Published in final edited form as:

J Investig Med. 2010 April ; 58(4): 612–620. doi:10.231/JIM.0b013e3181d473e7.

Identification of Diagnostic Urinary Biomarkers for Acute Kidney Injury

Sanju A. Varghese¹, T. Brian Powell¹, Michael G. Janech¹, Milos N. Budisavljevic^{1,3},
Romesh C. Stanislaus², Jonas S. Almeida², and John M. Arthur^{1,3}

¹Department of Medicine, Division of Nephrology, Medical University of South Carolina, Charleston, SC, USA

²Department of Biostatistics and Applied Mathematics, University of Texas MD Anderson Cancer Center, Houston, TX

³Department of Medicine, Ralph H. Johnson VA Medical Center, Charleston, SC, USA

Abstract

Acute kidney injury (AKI) is an important cause of death among hospitalized patients. The two most common causes of AKI are acute tubular necrosis (ATN) and prerenal azotemia (PRA). Appropriate diagnosis of the disease is important but often difficult. We analyzed urine proteins by 2-DE from 38 patients with AKI. Patients were randomly assigned to a training set, an internal test set or an external validation set. Spot abundances were analyzed by artificial neural networks (ANN) to identify biomarkers which differentiate between ATN and PRA. When the trained neural network algorithm was tested against the training data it identified the diagnosis for 16/18 patients in the training set and all 10 patients in the internal test set. The accuracy was validated in the novel external set of patients where 9/10 subjects were correctly diagnosed including 5/5 with ATN and 4/5 with PRA. Plasma retinol binding protein (PRBP) was identified in one spot and a fragment of albumin and PRBP in the other. These proteins are candidate markers for diagnostic assays of AKI.

Keywords

Acute kidney injury; Biomarkers; Diagnosis; Kidney; Urine

BACKGROUND

Mortality associated with acute kidney injury (AKI) is approximately 50% in spite of advances in medical therapy [1,2]. Timely recognition of AKI and its causes could lead to better understanding of the pathophysiology and improve outcomes. Unfortunately no reliable diagnostic markers which can predict the cause of AKI are currently available. The two most common causes of AKI are prerenal azotemia (PRA) and acute tubular necrosis (ATN) [1–3]. Identification of markers that predict which of these diseases is present would allow earlier and more appropriate therapy. The most common method to differentiate ATN from PRA is

Address correspondence to: John M. Arthur, Department of Medicine, Division of Nephrology, Medical University of South Carolina, Charleston, SC, 96 Jonathan Lucas Street, PO Box 250623, Charleston, SC 29425, USA. Phone: (843)792-4123; Fax: (843)792-8399; arthurj@musc.edu.

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.

to determine the fractional excretion of sodium (FENa) in the urine [4]. The FENa is typically low in PRA and higher in ATN. However, treatment with diuretics increases the FENa in PRA. In addition to the problem with diuretics, FENa often gives incorrect results in other situations [5–8]. Because of the difficulty differentiating between these conditions, new markers are necessary.

Several urine protein markers of tubular injury have been proposed including serum proteins like albumin, alpha-1 microglobulin, beta 2-glycoprotein and plasma retinol binding protein [9,10]. Proteins of renal tubular origin which are potential markers include NHE3 [11], N-acetyl-glucosaminidase (NAG) [12,13], neutrophil gelatinase-associated lipocalin (NGAL) [14], cytokines and proteases [15]. Kidney Injury Molecule-1 (KIM-1), a proximal tubular brush border protein which is upregulated by injury, was identified as a potential urinary marker of tubular injury [16,17] and a combination of KIM-1 and NAG have been used to predict the need for renal replacement therapy in patients with acute kidney injury [12]. Urine levels of the two proteins, along with four clinical variables, predicted the need for renal replacement therapy better than any of the factors alone. IL-18 is known to be a mediator of AKI in animal models. Urine levels of IL-18 correlate with the diagnosis of ATN compared to other causes of AKI [18]. Elevated levels of NGAL were highly correlated with AKI in a study of patients seen in an emergency room [19]. These markers have been used to identify early tubular injury or to provide prognostic information but have not been used to distinguish ATN from PRA. Traditionally, potential biomarkers have been identified from known proteins and screened individually for their diagnostic potential. Proteomic methodologies can quantify multiple different potential markers simultaneously in the sample to provide a more powerful analysis [20–22]. Previous studies of urine biomarkers have shown an advantage in sensitivity and specificity using combinations of markers rather than a single marker [12,23].

Proteomic methods have previously been used to identify proteins associated with tubular injury. Nguyen and colleagues used SELDI (Surface Enhanced Laser Desorption Ionization) to identify proteins which increased in postoperative AKI prior to an increase in serum creatinine [24]. These markers have not been confirmed to be reproducible or accurate on samples that were not used to train the algorithm. In a rat model of sepsis-induced AKI, two dimensional gel electrophoresis with DIGE (Difference Gel Electrophoresis) was used to show that 30 urine proteins changed in abundance including the renal tubular brush border protein meprin 1-alpha [25]. Proteomic analysis of urinary exosomes has also been used to identify urine protein changes associated with tubular injury. Exosomes are apical membrane vesicles that are secreted into the urine which contain both membrane proteins and intracellular fluid [26,27]. Analysis of urinary exosomes by DIGE in two models of renal injury demonstrated proteins that were changed in abundance prior to an increase in serum creatinine [28]. One of the proteins, fetuin A was also increased in three ICU patients with acute kidney injury compared to ICU patients without renal injury. These studies have used proteomic methods to identify individual proteins that are changed with AKI. They have not systematically examined the accuracy of these markers in complicated patients with coexisting diseases. Combinations of markers will most likely be required for diagnostic markers in acutely ill patients with confounding illnesses. The objective of this study is to identify biomarkers that can differentiate between ATN and PRA. In this report, we separate patients with ATN from those with PRA by defined criteria and by following them over time. This approach allows us to accurately separate the two groups. We then use artificial neural networks (ANN), which are computer algorithms that can define a nonlinear association, to identify a combination of two spots from 2D gels that can differentiate between patients with ATN and PRA and test the accuracy of the algorithm in a different set of patients.

