Fig. 5).

The predictive performance of the linear discriminant model was validated by the leave-one-out cross-validation method. Table II shows the average percentage obtained from 26 and 24 linear discriminant models for the CA and TX samples, respectively. The cross-validated prediction results show that for four-component models the differences between the control and RCC samples can be predicted accurately in 89.8% of RCC cases for the CA data set. The overall rate of correct classification was 88.31% within the CA set. In the case of TX samples, a poor classification was attained as expected due to the lack of separation between the pre- and postoperative samples as shown in clustering as well as PCAs shown earlier. The differences between the preoperative and postoperative



**FIG. 5. Prediction of group membership using LDA for the choice of  $k = 4$ .** LDA analysis was performed on all samples for CA samples (a and b) and TX samples (c and d). The posterior probabilities for the groups are shown on the y axis. The absolute posterior probability is the probability of assigning a sample to each group. *op*, operative.



samples are predicted accurately in 56.19% of preoperative samples. For most of the TX samples, the posterior probabilities of correctly classifying samples to their true group membership ranged around 50% (Fig. 5d).

**Estimation of Variance Components in Urology Clinic Patients**—Variations in metabolomics data can arise from several sources that are attributed to both biological and technical causes. Biological variations result from heterogeneity among samples due to disease status, sex, age, race, gene-environment interactions, and other factors that are real and of interest to investigators. Technical variations reflect changes in experimental conditions during data processing, which can significantly impact the quality of data and introduce bias to results. Hence variance component analysis was next performed to assess the relative contributions of various factors in our metabolomics experiment. We estimated the relative magnitudes of sources of variation using the 26 samples collected at the University of California, Davis and the 24 samples collected at the University of Texas at San Antonio separately because the two data sets had different sources of group variation. The group variation in the CA data set resulted from the presence/absence of disease, whereas the

group variation in the TX data set was due to the removal of the primary tumor.

For each metabolite within the CA data set, intensity levels were modeled as follows:  $Y_i = \mu_i + \text{Group}_i + \text{Sex}_i + \text{Age}_i + \text{Grade}(\text{Group})_i + \varepsilon_i$  where  $\text{Group} \sim N(0, \sigma_G^2)$  is the effect of disease group variation among measurement units,  $\text{Sex} \sim N(0, \sigma_S^2)$  is the effect of sex variation among measurement units,  $\text{Age} \sim N(0, \sigma_A^2)$  is the effect of age variation among measurement units,  $\text{Grade}(\text{Group}) \sim N(0, \sigma_{\text{Grade}}^2)$  is the effect of tumor grade variation nested within the disease group, and  $\varepsilon_i \sim N(0, \sigma_R^2)$  is the residual error, which is the portion of total variance that cannot be explained by the factors included in the model. The model was fit separately on the peak intensity level measurements of each of the metabolites. All factors were treated as random, and variance components were estimated. The distributions of relative magnitudes of different sources of variation showed that most of the metabolites had small biological variance, indicating little variability of measured intensity levels between the different groups (Fig. 6a). A small portion of the metabolites showed large variation between the two different disease groups: 112 metabolites (6%) (metabolites at the *upper tail* in Fig. 6a, *Group*) had relative

TABLE II  
Summary of classification percentages using PCA-linear discriminant models for each choice of  $k$

Data are presented as mean  $\pm$  S.D. The percentages reported here are the average of the number of cross-validated PCA-linear discriminant models.  $k$ , the number of the first principal components used in linear discriminant analysis. Percentage of total variation, the average percentage of total variance in data that is explained by the first  $k$  principal components.

