A database of naturally occurring human urinary peptides and proteins for use in clinical applications

[#]Joshua J. Coon[1,2], [#]Petra Zürbig[3*], Hartwig W. Bauer[4], Georg Behrens[5], Mohammed Dakna[6], Anna F. Dominiczak[7], Stéphane Decramer[7,8], Jochen H. H. Ehrich[9], Danilo Fliser[10,+], Moritz Frommberger[11], Arnold Ganser[12], Mark A. Girolami[13], Igor Golovko[3], David M. Good[1], Wilfried Gwinner[10], Marion Haubitz[10], Stefan Herget-Rosenthal[14,+], Holger Jahn[15], George Jerums[16], Bruce A. Julian[17], Markus Kellmann[18], Volker Kliem[19], Andrzej Krolewski[20,21], Walter Kolch[22], Mario Luppi[23], Michael Melter[24], Ziad Massy[25,+], Christian Neusüss[26], Jan Novak[16], Karlheinz Peter[27], Kasper Rossing[28], Harald Rupprecht[29], Joost P. Schanstra[7], Eric Schiffer[3], Jens-Uwe Stolzenburg[30], Lise Tarnow[28], Dan Theodorescu[31], Visith Thongboonkerd[32], Raymond Vanholder[33,+], Eva M. Weissinger[12,+], Harald Mischak[3,+], Philippe Schmitt-Kopplin[11]

[1] Department of Chemistry, University of Wisconsin-Madison, Madison, WI, USA

[2] Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

[3] Mosaiques diagnostics & therapeutics, Hannover, Germany

[4] Ludwig-Maximilians-University Munich, Munich, Germany

[5] Department of Clinical Immunology, Hannover Medical School, Hannover, Germany

[6] BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, UK

[7] Renal and cardiac remodeling, Inserm, U858/I2MR, Toulouse, France

[8] Pediatric Nephrology Unit, Hopital des Enfants, CHU de Toulouse, Toulouse, France

[9] Children's Hospital, Hannover Medical School, Hannover, Germany

[10] Department of Nephrology, Hannover Medical School, Hannover, Germany

[11] Institute for Ecological Chemistry, GSF - National Research Centre for Environment and Health, Neuherberg, Germany

[12] Department of Hematology, Hemostasis and Oncology, Hannover Medical School, Hannover, Germany

[13] Department of Computing Science, University of Glasgow, Glasgow, UK

[14] Department of Nephrology, University Hospital, University Duisburg-Essen, Germany

[15] Department of Psychiatry, University Hospital Hamburg-Eppendorf, Hamburg, Germany

[16] Department of Medicine, Austin Health & Northern Health, University of Melbourne, Melbourne, Australia

[17] University of Alabama at Birmingham, Birmingham, AL, USA

[18] Thermo Fisher Scientific, Bremen, Germany

[19] Department of Nephrology Lower, Transplantation Centre, Saxony Centre for Nephrology, Hann. Muenden, Germany

[20] Section on Genetics and Epidemiology, Joslin Diabetes Center, Boston, MA, USA

[21] Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

[22] The Beatson Institute for Cancer Research & Sir Henry Wellcome Functional Genomics Facility, University of Glasgow, Glasgow, UK

[23] Department of Oncology and Hematology, University of Modena and Reggio Emilia, Modena, Italy

[24] University of Regensburg, Children's Hospital, Regensburg, Germany

[25] INSERM, ERI-12, and Amiens University Hospital, UPJV, Amiens, France

[26] University of Aalen, Aalen, Germany

[27] Baker Heart Research Institute, Melbourne, Australia

[28] Steno Diabetes Centre, Gentofte, Denmark

[29] Department of Nephrology, Clinical Centre Bayreuth, Bayreuth, Germany

[30] Department of Diagnostic Radiology, University of Leipzig, Leipzig, Germany

[31] University of Virginia, Charlottesville, VA, USA

[32] Faculty of Medicine at Siriraj Hospital, Mahidol University, Bangkok, Thailand

[33] Department of Internal Medicine, University Hospital, Ghent, Belgium

[#]These authors have contributed equally to this work.