METHODS

Definition of Acute kidney injury

AKI was defined by an increase in serum creatinine of $\geq 50\%$ or 0.3 mg/dl over a period ≤ 48 hours. ATN was defined as AKI in a situation which could be expected to cause ATN that could not be attributed to another cause, did not respond to fluid hydration and had at least 2 of the following characteristics: 1) BUN:Cr ratio ≤ 20 ; 2) urinalysis with muddy brown granular and epithelial cell casts and free epithelial cells; 3) fractional excretion of sodium $> 1\%$, fractional excretion of uric acid $> 12\%$ or fractional excretion of urea $> 35\%$; 4) urine osmolality < 350 ; 5) rate of rise of serum creatinine > 0.3 mg/dl/day. PRA was defined as AKI in an appropriate clinical situation which could not be attributed to another cause, where urinary sediment did not show epithelial cell casts or free epithelial cells, responded to fluid hydration with improvement or no further worsening of serum creatinine and had at least two of the following characteristics: 1) BUN:Cr ratio ≥ 20 ; 2) fractional excretion of sodium $< 1\%$, fractional excretion of uric acid $< 12\%$ or fractional excretion of urea $< 35\%$; 3) urine osmolality > 500 mosm; 4) urine volume < 500 ml/day. In all cases we continued to follow the clinical course of the patients to confirm the diagnosis of ATN or PRA. Since the patients were followed over time we could differentiate the two conditions better than observations at a single time point. Patients were excluded for whom informed consent could not be obtained. In most cases, the investigators were made aware of the patients by the clinical Nephrology service. All patients who we identified with AKI that was caused by either ATN or PRA and from whom we obtained urine during the time from June 2003 to October 2004 were enrolled. Serum laboratory values were obtained from the clinical record. Fractional excretion of sodium was calculated as (urine sodium $\times$ serum creatinine)/(serum sodium $\times$ urine creatinine). Fractional excretion of urea or uric acid was calculated by replacing sodium in the formula with urea or uric acid.

Urine Collection and sample preparation

Urine was collected from patients at the Medical University of South Carolina or the Ralph H. Johnson VA Medical Center in accordance with a protocol approved by the appropriate institutional review boards. Random fresh urine was collected and processed immediately. We did not collect first morning voids because we attempted to collect the urine as soon as possible after recognition that the patient had AKI. In addition, many of the patients had bladder catheters. Cellular elements and debris were removed by centrifugation at $1000 \times g$ for 10 minutes. Samples were frozen without protease inhibitors at -80°C until they were prepared for two dimensional gel electrophoresis (2-DE). After thawing, an equal volume of acetone at 4°C was added to 10 ml of urine sample and incubated at 4°C for 10 minutes. The samples were centrifuged at $9000 \times g$ for 10 minutes and the precipitate was resuspended in 100 μl of a buffer containing 7 M urea, 2 M thiourea, 2% CHAPS and 1% Amidosulfobetaine-14 (ASB-14). Urine protein concentration was adjusted to 100 μg in 185 μl with a buffer containing 7 M urea, 2 M thiourea, 2% CHAPS, 1% ASB-14, 0.2% 3–10 ampholytes and 50 mM DTT. Two-dimensional electrophoresis was done as previously described over a pI range of 4–7 and on 8–16% gradient SDS-PAGE gels [29]. Gels were stained with Sypro Ruby (Molecular Probes) and imaged on an FX Pro Plus fluorescent imager (Bio-Rad). Gel images were matched into a set using PDQuest software. Protein spots were aligned across the 38 gels using PDQuest software (Bio-Rad). Automated spot alignment by the software was followed by manual confirmation and correction of alignments by an experienced user.

Data conversion, normalization and protein clustering with bioinformatic tools

We used an exploratory multivariate statistical cluster algorithm to look for general patterns of similarity. Protein abundance data was converted to a usable format by exporting the list of protein coordinates, matches and intensities from PDQuest as a text file. The text file was

transformed to an XML form called annotated gel markup language (AGML) as previously described [30]. This data structure enabled the assembly of a bioinformatics infrastructure, AGML central, which is a web-based open-source public infrastructure for dissemination of 2-DE data in XML format [31]. The AGML formatted structure describing the urine protein spot abundance was processed by code written in MATLAB (MathWorks Inc) for data analysis and graphic visualization. The spots were ranked by intensity and expressed as quantiles. This is a non-parametric approach that replaces spot intensity by its quantile within the individual gel [32]. Quantile rank rather than spot intensity was used for analysis. To determine if groupings of patients could be observed based on disease; we performed an exploratory analysis using multivariate statistical methods with code written making use of the MATLAB statistics toolbox. Bottom-up exploratory analyses was performed by unweighted pair group average (UPGA), using Euclidean distance between spot quantile vectors as a dissimilarity metric.

Analysis by artificial neural network to differentiate etiology of AKI

The protein spot intensities were ranked by intensity and expressed as quantiles. The 38 patients in the set were randomly sorted into one of three groups: training (18 patients), internal testing (10 patients) and external validation (10 patients). The random selection to each group was performed to keep a balanced proportion between the two outcomes, that is, independent selection from each of the two groups of patients. The identification of artificial neural network models was performed by Matlab code written along the guidelines previously proposed [33], which includes bootstrapped cross-validation as an early stop criteria, and screening for optimal topology. The predictive value of each spot was evaluated by sensitivity analysis by determining, for each i^{th} spot in each j^{th} gel/patient, $S_{i,j} = (d_{Oj}/d_{li,j}) \cdot (l_{i,j}/O_j)$. The TRAINING dataset was used to parameterize the ANN which was pursued with its own internal testing by cross-validation, with each run including an early stopping and topology optimization. A cross-validation scheme with leave 1/9th out was used where the median performing ANN was selected from each run. A more detailed description of the topology optimization, cross-validated, with early stopping and training procedure is described in [33].