$k$	Percentage of total variation	Overall percentage of correct predicted classifications	Percentage of correctly classified RCCs <sup>a</sup>	Percentage of misclassified RCCs <sup>b</sup>
CA (control vs. RCC)				
2	30.1 $\pm$ 0.2	88.0 $\pm$ 2.5	89.4 $\pm$ 4.2	12.8 $\pm$ 3.1
3	40.1 $\pm$ 0.3	89.6 $\pm$ 2.6	85.5 $\pm$ 5.6	7.5 $\pm$ 2.8
4	50.4 $\pm$ 0.6	88.3 $\pm$ 1.6	89.8 $\pm$ 3.9	12.8 $\pm$ 2.6
5	57.8 $\pm$ 0.7	89.7 $\pm$ 3.4	90.8 $\pm$ 5.1	11.5 $\pm$ 4.4
6	64.5 $\pm$ 0.6	98.5 $\pm$ 3.0	98.2 $\pm$ 3.8	1.4 $\pm$ 3.5
TX (preop vs. postop)				
2	41.5 $\pm$ 3.4	56.4 $\pm$ 5.2	56.4 $\pm$ 13.4	39.5 $\pm$ 20.0
3	59.0 $\pm$ 2.9	59.7 $\pm$ 7.1	58.9 $\pm$ 23.9	36.0 $\pm$ 25.6
4	63.5 $\pm$ 1.4	57.5 $\pm$ 6.0	56.2 $\pm$ 17.1	37.8 $\pm$ 23.6
5	69.7 $\pm$ 1.2	59.4 $\pm$ 6.3	59.0 $\pm$ 12.5	36.0 $\pm$ 15.7
10	89.0 $\pm$ 0.6	59.7 $\pm$ 6.2	66.9 $\pm$ 8.8	42.7 $\pm$ 11.9
15	96.7 $\pm$ 0.5	79.9 $\pm$ 4.2	85.1 $\pm$ 7.0	18.5 $\pm$ 4.7
20	99.7 $\pm$ 0.1	92.5 $\pm$ 1.7	93.1 $\pm$ 3.6	0 $\pm$ 0
21	99.9 $\pm$ 0.1	95.5 $\pm$ 1.2	99.3 $\pm$ 2.4	0 $\pm$ 0

<sup>a</sup> Percentage of correctly classified RCCs is the percentage of RCC samples correctly classified as RCC (preoperative (preop)).

<sup>b</sup> Percentage of misclassified RCCs is the percentage of controls (postoperative (postop)) misclassified as RCC (preoperative).

proportions of disease group variation greater than 0.5, suggesting that for those metabolites the disease group was the major contribution to variation in the data; these metabolites are most likely to result in viable biomarkers for RCC. Hence we focused our attention on these metabolites, which particularly likely play a major role in classification. We performed PCA and LDA for only those 112 metabolites and assessed the predictive performance as described above. The 112 metabolites yielded a similar overall rate of correct classification compared with that of using the entire 1766 metabolites, suggesting that the metabolites with greater disease variation are potent factors in the classification. For the majority of the metabolites, the greatest source of variation was residual error such that these metabolites will not be useful as biomarkers. Table III summarizes the descriptive statistics of each variance component.

For the TX data set, intensity levels for each metabolite were modeled as follows:  $Y_i = \mu_i + \text{Surgery}_i + \text{Sex}_i + \text{Age}_i + \text{Grade}_i + \varepsilon_i$  where  $\text{Surgery} \sim N(0, \sigma_{\text{Surgery}}^2)$  is the effect of surgery status variation among measurement units,  $\text{Sex} \sim N(0, \sigma_{\text{Sex}}^2)$  is the effect of sex variation among measurement units,  $\text{Age} \sim N(0, \sigma_{\text{Age}}^2)$  is the effect of age variation among measurement units,  $\text{Grade} \sim N(0, \sigma_{\text{Grade}}^2)$  is the effect of tumor grade variation nested within the disease group, and  $\varepsilon_i \sim N(0, \sigma_{\text{Res}}^2)$  is the residual error. The model was fit separately on the peak intensity level measurements of each of the metabolites. All factors were treated as random, and variance components were estimated. The distributions of relative magnitudes of different sources of variation for the TX data set were similar to that for the CA data set (Fig. 6b). The results indicated that most of the metabolites had very small biological variance;