[+]On behalf of the EUTox consortium

[*]Corresponding author at:

Mosaiques diagnostics & therapeutics AG

Mellendorfer Str. 7-9

30625 Hannover

Germany

Tel: +49 511 554744 15

Fax: +49 511 554744 31

E-mail: zuerbig@mosaiques-diagnostics.com

SUMMARY

Owing to its availability, ease of collection and correlation with (patho-) physiology, urine is an attractive source for clinical proteomics. However, the lack of comparable datasets from large cohorts has greatly hindered development in this field. Here we report the establishment of a high resolution proteome database of naturally occurring human urinary peptides and proteins - ranging from 800–17,000 Da - from over 3,600 individual samples using capillary electrophoresis coupled to mass spectrometry, yielding an average of 1,500 peptides per sample. All processed data were deposited in an SQL database, currently containing 5,010 relevant unique urinary peptides that serve as classifiers for diagnosis and monitoring of diseases, including kidney and vascular diseases. Of these, 352 have been sequenced to date. To demonstrate the applicability of this database, two examples of disease diagnosis were provided: For renal damage diagnosis, patients with a specific renal disease were identified with high specificity and sensitivity in a blinded cohort of 131 individuals. We further show definition of biomarkers specific for immunosuppression and complications after transplantation (Kaposi‘s sarcoma). Due to its high information content, this database will be a powerful tool for the validation of biomarkers for both renal and non-renal diseases.

INTRODUCTION

Analysis of urine plays a central role in clinical diagnostics. The human urinary proteome/peptidome has been extensively investigated to gain information on disease processes affecting the kidney and the urogenital tract[1-3]. Urinary proteins and peptides originate not only from glomerular filtration, but also from tubular secretion, epithelial cells shed from the kidney and urinary tract, secreted exosomes, and semen[4-6]. Thus, in principle, urine is a rich source of biomarkers for a wide range of diseases due to specific changes in its proteome/peptidome[7-10]. To test the feasibility of urinary proteomics as a non-invasive diagnostic tool, large-scale studies are needed to analyze urine with reliable and quantitative methods. A variety of techniques have been applied to this effort, including two-dimensional electrophoresis combined with mass spectrometric (2-DE-MS) and/or immunochemical identification of proteins[11-13], liquid chromatography coupled to mass spectrometry (LC-MS)[14, 15], and surface-enhanced laser desorption ionization mass spectrometry (SELDI-MS)[16].

Due mostly to technical challenges, studies of urine based on proteomic methods often included only two groups of subjects (*i.e.*, healthy controls *versus* patients with one disorder), and low numbers of patients and lack of comparability severely limit the suitability of such data for a meta-analysis to define biomarkers. Consequently, the findings from several studies cannot be combined to construct a broad database. Ideally, the analysis must be accomplished within a reasonably short time with high resolution, enabling profiling of an adequate number of features from sufficient samples to yield robust diagnostic panels.

Capillary electrophoresis coupled to mass spectrometry (CE-MS) enables reproducible and robust high-resolution analysis of several thousand low-molecular-weight urinary proteins/peptides in less than an hour[1]. This approach has been used to analyze

urine samples from healthy volunteers and patients with various diseases[17-24]. The high number of datasets analyzed using identical conditions on the same instruments allows the establishment of a low-molecular-weight proteome database that can serve as a basis for the diagnosis, classification, and monitoring of a wide range of diseases using proteome analysis. Here, we report on the establishment of a database of urinary peptides analyzed by CE-MS and illustrate the application of this database to chronic renal diseases (CRD) and to non-renal diseases.

RESULTS

All urine samples were prepared identically and analyzed by CE-MS analysis, using identical instruments. Analysis resulted in individual data sets containing information on generally 1,200-2,000 peptides and proteins per sample. All information recommended by the "minimum information about proteomics experiments" (MIAPE) guidelines[25] about proteome analysis using CE and MS is recorded, and available upon request. The data were evaluated using MosaiquesVisu[26] (see **methods**), resulting in a list of peptides/proteins defined by mass, migration time, and ion-counts, serving as a measure of relative abundance. Different charge states of identical peptides/proteins were combined and included as a single identification in the database. A list of tentatively identified peptides of any sample is obtained and subsequently calibrated using "internal standards", peptides generally present in urine[21] (see **methods**). This allows the digital compilation of individual datasets into a specific "disease group" that can be compared to any desired "control group", enabling the identification of statistically significant changes that result in the definition of potential biomarkers. Only datasets that fulfilled all quality control criteria (see **methods**) were subsequently utilized and included in the database,

which currently includes CE-MS data from 28 different pre-selected pathophysiological conditions (see **figure 1**).