The TESTING dataset was used strictly for variable addition – for all possible n-s variables to be added to a pool of s variables already selected from a total of n, the trained ANN that best predicted this second dataset would designate which variable to be selected by the s^{th} iteration of the variable addition procedure.

Finally, the EXTERNAL validation dataset was kept completely independent from both the training and variable selection procedures in order to have an unbiased assessment of the model predictability which was used to determine the point of the variable selection procedure at which we find the optimal set of variables. Comparison of the diagnostic results of the external set with those of the training and internal testing sets were only done after the training was completed.

The threshold value to discriminate ATN from PRA was determined in the training set and applied independently to the external validation set. Therefore, since the ANN was trained and the threshold value was set independently of the external validation set, the classification is unbiased. Outputs greater than the threshold predict a diagnosis of ATN and outputs less than the threshold predict PRA. Sensitivity was defined as True positive ATN/All patients with ATN and specificity was defined as True negative ATN/All patients without ATN. Overall accuracy was calculated in the external set by true discovery rate, that is, by all correct diagnoses/number of patients.

Protein identification

Protein spots were picked from the gel using a Proteome Works spot picking robot (Bio-Rad) and digested with trypsin as previously described [29]. Digests were concentrated using C18 ziptips (Millipore). Digests were initially analyzed on a Micromass MALDI-TOF mass spectrometer. Peptide mass fingerprinting was done using the Mascot search engine against the MSDB database with the following assumptions: Fixed carbamidomethyl modifications of cysteine residues, variable oxidation modifications of methionine and peptide mass tolerance ± 100 ppm. For the spots that could not be identified by MALDI-TOF, tryptic digests of gel plugs were loaded onto a 15 cm, 75 μ m internal diameter reversed phase C18 column (Microtech Scientific, Vista, Ca) for liquid chromatography/mass spectrometry analysis. Peptides were eluted with a gradient of 2–80% buffer B (acetonitrile/0.1% formic acid) directly into a Finnigan LTQ linear ion trap mass spectrometer in nanospray, positive ion mode. Full scans, zoom scans and MS/MS scans were done. Protein identification was done using the Turbo-SEQUEST algorithm in Bioworks 3.1 software and the MSDB database. Criteria for identification of peptides was XCorr >1.5 for singly charged ions, >2.0 for doubly charged and > 2.5 for triply charged ions.

RESULTS

Mean characteristics of the 38 patients are shown in Table 1. Although we were comparing urine protein differences between two diseases, the patients in the groups had very heterogeneous characteristics. Many coexisting diseases were present among patients in both groups. Table 2 shows the individual patients in each group with their renal disease, gender, urine specific gravity, serum creatinine, fractional excretion of sodium, coexisting disease processes, and whether or not they were septic at the time the urine was obtained. The table also shows whether they required dialysis or not. The patients were a diverse group and it was often difficult to tell exactly when the injury occurred. This is typical of the clinical presentation of AKI. Proteins were separated by 2-DE and 231 spots were aligned across the gels from the 38 patients. Abundance values of seven spots representing three proteins (α -1 antitrypsin, gelsolin and transferrin) were statistically different between the groups but none could consistently predict the etiology of the disease (data not shown).

Exploratory multivariate statistical analysis

To determine if unsupervised groupings of proteins was correlated with unrelated factors such as patient age, the amount of protein in the urine, the batch in which the sample was prepared and separated or the serum creatinine, we performed exploratory multivariate statistical analysis by the bottom-up approach of unsupervised simultaneous clustering of gels and spots by unweighted pair group average (UPGA). This approach did not reveal any clustering by any of the factors (Figure 1) and demonstrates that factors other than the cause of AKI or the other measured variables are causing most of the variability in spot intensity on the gels.

Artificial Neural Network analysis

We used a machine learning algorithm (ANN) to identify nonlinear patterns that could differentiate the diseases. The rank ordered spot intensity data from the 38 patients was used to train the ANN. Networks were repetitively trained with the training set and iteratively checked against the test set. A combination of two markers had the maximum predictive independent performance. The two spots are shown in Figure 2.

Evaluation of the output of the ANN in the novel set of patients

To test the ability of the algorithm derived from the ANN to identify the etiology of AKI, we evaluated the output of the ANN in the external validation set which was not used for training

the algorithm. Table 3 shows the output values of individual patients in each of the three sets. The table compares the observed diagnosis (0 for PRA or 1 for ATN) to the predicted diagnosis value for each of the three sets (a number between 0 and 1). The output value is assigned by the ANN and represents the similarity of the data from each patient to data in the training set. A segmentation threshold value at which a diagnosis of ATN or PRA was made was set at 0.21 based on the training set. Since this value was selected in the training set and applied to the external set, no violation of the independent assessment for the external dataset takes place. By comparing the output value to the threshold value, a predicted diagnosis is obtained. The diagnoses for 16/18 patients in the training set and for all 10 patients in the internal test set were correctly identified. Most importantly, all 5 of the patients with ATN in the external validation set were correctly diagnosed (100% sensitivity) as were 4 of 5 patients with PRA (80% sensitivity). An ROC curve was generated for the group of 10 patients in the external validation set where a total area under the curve (AUC) of 0.88 was observed. The 90% accuracy and high AUC value in the validation set indicate a very good quality test since the validation set was not used in either training or variable selection for the algorithm. The algorithm was not able to predict death or dialysis in these patients.