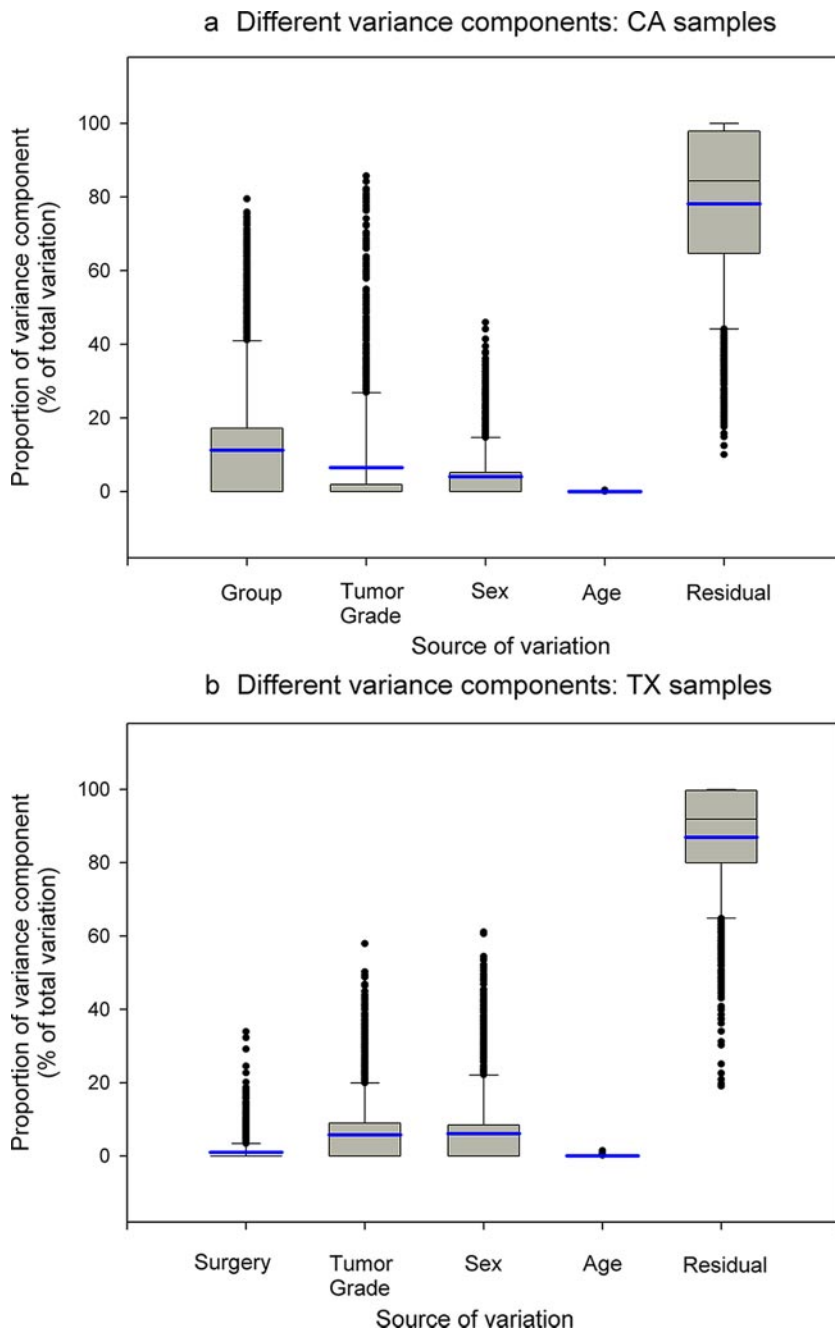
specifically of interest is that less than 1% of variance was caused by surgery status. For the majority of the metabolites, the greatest source of variation was residual error. Table IV summarizes descriptive statistics of each variance component for the TX data set.