To improve the mass accuracy, several samples were analyzed using CE-online coupled to Fourier transform-ion cyclotron resonance mass spectrometry (FT-ICR-MS). Due to the high cost and the lower sensitivity of the FT-ICR MS instrument used in comparison to TOF-instruments, it was not practical to analyze all samples using CE-FT-ICR MS, and only a set of 20 samples was re-analyzed using this technique. Owing to the higher detection limit of the FT-ICR instrument (in comparison to the ESI-TOF used), the number of FT-ICR-traceable peptides was significantly lower (a factor of ~10). However, the analysis using CE-FT-ICR resulted in the definition of over 300 urinary peptides, and 80 of these precise masses (mass deviations <0.5 ppm) were utilized to calibrate the TOF-MS derived masses. The high FT-ICR MS resolution also enabled an accurate analysis of the first isotope signal of ions with z>6, which is crucial for the exact mass determination of proteins and high molecular weight peptides. These data were used to refine the TOF-MS masses in the human urinary proteome database. Consequently, 'FT-ICR-calibrated' TOF masses of most peptides revealed a deviation from the theoretical mass of 3±9 ppm.

All detected peptides and proteins in the 3,687 human urine samples that passed all quality control criteria (on average, 1,724 peptides/proteins were detected in each individual urine sample, ranging from 983 to 4,094) were deposited in a Microsoft SQL database and subsequently matched for further analysis and comparison of individual samples. This process resulted in the definition of 116,869 different peptides and proteins. Each peptide was assigned a unique identification number (Protein ID). As described previously[22, 27], several of these peptides appeared sporadically, being observed in only one or a few samples. To eliminate such peptides of apparently low significance, only those present in more than 20% of the urine samples in at least one group (samples from

patients with same disease) were further investigated. This noise-filtering process reduced the number of peptides available for analysis significantly: applying these limits, 5,010 “relevant” different peptides characterized by molecular mass [Da] and normalized CE-migration time [min] could be detected. The filtered data of all individual samples can be accessed in the **supplementary table 1**.

We sequenced an array of the peptides with a variety of tandem MS technologies (MS/MS). As described previously[28, 29], the migration time in CE depends on the size and the number of charges at pH 2.2, equaling the number of free amino groups (N-terminus and basic amino acids). Therefore, it is not a prerequisite to use CE-separation for MS/MS sequencing, as the number of basic amino acids, as well as the exact mass, serves to correlate sequences with a signal in the CE-MS run with high confidence. To date, we have identified 352 peptides by sequence analysis with different MS/MS platforms. The most commonly sequenced peptides were fragments of different collagen types and of uromodulin (**table 1** and **supplementary table 2**).

Application of the human urinary proteome database

The main purpose of this database is to serve as a universal platform for identification and validation of biomarkers for a variety of diseases/pathophysiological changes. This process is demonstrated in the following two examples:

Renal Disease

For the selection of disease-specific biomarkers, the data from individual samples were compiled as described previously[17, 21] and grouped according to the patients’ profiles (diagnostic group). These were healthy subjects (N=386) and patients with various biopsy-proven renal diseases (N=226) (for details, see **table 2**). For biomarker definition, all peptides and proteins were statistically analyzed and corrected for multiple testing[30]. This

resulted in the tentative identification of 35 peptides, listed in **supplementary table 3**. In an additional step, an SVM-based classification model was built upon these peptides (**figure 2** and **supplementary table 3**), enabling distinguishment of both groups in the training set with 92% sensitivity and 99.5% specificity [for Receiver Operating Characteristic (ROC) analysis, see also **figure 3A**].

To examine the value of these biomarkers[31], we validated the biomarker pattern from the reference cohort in a second blinded cohort. The 'chronic renal disease model' was assessed in a blinded cohort of 131 urine samples including patients with different renal diseases and healthy controls. Upon unblinding, 33/35 healthy controls and 86/96 patients with chronic renal disease were correctly classified, resulting in a sensitivity of 89.6% [95% CI: 81.7-94.9] and specificity of 94.3% [95% CI: 80.8-99.1]) (see **figure 3B**).

Next, two sets of biomarkers were established that discriminated patients with focal segmental glomerulosclerosis (FSGS) or with membranous glomerulonephritis (MNGN) from patients with other renal diseases (see **table 2**) ('FSGS' biomarker set; 24 biomarkers and 'MNGN' biomarker set; 19 biomarkers, respectively, see **supplementary table 3**). These biomarkers were again combined in an SVM-based model. Upon complete crossvalidation, the models enabled correct classification of the FSGS patients with 100% sensitivity and 95.5% specificity, and the MNGN patients with 100% sensitivity and 90.3% specificity. Application of the model onto the blinded test set resulted in the separation of all 3 patients with FSGS from patients with other renal diseases in the blinded set (N=88) and the 'MNGN' biomarker set detected all 4 MNGN patients (for the differential analysis results see **figure 4**).