Relationship of diagnostic protein spots

The two protein spots that were identified as markers were spot numbers 40 (pI 4.9, Mr 13, Spot 2103 in AGML) and 133 (pI 5.4, Mr 20, Spot 5305 in AGML). Figure 3 plots the quantile ranks for intensity of the spots in the validation set. The correlation of protein rank abundances for the two spots with the disease is nonlinear, more specifically it indicates an XOR (exclusive OR) interdependency, where either high or low values of spot 133 can correlate with presence of ATN depending on the value of spot 40. Intermediate abundances of spot 133 are associated with PRA only for low values of spot 40. The abundance of both proteins spots must be known to differentiate between the two outcomes. The XOR distribution of intensities observed also suggests a possible reason that linear approaches to identify biomarkers failed: the ANN predictor is relying on inter-dependencies that would not be captured by linear discriminant analysis tools.

After determining that analysis of abundance of two spots on the gel could differentiate between ATN and PRA, we attempted to identify the proteins that make up those spots. Digested spots were analyzed by MALDI-TOF MS and MALDI-TOF/TOF tandem MS. Spot 133 was identified as PRBP by peptide mass fingerprinting with a MASCOT score of 81 (> 63 significant, $p < 0.05$). Ten peptides were matched from 43 searched. 53% of the protein was covered by the expected peptides (figure 4A). To confirm the identification, peptide sequencing was done by MALDI-TOF/TOF MS. Based on three peptide ions sequenced (two were overlapping), the protein was again identified as PRBP with a total ions score of 167 where a total ion score of 60 is considered significant. Bold sequence in figure 4A shows sequenced peptides.

Spot 40 was not identified by MALDI-TOF mass spectrometry with peptide mass fingerprinting nor by tandem mass spectrometry using the MALDI-TOF/TOF mass spectrometer. To identify this spot, we took advantage of the higher sensitivity of the linear ion trap mass spectrometer. We identified sequence from two different proteins in the gel plug. One peptide from PRBP was sequenced from both scan 1755 (XCorr 2.532, charge 2) and scan 1769 (XCorr 2.509, charge 2). The precursor ion had a mass of 1199.38 (YWGVSFLQK). Two different peptides were sequenced which mapped to albumin protein. They were an ion from scan 1324 (XCorr 2.968, charge 3) with a precursor mass of 1640.91 (KVPQVSTPTLVEVSR) and an ion from scan 1654 (XCorr 2.263, charge 2) with a precursor mass of 1312.54 (HPDYSVLLLR). Since the molecular weight of the original protein spot was 13 kDa and the molecular weights of PRBP and albumin are 21 and 66 kDa, the proteins

identified in the spot are likely fragments. To confirm that these proteins were in the gel spot and to refine the likely sequence of the peptide fragments, we compared the spectrum obtained from MALDI-TOF analysis of the tryptic peptides to predicted tryptic peptides from the two proteins (Figure 5 and Table 4). All but six of the sixteen most abundant peptides in the MALDI spectrum could be accounted for by either PRBP or albumin or by the contaminant keratin or autolysis of trypsin. Three peptides were identified whose masses matched to predicted tryptic fragments of PRBP. The peptides covered a region of PRBP extending from amino acid residue 90–139 of the circulating form of PRBP (Figure 4B). All 50 amino acids of this fragment were accounted for yielding 100% coverage. The predicted mass of this peptide is 5.8 kDa. Since the gel plug was cut from a position on the gel corresponding to 13 kDa, this sequence must represent only a portion of the sequence of the protein fragment in the spot. Four peptides were identified whose mass matched to predicted tryptic peptides from a fragment of serum albumin extending from 324–428 (Figure 4C). The predicted size of the fragment is 12.2 kDa which correlates well with the position of the protein on the gel implying that all or most of this fragment can be accounted for by the postulated sequence.

DISCUSSION

The two most common definitions of AKI are the RIFLE and the AKIN definitions. The change in creatinine criteria for the first stage of RIFLE is an increase in serum creatinine of at least 50% and the first stage AKIN is an increase of 50% or 0.3 mg/dl. All patients in this study met the criteria for at least the first stage of both AKI criteria. ATN and PRA are the two most common causes of AKI. Differentiating between these two causes is a common clinical problem. Tools such as FENa that are most commonly used are not reliable. Using artificial neural network analysis we identified a set of two markers that can differentiate between ATN and PRA 90% of the time in patients that were not used to train the network. Not only was more than one marker required for the diagnostic ability, but the pattern of spot rank abundances for the two spots suggests a plausible explanation why data analysis methods that rely on assumptions of linear independence have been inefficient at uncovering biomarkers. The predictive association relies on the covariance between the two candidate markers, and specifically on a particularly complicated type of interdependency, the Exclusive OR relationship (XOR), a well established litmus test for machine learning tools [34]. This study demonstrates that urine protein abundance can be used to predict the cause of AKI. We have recently used artificial neural network analysis of urine protein abundances to identify biomarkers of glomerular diseases [23]. In that study of patients with biopsy proven diseases, we showed that altered glomerular permeability to glycosylated proteins can be used to predict the cause of the glomerular disease. The current study extends these findings to demonstrate that urine proteins can be used to predict the cause of AKI.

We did not identify several previously suggested candidate proteins for the occurrence of AKI such as NGAL and KIM-1 because they were not observed to be differentially expressed in our 2D gels. The proteins we identified in the diagnostic spots were PRBP and a spot that contained fragments of PRBP and albumin. PRBP has been suggested previously as a marker for renal tubular dysfunction [10,35,36]. As we found for a number of potential protein markers, however, levels of PRBP alone had too much overlap between tubular and non tubular diseases to be useful. In fact, mean abundances of PRBP were not statistically different between patients with the two diseases. PRBP was diagnostic only when it was analyzed in a complex algorithm with the abundance of a second spot. As shown in figure 3, the relationship of spot abundance to the disease was not linear. In fact, either high or low levels of plasma retinol-binding protein were associated with a diagnosis of ATN which only become informative when the abundance of the albumin component marker is also known. Intermediate levels were associated with a diagnosis of PRA. We propose an hypothesis for the reason that these urine proteins change in this way. PRBP and albumin are filtered at the glomerulus and reabsorbed and metabolized

in the renal proximal tubule. When the tubules are dysfunctional, as in ATN, larger amounts of PRBP will appear in the urine. In severe ATN the fall in GFR is greater than that seen in PRA, therefore filtration of plasma proteins declines leading to a lower level in the urine. The value of the abundance of the second spot which is a fragment of a filtered serum protein is required to interpret the value for PRBP and predict the diagnosis. Since there may be a transition state between ATN and PRA this relationship may not always hold up diagnostically.