## DISCUSSION

Although kidney cancer is the sixth leading cause of cancer deaths and represents 3% of cancer incidence, it is the direct cause of death in  $\sim$ 11,000 patients per year in the United States. Not only is the disease notoriously chemotherapy-resistant, but it is often found incidentally such that one-third of cases are metastatic at diagnosis. When detected with symptoms, prognosis of RCC is poor (15); once metastatic, RCC has a 5% 5-year survival (16). For these reasons, novel, convenient, and non-invasive approaches are needed for identifying RCC at an earlier stage prior to metastasis.

The new science of metabolomics, in which the entire suite of metabolites generated by an organism is examined, holds promise for the development of diagnostic tests based on metabolite profiling as well as for the discovery of individual biomarkers (so-called "slow" and "fast-track" (3)). Given the ease of its collection, urine is a logical biofluid to examine in this regard; in light of the fact that RCC is a urothelial tumor, this malignancy was chosen for an initial application of this technology. In a pilot study using a limited number of samples, we have shown previously that urine metabolomics analysis has the capability of separating RCC from control patients (3). The present study extends that work, provides a means of selecting potential urinary biomarkers, and identifies

**FIG. 6. Assessment of sources of variation in metabolomic profiles.** Box plots of relative magnitudes of different sources of variation are shown for CA samples (a) and TX samples (b). The proportions of different variance components are shown on the y axis. The blue line represents the mean. The box is drawn from the 25th to 75th percentiles in the distribution of proportions of each variance component. The median, or 50th percentile, is drawn as a black horizontal line inside the box. The whiskers (lines extending from the box) describe the spread of the data within the 10th and 90th percentiles. The dots display any points beyond the 10th and 90th percentiles.



**TABLE III**  
Proportions of different sources of variation: CA data set  
CV, coefficient of variation; Min, minimum; Max, maximum.

	Source group	Tumor grade	Sex	Age	Residual
Mean	11.27	6.51	4.05	0.02	78.15
S.D.	18.01	15.16	7.48	0.05	21.65
CV (%)	159.82	232.91	184.42	233.15	27.70
Median	0	0	0	0	84.38
Min	0	0	0	0	10.05
Max	79.50	85.72	45.95	0.36	100

confounding issues that need to be considered in interpreting data from such studies.

In addition to confirming our pilot study and demonstrating that patients with RCC can be differentiated from control patients by urine metabolomics analysis, it is shown that institutional site of sample collection caused substantial variability in metabolomics analysis. Despite attempts to bring uniformity into the collection process and notwithstanding the fact that all samples were analyzed (although not collected) in the same laboratory by the same personnel, significant features of samples collected at the University of California,

TABLE IV  
Proportions of different sources of variation: TX data set  
CV, coefficient of variation; Min, minimum; Max, maximum.

	Source surgery	Tumor grade	Sex	Age	Residual
Mean	1.04	5.79	6.14	0.06	86.97
S.D.	3.18	9.69	11.16	0.13	14.72
CV (%)	305.83	167.25	181.66	218.51	16.93
Median	0	0	0	0	91.86
Min	0	0	0	0	19.07
Max	33.94	57.92	61.12	1.46	100

Davis did not intersect with those collected at the University of Texas at San Antonio consistent with previous reports in other cancers examined using proteomics (17). Thus, future investigations need to be directed at the source of variability created by the use of more than one institution, a serious problem for generalizability of a screening test; until this is determined, urine metabolomics studies should be confined to a single collection site. Similar results have been observed with other omics techniques (18).

We have also now addressed important issues regarding the urine metabolome in normal subjects. Some investigators have shown large intraindividual variability in both normal and animal control populations looking at specific metabolites (19), yet others have shown, at least in dogs, that intraindividual variability is relatively small (20). A surprising result of our study is that there was minimal variation in urine metabolomics analysis when examined relative to the time of the patient's last meal. For the majority of the metabolites, the influence of meals on their metabolomic profiles was not significant, and thus it can be concluded based on this pilot study that it is probably not necessary to tightly control for normal meals in urinary metabolomics analysis. However, we have not studied the influence of diet composition on the urine metabolome, an important area of focus for future studies, nor have we studied the effects of mealtimes on the urine metabolome in diseased patients. In any case, in our study of RCC patients and urology clinic controls, a single urine sample was obtained from each patient. Between-individual correlations within the same collection site ranged from 0.82 to 0.98; thus interindividual variability in these metabolomic profiles seems to be minimal. Intraindividual variation, as well as inter- and intralaboratory variability, will need to be investigated in future studies.