Non-renal Disease

An example for the application of the urinary database to non-genitourinary diseases is the identification and validation of biomarkers indicating immunosuppression in organ or stem

cell transplantation. In a reference cohort of 395 urine samples (135 patients on immunosupression and 260 controls, see **table 2**), 10 indicative and statistically significant biomarkers were selected (**supplementary table 3**) that enabled classification of the training set with 90.4% sensitivity and 87.7% specificity upon complete crossvalidation. Application of this panel to a test cohort of 1,304 samples resulted in correct identification of 322/400 transplant patients (sensitivity 80.5%) and 746/904 controls (specificity 82.5%) (**table 2**). The control samples consisted of 298 healthy controls and 368 chronic renal diseases and covered a broad range of non-transplanted controls. In a further step, we used CE-MS data from patients after liver (N=5) or renal transplantation (N=18) without Kaposi's sarcoma (KS) and transplant patients developing KS (N=20). After rigorous multiple testing using Tmax statistics, we tentatively identified 8 significant biomarker candidates (**supplementary table 3**). These 8 peptides were utilized in an SVM-based model using the samples as the reference set. Upon total cross validation, 22/23 patients without KS and 18/20 patients with KS of the reference set were classified correctly (96.7% sensitivity and 91.3% specificity).

DISCUSSION

Here we report on the establishment of a database of naturally occurring urinary peptides and proteins and demonstrate its application to the definition of biomarkers of human diseases. These biomarkers apparently reflect primary pathogenetic changes as well as the reaction to the disease. Hence, their usefulness extends beyond the applicability to diseases of the urogenital tract, and the approach may be universally applicable to diseases that result in systemic changes. While genetic analysis can predict the risk of a disease, proteomics, with its potential to monitor dynamic processes, may more clearly

show at which point the risk manifests itself as disease and also facilitate monitoring of the response to therapy. Thus, these methods are complementary in personalized medicine[32].

Recently, several groups reported on the sequencing of an array of urinary proteins[12, 33]. While these data impressively demonstrate the large amount of proteins – potential information – contained in urine, critical or even mandatory information for their application in the definition of biomarkers is missing:

All reports are on tryptic digests of urinary proteins, and the sequences obtained from the tryptic peptides allow the authors to tentatively assign a protein to this sequence with variable confidence. Unfortunately, due to the *in vitro* manipulation of samples *via* digestion, it is not possible to define which species are actually present in the urine at the time of sampling. The actually occurring protein(s) will generally not be the protein in the database (*e.g.* Albumin precursor), but one or several variably post-translationally modified (PTM) proteins. In fact, PTMs are often the hallmark of the potential biomarker, *e.g.* advanced glycation endproducts as markers for uremia[34]. The information that is required is the definition of the peptide/protein present in urine. Further, if biomarkers are to be defined, information on the relative abundance of proteins/peptides is mandatory.

We have therefore attempted to obtain such critical information: the naturally occurring protein/peptide is defined by mass and migration time, and the relative abundance is defined on ion-counting, set in relation to “internal standards”, mostly specific collagen fragments that are present in almost every sample and appear not to change significantly in all the samples and disease-groups investigated to date. While this approach does not initially allow identification via sequence, it does allow for tentative identification based on mass and migration time. Sequencing was performed in a second step, but does meet several obstacles associated with sequencing of naturally occurring peptides [tryptic digests cannot be utilized, as these would result in a loss of connectivity to the original

identification parameters[35]]. Major obstacles are the frequently occurring PTMs that change the mass, which then is different from the theoretical mass present in the database, and the higher degree of freedom, as not pre-set terminal Arginine or Lysine can be employed in the database search. Further, search algorithms are generally adapted to the needs of tryptic digests, which differ greatly from the requirements for *de novo* sequencing of naturally occurring peptides or proteins (see also *e.g.*[1, 27]).

CE-MS analysis of urine enables tentative identification of biomarkers for a variety of diseases of the kidney and the urogenital tract[17-19, 21, 22, 36], although the high biological variability of peptides/proteins presents a serious methodological impediment. Therefore, it appears imperative to evaluate clinical conditions not on the basis of single peptide/protein markers, but rather on the basis of a panel of biomarkers that must be derived from distinct and clearly defined molecules. A panel of biomarkers will tolerate changes in individual analytes without jeopardizing the diagnostic precision, *i.e.* such variability will not result in gross changes of the diagnostic result.