A frequent problem in interpretation of urine protein abundance is how to normalize protein abundance in the urine. Several methods have been proposed including loading equal amounts of protein, adjusting for the amount of creatinine in the sample or for the volume of urine. In the ANN analysis we normalized protein abundance by assigning proteins a quantile value based on the rank of their abundance within the gel. The most abundant protein has a rank of 1 and the least abundant has a rank of 0. We have recently shown that this nonparametric method of normalization to be particularly advantageous for gel electrophoresis data where strong variability in the staining and exposure procedures are compounded by a spot identification error [32]. These data show that it is also a useful method for normalizing urine protein abundance because it expresses protein abundance as it relates to other proteins in the sample.

These studies have defined candidate markers that can be used to differentiate ATN from PRA. The findings were confirmed in a set of ten patients that were not used to train the algorithm demonstrating the validity of the test. Further validation using samples obtained from other centers will be necessary to confirm the usefulness of these markers.

Acknowledgments

This project has been funded by R01DK080234 and with Federal funds as part of the NHLBI Proteomics Initiative from the National Heart, Lung, and Blood Institute, National Institutes of Health, under Contract No. N01-HV-28181. Additional support for this project came from the Department of Veterans Affairs. 2D gel electrophoresis was done by Alison Bland, Lou D'Eugenio and Melissa Dugan. Tim Taylor assisted with sample collection and preparation. Tandem mass spectrometry was done in the MUSC mass spectrometry facility with the assistance of Dr. Kevin Schey and Jennifer Bethard. We are grateful to the Nephrology fellows at MUSC for help with identification of patients and sample collection.

References

- (1). Obialo CI, Okonofua EC, Tayade AS, Riley LJ. Epidemiology of De Novo Acute Renal Failure in Hospitalized African Americans: Comparing Community-Acquired Vs Hospital-Acquired Disease. *Arch. Intern. Med* 2000;160:1309–1313. [PubMed: 10809034]
- (2). Brivet FG, Kleinknecht DJ, Loirat P, Landais PJ. Acute Renal Failure in Intensive Care Units--Causes, Outcome, and Prognostic Factors of Hospital Mortality; a Prospective, Multicenter Study. French Study Group on Acute Renal Failure. *Crit Care Med* 1996;24:192–198. [PubMed: 8605788]
- (3). Liano F, Pascual J. Epidemiology of Acute Renal Failure: a Prospective, Multicenter, Community-Based Study. Madrid Acute Renal Failure Study Group. *Kidney Int* 1996;50:811–818. [PubMed: 8872955]
- (4). Miller TR, Anderson RJ, Linas SL, Henrich WL, Berns AS, Gabow PA, Schrier RW. Urinary Diagnostic Indices in Acute Renal Failure: a Prospective Study. *Ann. Intern. Med* 1978;89:47–50. [PubMed: 666184]
- (5). Vaz AJ. Low Fractional Excretion of Urine Sodium in Acute Renal Failure Due to Sepsis. *Arch. Intern. Med* 1983;143:738–739. [PubMed: 6838295]
- (6). Steinhauslin F, Burnier M, Magnin JL, Munafo A, Buclin T, Diezi J, Biollaz J. Fractional Excretion of Trace Lithium and Uric Acid in Acute Renal Failure. *J Am Soc Nephrol* 1994;4:1429–1437. [PubMed: 8161725]
- (7). Zarich S, Fang LS, Diamond JR. Fractional Excretion of Sodium. Exceptions to Its Diagnostic Value. *Arch. Intern. Med* 1985;145:108–112. [PubMed: 3970621]