We have found that, although there is a high degree of group variation in the urine metabolomics analyses (*i.e.* between RCC and control), it is apparent that there is substantial residual variation as well. Those metabolites that strongly contributed to the group variation (see Fig. 6) will be identified in future studies as potential biomarkers, whereas the sources of residual variation will need to be determined to maximize the discriminatory potential of any future urine test. Sources of residual variation may include dietary issues, medications,

length of time of urine storage, and handling, etc. In addition, our choice of HILIC-MS instead of other types of separation and molecular analysis (*e.g.* gas chromatography-MS or LC-MS; see Ref. 3) may yield different outcomes for variance component analysis and cluster analysis due to differences in techniques.

Although it was hypothesized that postoperative samples would serve as controls for preoperative samples from the same patients, a surprising yet biologically and clinically significant finding was that postoperative urine metabolomic profiles did not segregate from preoperative samples from the same institution. This suggests that metabolic changes persist in patients at least several months after removal of the primary tumor, a finding that may be explained by the existence of known paraneoplastic syndromes that accompany RCC (16). In addition, our finding of close clustering of pre- and postoperative samples from the University of Texas at San Antonio serves as a control for reproducibility of collection parameters as well as analytical technique.

Our results demonstrate the feasibility of using metabolomic profiles to identify metabolomic profiles and biomarkers predictive of cancer. However, the current study has limitations. We identified sources of experimental variations and assessed their magnitude by variance component analysis. The analysis identified factors that would contribute to the success of future metabolomics studies. When designing a metabolomics study, researchers should adhere to the principles of study design: matching the experimental variables of cases and controls to the fullest extent possible, selecting clinically homogenous sample populations, and balancing a design with respect to all factors that can confound results among the comparison groups. Violation of these principles will lead to biased results and can cause a loss in power. A large sample size is necessary to achieve a good power to demonstrate significance of findings, particularly with thousands of metabolites to test, but the number of cancer samples available at a single clinic given the period of study is limited. Therefore, as in the current study, collaborative efforts on recruitment between several clinics may be necessary. In this case, the success of metabolomics studies depends on careful selection of sample populations and collaborative analytical approaches. Moreover to ensure highly reproducible metabolomics results, technical variation should be minimized, in the planning of experiments, by controlling the quality of the urine samples and by efficient and uniform data collection procedures.

#### CONCLUSION

We have addressed issues relating to urine sample collection and have confirmed the utility of urine metabolomics analysis for differentiating kidney cancer from control patients. We have shown, using variance analysis, that potential biomarkers are identifiable. However, there remain potential sources of variability and confounding factors that need to be

addressed when carrying out such studies and in searching for biomarkers in the future.

**Acknowledgments**—We thank Deborah Hensley at the University of Texas at San Antonio and Dr. Chris Evans, Dr. Ralph de Vere White, and Stephanie Soares at the University of California, Davis for assistance in obtaining urine samples and Dr. Vladimir Tolstikov of the Genome Center Metabolomics Core at the University of California, Davis for technical assistance with the HILIC-MS.

\* This work was supported, in whole or in part, by National Institutes of Health Grant 5U01CA86402 from the Early Detection Research Network, NCI. This work was also supported by the Research Service of the United States Department of Veterans Affairs, Morris Animal Foundation Grant D06CA-065, and a grant from Dialysis Clinics, Inc.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

§§ To whom correspondence should be addressed: Division of Nephrology, Dept. of Internal Medicine, Genome and Biomedical Sciences Bldg., Rm. 6312, University of California, One Shields Ave., Davis, CA 95616. Tel.: 530-752-4010; Fax: 530-752-3791; E-mail: rhweiss@ucdavis.edu.