While the exact sequences of biomarkers are not required to exploit their diagnostic potential, the sequences may offer further insight into the pathogenesis of a disease, (patho-) physiological mechanisms, and design of relevant therapeutics. Hence, sequence analysis of naturally occurring peptides in urine completes the content of this database. Most of the original proteins have also been identified by other research groups[12, 33], but the majority of naturally occurring peptides have not been defined. This is also reflected by the finding that collagen fragments represented the most abundant peptides in urine (see **table 1**).

Most of these naturally occurring urinary peptides are the result of proteolytic activity. Extracellular proteases may reflect the presence of the disease and its progression[37]. Complex changes in protease activities may be more readily recognized by

the pattern of proteolytic fragments generated than by direct assessment of the specific protease activity[38]. CE-MS analysis may be suitable to display the regulated activity of proteases and protease inhibitors by displaying potential products and enabling monitoring of their concentrations.

To assess the value of the urinary proteome database, we utilized the data for the diagnosis of CRD as a representative example for diseases that are related to direct involvement of the urogenital tract. Using this database, biomarkers could be defined that allowed classification of a blinded cohort of urine samples, which distinguished 35 healthy controls from 96 patients with CRD (sensitivity 89.6%, specificity 94.3%) (ROC analysis is shown in **figure 3B**). The potential of the approach was further underlined by the assessment of the same dataset using multiple panels. Without any additional measurements, the 131 subjects were reassessed using two additional disease-specific peptide panels, for 'FSGS' and 'MNGN' (see **figure 4**). With this approach, the FSGS and MNGN patients were identified with high sensitivity and specificity. The clear identification of these FSGS and MNGN patients, representing only a few individuals of this heterogeneous population (see **table 2**), demonstrates the feasibility of the use of the urinary proteome database for clinical purposes. The approach can be extended systematically to a variety of other CRD to generate a differential diagnosis in a noninvasive manner without the inherent risks of renal biopsy[18]. Interestingly, many of the sequenced biomarkers for CRD are fragments of collagens that are down-regulated. This down-regulation may be due to the fact that the activities of collagenases, such as matrix metalloproteases (MMP) are decreased. Regardless of the primary etiology, CRD is characterized by tubular atrophy, interstitial fibrosis, and glomerulosclerosis. Hence, it has been assumed that diminished MMP activity is responsible for the accumulation of the extracellular matrix (ECM) proteins and collagens that typify the fibrotic kidney[39]. This

effect may be the response of an increased concentration of inhibitors of MMP, such as tissue inhibitors of matrix metalloproteinase (TIMPs).

The application of the urinary proteome analysis is not restricted to genitourinary diseases. For example, we validated a peptide pattern indicative of immunosuppression in organ or cell transplantation in a test set of 1,304 samples. The identified biomarkers are most likely a result of immunosuppressive drugs, such as calcineurin-inhibitors. A detailed reevaluation of the immunosuppressive therapy of the transplanted patients focusing on the proteomic results may provide novel insights into the molecular mechanisms of the side effects of treatment. The pathogenic mechanisms of such undesired effects are rapidly gaining attention, because the focus in transplantation medicine has shifted from maintaining graft function over a short post-transplantation interval to optimizing quality of life for the long term[23].

As we learn to better appreciate the huge individual differences in the responses of patients to therapy, objective methods to measure treatment responses will become of prime importance so as to tailor the therapy to the individual. Noninvasive urinary proteomics has an advantage in that such monitoring is possible in real time, and adjustments can be made accordingly. This vision is within reach, but its realization entirely depends on establishment of databases that allow a quick and robust comparison of the patient's profiles against that of healthy controls and other patients. Thus, we contend that the urinary proteome database presented here is a major step forward in this direction. We anticipate that the availability of such databases in the future will significantly improve the options for patients with respect to diagnosis and therapy.

METHODS

Samples were collected at >20 clinical centers according to established protocols[20, 21]. Informed consent was obtained from all patients and ethical approval was obtained from the appropriate boards for all samples included.