- (8). Werb R, Linton AL. Aetiology, Diagnosis, Treatment and Prognosis of Acute Renal Failure in an Intensive Care Unit. *Resuscitation* 1979;7:95–100. [PubMed: 542732]
- (9). Lapsley M, Flynn FV, Sansom PA. Beta 2-Glycoprotein-1 (Apolipoprotein H) Excretion and Renal Tubular Malfunction in Diabetic Patients Without Clinical Proteinuria. *J Clin. Pathol* 1993;46:465–469. [PubMed: 8320329]
- (10). Flynn FV, Lapsley M, Sansom PA, Cohen SL. Urinary Excretion of Beta 2-Glycoprotein-1 (Apolipoprotein H) and Other Markers of Tubular Malfunction in “Non-Tubular” Renal Disease. *J Clin. Pathol* 1992;45:561–567. [PubMed: 1381383]
- (11). duCheyron D, Daubin C, Poggioli J, Ramakers M, Houillier P, Charbonneau P, Paillard M. Urinary Measurement of Na⁺/H⁺ Exchanger Isoform 3 (NHE3) Protein As New Marker of Tubule Injury in Critically Ill Patients With ARF. *Am J Kidney Dis* 2003;42:497–506. [PubMed: 12955677]
- (12). Liangos O, Perianayagam MC, Vaidya VS, Han WK, Wald R, Tighiouart H, Mackinnon RW, Li L, Balakrishnan VS, Pereira BJ, Bonventre JV, Jaber BL. Urinary N-Acetyl-(Beta)-(D)-Glucosaminidase Activity and Kidney Injury Molecule-1 Level Are Associated With Adverse Outcomes in Acute Renal Failure. *J Am Soc Nephrol* 2007;18:904–912. [PubMed: 17267747]
- (13). Westhuyzen J, Endre ZH, Reece G, Reith DM, Saltissi D, Morgan TJ. Measurement of Tubular Enzymuria Facilitates Early Detection of Acute Renal Impairment in the Intensive Care Unit. *Nephrol Dial. Transplant* 2003;18:543–551. [PubMed: 12584277]
- (14). Mishra J, Ma Q, Prada A, Mitsnefes M, Zahedi K, Yang J, Barasch J, Devarajan P. Identification of Neutrophil Gelatinase-Associated Lipocalin As a Novel Early Urinary Biomarker for Ischemic Renal Injury. *J Am Soc Nephrol* 2003;14:2534–2543. [PubMed: 14514731]
- (15). Rabb H. Novel Urinary Markers for Early Diagnosis of ARF. *Am J Kidney Dis* 2003;42:599–600. [PubMed: 12955692]
- (16). Han WK, Bailly V, Abichandani R, Thadhani R, Bonventre JV. Kidney Injury Molecule-1 (KIM-1): a Novel Biomarker for Human Renal Proximal Tubule Injury. *Kidney Int* 2002;62:237–244. [PubMed: 12081583]
- (17). Ichimura T, Hung CC, Yang SA, Stevens JL, Bonventre JV. Kidney Injury Molecule-1: a Tissue and Urinary Biomarker for Nephrotoxicant-Induced Renal Injury. *Am J Physiol Renal Physiol* 2004;286:F552–F563. [PubMed: 14600030]
- (18). Parikh CR, Jani A, Melnikov VY, Faubel S, Edelstein CL. Urinary Interleukin-18 Is a Marker of Human Acute Tubular Necrosis. *Am J Kidney Dis* 2004;43:405–414. [PubMed: 14981598]
- (19). Nickolas TL, O'Rourke MJ, Yang J, Sise ME, Canetta PA, Barasch N, Buchen C, Khan F, Mori K, Giglio J, Devarajan P, Barasch J. Sensitivity and Specificity of a Single Emergency Department Measurement of Urinary Neutrophil Gelatinase-Associated Lipocalin for Diagnosing Acute Kidney Injury. *Ann. Intern. Med* 2008;148:810–819. [PubMed: 18519927]
- (20). Janech MG, Raymond JR, Arthur JM. Proteomics in Renal Research. *Am J Physiol Renal Physiol* 2007;292:F501–F512. [PubMed: 17107941]
- (21). Pang JX, Ginanni N, Dongre AR, Hefta SA, Opitek GJ. Biomarker Discovery in Urine by Proteomics. *J Proteome. Res* 2002;1:161–169. [PubMed: 12643536]
- (22). Thongboonkerd V, Malasit P. Renal and Urinary Proteomics: Current Applications and Challenges. *Proteomics* 2005;5:1033–1042. [PubMed: 15669002]
- (23). Varghese SA, Powell TB, Budisavljevic MN, Oates JC, Raymond JR, Almeida JS, Arthur JM. Urine Biomarkers Predict the Cause of Glomerular Disease. *J Am Soc Nephrol* 2007;18:913–922. [PubMed: 17301191]
- (24). Nguyen MT, Ross GF, Dent CL, Devarajan P. Early Prediction of Acute Renal Injury Using Urinary Proteomics. *Am. J Nephrol* 2005;25:318–326. [PubMed: 15961952]
- (25). Holly MK, Dear JW, Hu X, Schechter AN, Gladwin MT, Hewitt SM, Yuen PS, Star RA. Biomarker and Drug-Target Discovery Using Proteomics in a New Rat Model of Sepsis-Induced Acute Renal Failure. *Kidney Int* 2006;70:496–506. [PubMed: 16760904]
- (26). Pisitkun T, Shen RF, Knepper MA. Identification and Proteomic Profiling of Exosomes in Human Urine. *Proc. Natl. Acad. Sci. U. S. A* 2004;101:13368–13373. [PubMed: 15326289]
- (27). Zhou H, Yuen PS, Pisitkun T, Gonzales PA, Yasuda H, Dear JW, Gross P, Knepper MA, Star RA. Collection, Storage, Preservation, and Normalization of Human Urinary Exosomes for Biomarker Discovery. *Kidney Int* 2006;69:1471–1476. [PubMed: 16501490]

- (28). Zhou H, Pisitkun T, Aponte A, Yuen PS, Hoffert JD, Yasuda H, Hu X, Chawla L, Shen RF, Knepper MA, Star RA. Exosomal Fetuin-A Identified by Proteomics: a Novel Urinary Biomarker for Detecting Acute Kidney Injury. *Kidney Int* 2006;70:1847–1857. [PubMed: 17021608]
- (29). Lefler DM, Pafford RG, Black NA, Raymond JR, Arthur JM. Identification of Proteins in Slow Continuous Ultrafiltrate by Reversed-Phase Chromatography and Proteomics. *J. Proteome. Res* 2004;3:1254–1260. [PubMed: 15595735]
- (30). Stanislaus R, Jiang LH, Swartz M, Arthur J, Almeida JS. An XML Standard for the Dissemination of Annotated 2D Gel Electrophoresis Data Complemented With Mass Spectrometry Results. *BMC. Bioinformatics* 2004;5:9. [PubMed: 15005801]
- (31). Stanislaus R, Chen C, Franklin J, Arthur J, Almeida JS. AGML Central: Web Based Gel Proteomic Infrastructure. *Bioinformatics* 2005;21:1754–1757. [PubMed: 15647304]
- (32). Almeida JS, Stanislaus R, Krug EL, Arthur JM. Normalization and Analysis of Residual Variation in 2D Gel Electrophoresis for Quantitative Differential Proteomics. *Proteomics* 2005;5:1242–1249. [PubMed: 15732138]
- (33). Almeida JS. Predictive Non-Linear Modeling of Complex Data by Artificial Neural Networks. *Curr. Opin. Biotechnol* 2002;13:72–76. [PubMed: 11849962]
- (34). Minsky, M.; Papert, PA. *Perceptrons*. MIT Press; Cambridge: 1969.
- (35). Nedelkov D, Nelson RW. Analysis of Human Urine Protein Biomarkers Via Biomolecular Interaction Analysis Mass Spectrometry. *Am J Kidney Dis* 2001;38:481–487. [PubMed: 11532678]
- (36). Sarasua SM, Mueller P, Kathman S, Campagna D, Uddin MS, White MC. Confirming the Utility of Four Kidney Biomarker Tests in a Longitudinal Follow-Up Study. *Ren Fail* 2003;25:797–817. [PubMed: 14575288]

Figure 1. Unsupervised simultaneous clustering of gels and spots by UPGA. Individual gels (patients) are shown on the y axis. Spots are shown on the x axis. Protein intensity is shown in the color map. UPGA did not reveal any consistent patterns of similarity between patients with the same disease, age, urine protein concentration, run batch for 2-DE or serum creatinine concentration.