### REFERENCES

- Dettmer, K., and Hammock, B. D. (2004) Metabolomics—a new exciting field within the “omics” sciences. *Environ. Health Perspect.* **112**, A396–A397
- Schmidt, C. (2004) Metabolomics takes its place as latest up-and-coming “omic” science. *J. Natl. Cancer Inst.* **96**, 732–734
- Kind, T., Tolstikov, V., Fiehn, O., and Weiss, R. H. (2007) A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal. Biochem.* **363**, 185–195
- Han, W. K., Waikar, S. S., Johnson, A., Betensky, R. A., Dent, C. L., Devarajan, P., and Bonventre, J. V. (2008) Urinary biomarkers in the early diagnosis of acute kidney injury. *Kidney Int.* **73**, 863–869
- Waikar, S. S., and Bonventre, J. V. (2007) Biomarkers for the diagnosis of acute kidney injury. *Curr. Opin. Nephrol. Hypertens.* **16**, 557–564
- Han, W. K., Alinani, A., Wu, C. L., Michaelson, D., Loda, M., McGovern, F. J., Thadhani, R., and Bonventre, J. V. (2005) Human kidney injury molecule-1 is a tissue and urinary tumor marker of renal cell carcinoma. *J. Am. Soc. Nephrol.* **16**, 1126–1134
- Hartigan, J. A. (1975) *Clustering Algorithms*, Wiley, New York
- Suzuki, R., and Shimodaira, H. (2006) Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542
- Jolliffe, I. T. (2002) *Principal Component Analysis*, 2nd Ed., Springer, New York
- Fisher, R. (1936) The use of multiple measurements in taxonomic problem. *Ann. Eugenics* **7**, 179–188
- Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300
- Allison, D. B., Gadbury, G. L., Heo, M., Fernandez, J. R., Lee, C., Prolla, T. A., and Weindruch, R. (2002) A mixture model approach for the analysis of microarray gene expression data. *Comput. Stat. Data Anal.* **39**, 1–20
- Zakharkin, S. O., Kim, K., Mehta, T., Chen, L., Barnes, S., Scheirer, K. E., Parrish, R. S., Allison, D. B., and Page, G. P. (2005) Sources of variation in Affymetrix microarray experiments. *BMC Bioinformatics* **6**, 214
- Tolstikov, V. V., and Fiehn, O. (2002) Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* **301**, 298–307
- Leslie, J. A., Prihoda, T., and Thompson, I. M. (2003) Serendipitous renal cell carcinoma in the post-CT era: continued evidence in improved outcomes. *Urol. Oncol.* **21**, 39–44
- Weiss, R. H., and Lin, P.-Y. (2006) Kidney cancer: identification of novel targets for therapy. *Kidney Int.* **69**, 224–232
- McLerran, D., Grizzle, W. E., Feng, Z., Thompson, I. M., Bigbee, W. L., Cazares, L. H., Chan, D. W., Dahlgren, J., Diaz, J., Kagan, J., Lin, D. W., Malik, G., Oelschlager, D., Partin, A., Randolph, T. W., Sokoll, L., Srivastava, S., Srivastava, S., Thorquist, M., Troyer, D., Wright, G. L., Zhang, Z., Zhu, L., and Semmes, O. J. (2008) SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate cancer. *Clin. Chem.* **54**, 53–60
- Chen, J. J., Hsueh, H. M., Delongchamp, R. R., Lin, C. J., and Tsai, C. A. (2008) Reproducibility of microarray data: a further analysis of microarray quality control (MAQC) data. *BMC Bioinformatics* **25**, 412
- Saude, E. J., Adamko, D., Rowe, B. H., Marrie, T., and Sykes, B. D. (2007) Variation of metabolites in normal human urine. *Metabolomics* **3**, 439–451
- Viant, M. R., Ludwig, C., Rhodes, S., Guenther, U. L., and Allaway, D. (2007) Validation of a urine metabolome fingerprint in dog for phenotypic classification. *Metabolomics* **3**, 453–463