Figure 2.

Two dimensional gel electrophoresis of urine protein from a patient with ATN. The two spots necessary for predicting the cause of AKI are shown. Two proteins were identified in spot 40 is a fragment of albumin and a fragment of plasma retinol binding protein. Spot 133 is intact plasma retinol binding protein.

Figure 3.

Plot of quantile ranks for two spots which can distinguish between ATN and PRA. X and Y planes show quantile ranks for spot 133 (PRBP) and 40 (fragments of PRBP and albumin). Spots in the shaded area on the graph predicts PRA and in the white area predicts ATN. The prediction algorithm was derived from the training set. Only patients in the external validation set are plotted on the graph. All patients were predicted correctly except the patient with PRA at the far left of the figure.

A. Spot 133

1 ERDCRVSSFR VKENFDKARF **SGTWYAMAK**K DPEGLFLQDN IVAEFSVDET GQMSATAKGR
 61 VRLLNNWDVC ADMVGTFDT EDPKFK**MKY WGVASFLQK**G NDDHWIVDTD YDTYAVQYSC
 121 RLLNLDGTCA DSYSFVFSRD PNGLPPEAQK IVRQRQEELC LARQYRLIVH NGYCDGRSER
 181 NL

B. Spot 40, Protein 1

1 **YWGVASFLQK** GNDDHWIVDT DYDTYAVQYS CRLLNLDGTC ADSYSFVFSR

C. Spot 40, Protein 2

1 DVFLGMFLYE YARR**HPDYSV VLLLR**LAKTY ETTLEKCAA ADPHECYAKV FDEFKPLVEE
 61 PQNLIKQNC LFEQLGEYKF QNALLVRYTK **KVPQVSTPTL VEVSR**

Figure 4.

Match of peptides and sequence to identified proteins. Underlined sequence was matched to predicted peptides from MALDI-TOF spectrum. Red sequence was identified by tandem mass spectrometry. A. Spot 133 was identified as Plasma retinol binding protein. 53% of the intact protein was covered by the assigned peptides. B. Spot 40, Protein 1. An internal sequence of plasma retinol binding protein was identified by tandem MS sequence of one peptide and predicted peptide masses of three peptides. The corresponding peptides covered 100% of a 5.8 kDa protein fragment of plasma retinol binding protein. C. Spot 40, protein 2. An internal sequence of serum albumin was identified by tandem MS sequence of two peptides and predicted masses of 4 peptides. The corresponding peptides covered 50% of a 12.2 kDa protein fragment of serum albumin.

Figure 5.

MALDI-TOF MS spectrum of tryptic digest of protein spot 40. Blue peaks match to a fragment of plasma retinol binding protein. Green peaks match to a fragment of albumin. Ten of the 12 most abundant peptides in this view can be attributed to one of the two proteins or a known contaminant. Assignment of the masses to the appropriate sequence is shown in figure 4B and C and in table 4.

Table 1

Mean characteristics of groups

	ATN	PRA	p value
Patients	19	19	
Mean age	52±4	60±4	0.08
M/F	13/6	10/9	0.33
Caucasian/AA	10/9	8/11	0.52
Serum Creatinine	4.4±0.5	2.1±0.2	<0.001
Baseline Creatinine	1.0±0.10	1.1±0.09	0.50

Table 2

Individual characteristics of patients

Gel name	DISEASE	SEX	S.Gravity	S. Creat	FeNa	Coexisting Diseases	Septic	Dialysis
ATN1_34	ATN	Male	1.020	5.9	13.0	Situs inversus, Afib, Hypotension	No	Yes
ATN2_78	ATN	Male	1.005	3.0	2.5	Cellulitis, Elevated Creatinine kinase	No	No
ATN3_79	ATN	Male	1.030	4.3	1.0	Diabetes, lactic acidosis, Hypotension	No	Yes
ATN4_142	ATN	Male	1.030	3.1	1.3	HIV, Pneumonia, Hypotension	Yes	No
ATN5_24	ATN	Female	1.015	8.1	9.7	Diabetes, Acute on chronic renal failure, gangrene	Yes	No
ATN6_89	ATN	Female	1.025	2.0	1.1	Gastric bypass, sepsis	Yes	No
ATN7_114	ATN	Female	1.020	4.0	1.4	Post CABG kidney injury	No	No
ATN8_366	ATN	Male	1.030	3.2	1.3	Unknown injury, biopsy proven ATN	No	No
ATN9_314	ATN	Female	1.025	8.9	6.7	Liver transplant, seizures	No	Yes
ATN10_485	ATN	Male	1.010	4.3	5.0	Sickle Cell disease, Hypotension	No	No
ATN11_329	ATN	Male	1.025	2.9	1.1	Motor Vehicle Accident, Hypotension	No	No
ATN12_321	ATN	Female	1.005	2.4	3.4	Short gut syndrome, Sepsis	Yes	No
ATN13_354	ATN	Male	1.010	6.3	9.6	HIV, Colitis, sepsis	Yes	No
ATN14_487	ATN	Female	1.030	2.4	0.7	Motor Vehicle Accident, Contrast Injury, Heart Failure	No	Yes
ATN15_509	ATN	Male	1.010	5.6	3.3	Sepsis, Contrast Administration	Yes	Yes
ATN16_510	ATN	Male	1.030	2.4	1.1	CHF, Hypotension	No	No
ATN17_30	ATN	Male	1.015	4.2	0.7	Cardiac arrest, Diabetes	No	No
ATN18_223	ATN	Male	1.030	6.5	0.7	Post Aneurysm repair kidney injury	No	Yes
ATN19_116	ATN	Male	1.030	4.4	0.3	Respiratory arrest, GI Bleed	No	Yes
PRA1_54	PRA	Female	1.015	1.8	0.6	Pregnant, Fungal sepsis, improved with hydration	Yes	No
PRA2_115	PRA	Male	1.025	1.9	0.1	CHF	No	No
PRA3_189	PRA	Female	1.030	1.6	0.1	CHF, Afib, Hypothyroid	No	No
PRA4_353	PRA	Female	1.020	2.5	0.4	Diabetes, volume depletion, Ace inhibitor treatment	No	No
PRA5_272	PRA	Female	1.025	2.2	0.7	Diabetes, Non-small cell lung ca	No	No
PRA6_378	PRA	Male	1.020	3.2	2.0	CHF, Cellulitis, Diabetes	No	No
PRA7_415	PRA	Male	1.025	2.6	0.2	CHF, Sarcoidosis, COPD, Diabetes	No	Yes
PRA8_417	PRA	Male	1.010	1.0	0.1	Volume Depletion, Diabetes	No	No

Gel name	DISEASE	SEX	S.Gravity	S. Creat	FeNa	Coexisting Diseases	Septic	Dialysis
PRA9_488	PRA	Male	1.030	1.0	0.5	Motor Vehicle Accident, Fever	No	No
PRA10_37	PRA	Female	1.030	3.5	0.4	Hip Replacement surgery, Volume depletion	No	No
PRA11_413	PRA	Male	1.015	1.8	0.5	Stroke, Alcohol abuse	No	No
PRA12_412	PRA	Female	1.025	1.7	0.1	Uterine Leiomyosarcoma, Nephrolithiasis	No	No
PRA13_506	PRA	Female	1.030	2.4	0.3	Gastric Bypass	No	No
PRA14_507	PRA	Female	1.010	2.6	0.5	Diabetes, CHF, Aggressive Diuresis	No	No
PRA15_511	PRA	Male	1.030	1.9	0.2	Small cell lung Ca, Gout	No	No
PRA16_53	PRA	Male	1.015	2.2	3.0	Adrenal insufficiency, seizure disorder, hypotension	No	No
PRA17_38	PRA	Male	1.020	2.5	2.5	Infected knee replacement, Diabetes	No	No
PRA18_184	PRA	Female	1.020	2.1	6.1	Post operative CABG, Diabetes, Diuresis	No	No
PRA19_496	PRA	Male	1.010	1.6	6.4	Acute MI, VFib, CHF, aggressive diuresis	No	No

Table 3

Observed and predicted diagnoses for cause of AKI.

Training Set			Test Set			External Set		
Observed Diagnosis	ANN Output	Predicted Diagnosis	Observed Diagnosis	ANN Output	Predicted Diagnosis	Observed Diagnosis	ANN Output	Predicted Diagnosis
PRA	0.010	PRA	PRA	0.036	PRA	PRA	0.100	PRA
PRA	0.006	PRA	PRA	0.035	PRA	PRA	0.061	PRA
PRA	0.029	PRA	PRA	0.011	PRA	PRA	0.000	PRA
PRA	0.071	PRA	PRA	0.038	PRA	PRA	0.074	PRA
PRA	0.206	PRA	PRA	0.068	PRA	PRA	0.992	ATN
PRA	0.952	ATN	ATN	0.995	ATN	ATN	0.847	ATN
PRA	0.003	PRA	ATN	0.998	ATN	ATN	1.000	ATN
PRA	0.470	ATN	ATN	0.998	ATN	ATN	0.673	ATN
PRA	0.016	PRA	ATN	0.989	ATN	ATN	0.966	ATN
ATN	0.230	ATN	ATN	1.000	ATN	ATN	0.997	ATN
ATN	0.995	ATN						
ATN	0.996	ATN						
ATN	0.989	ATN						
ATN	0.999	ATN						
ATN	1.000	ATN						
ATN	0.974	ATN						
ATN	0.994	ATN						
ATN	1.000	ATN						
Correct: 16/18			Correct: 10/10			Correct: 9/10		

Table 4

Major peptide masses from MALDI-TOF spectrum of digested proteins in spot 40 displayed in order of the abundance of the monoisotopic peak. The MALDI-TOF spectrum from spot 40 is shown in figure 5.

Measured Mass	Protein	Calculated mass	Mass error (ppm)	Alignment to Protein Fragment	Predicted sequence
2065.2366	RBP	2064.9695	129	122-139	LLNLDGTCADSYSFVFSR
1198.8263	RBP	1198.6255	167	90-99	YWGVSFLQK
1311.9508	Albumin	1311.7419	159	338-348	HPDYSVVLLLR
1624.0266	Albumin	1623.9580	42	324-336	DVFLGMFLYEYAR
2693.4932	RBP	2693.1208	138	100-121	GNDDHWIVDTDYDTYAVQYSCR
1220.8237	Unknown				
2272.2358	Keratin	2272.0400	86		
2211.3911	Trypsin	2211.0900	136		
2808.7109	Unknown				
1236.8269	Unknown				
1468.0714	Albumin	1467.8430	156	337-348	RHPDYSVVLLLR
2070.1919	Unknown				
1302.8707	Keratin	1302.7150	120		
2225.3914	Unknown				
2290.2295	Unknown				
1640.0530	Albumin	1639.9377	70	414-428	KVPQVSTPTLVEVSR