Sample preparation

All samples for CE-MS analysis were from spontaneously voided urine and were stored at -20°C until analysis. For proteomic analysis, a 0.7 mL aliquot was thawed immediately before use and diluted with 0.7 mL 2 M urea, 10 mM NH_4OH containing 0.02 % SDS. To remove proteins of higher molecular mass, the sample was filtered with Centricon ultracentrifugation filter devices (30 kDa molecular weight cut-off; Millipore, Billerica, MA, USA) at 3,000 g until 1.1 ml of filtrate was obtained. The filtrate was then applied onto a PD-10 desalting column (Amersham Bioscience, Sweden) equilibrated in 0.01% NH_4OH in HPLC-grade water to remove urea, electrolytes and salts. Finally, all samples were lyophilized, stored at 4°C, and resuspended in HPLC-grade water shortly before CE-MS analysis, as described[21]. The re-suspension volume was adjusted to 0.8 µg/µL, according to the peptide content of the sample as measured by BCA assay (Interchim, Montlucon, France).

CE-MS analysis

CE-MS analysis was performed with a P/ACE MDQ capillary electrophoresis system (Beckman Coulter, USA) coupled online to a Micro-TOF MS (Bruker Daltonic, Germany)[21] The ESI sprayer (Agilent Technologies, USA) was grounded, and the ion spray interface potential was set between -4.0 and -4.5 kV. Data acquisition and MS acquisition methods were automatically controlled by the CE via contact-close-relays. Spectra were

accumulated every 3 seconds, over a range of m/z 350 to 3,000. The average recovery of the sample preparation procedure was approximately 85% with a detection limit of ~1 fmol[21]. The monoisotopic mass signals could be resolved for z≤6. The mass accuracy of the CE-TOF-MS method was determined to be <25 ppm for monoisotopic resolution and <100 ppm for unresolved peaks (z>6). The precision of the analytical method was determined by assessing (a) the reproducibility achieved for repeated measurement of the same aliquot and (b) by the reproducibility achieved for repeated preparation and measurement of the same urine sample. The 200 most abundant peptides ("internal standard" peptides) were detected with a rate of 98%. The performance of the analytical system over time was assessed with consecutive measurements of the same aliquot over a period of 24 h. No significant loss of peptides and proteins was observed implying the stability of the CE-MS set up, the post-preparative stability of the urine samples at 4°C and their resistance to oxidizing processes or precipitation[21, 26].

Data were accepted only if the following quality control criteria were met: A minimum of 950 peptides/proteins (mean number of peptides/proteins minus one standard deviation) must be detected with a minimal MS resolution of 8,000 (required resolution of peaks with z=6) in a minimal migration time interval (the time window, in which separated peptides can be detected) of 10 minutes. After calibration, the deviation of migration time must be below 0.35 minutes.

CE-FT-ICR-MS analysis

For CE-FT-ICR-MS, a Bruker Daltonic Apex Qe instrument equipped with a 12-T magnet and an Apollo II ion source was used. Coupling of the P/ACE 5510 capillary electrophoresis system (Beckman Coulter, USA) via the Agilent ESI sprayer was performed as above. The instrument was tuned with a peptide standard mix[21] and

externally mass calibrated on arginine clusters (< 0.1 ppm calibration errors). Mass spectra were acquired over an m/z range of 300-2,000. Ions were stored in the collision cell for 500 ms and 5 spectra were accumulated for each scan, resulting in a scan rate of 5 s.

Data processing

Mass spectral ion peaks representing identical molecules at different charge states were deconvoluted into single masses using MosaiquesVisu software[40] (www.proteomiques.com). Only signals observed in a minimum of 3 consecutive spectra with a signal-to-noise ratio of at least 4 were considered. MosaiquesVisu employs a probabilistic clustering algorithm and uses both isotopic distribution as well as conjugated masses for charge-state determination of peptides/proteins. The resulting peak list characterizes each protein/peptide by its molecular mass and its normalized migration time. TOF-MS data were calibrated utilizing FT-ICR-MS data as reference masses applying linear regression. Both CE-migration time and ion signal intensity (amplitude) showed high variability, mostly due to different amounts of salt and peptides in the sample. Consequently, CE-migration time and ion signal intensity were normalized based on reference signals by 200 abundant "housekeeping" peptides generally present in urine, which serve as internal standards[20, 21]. These "internal standards" were present in at least 90% of all urine samples with a relative standard deviation less than 100%. For calibration, a weighted regression was performed. The resulting peak list characterizes each protein and peptide by its molecular mass [Da], normalized CE migration time [min] and normalized signal intensity. All detected peptides were deposited, matched, and annotated in a Microsoft SQL database, allowing further analysis and comparison of multiple samples (patient groups). Proteins and peptides within different samples were considered identical, if the mass deviation was lower than 50 ppm for small peptides or 75 ppm for larger

peptides and proteins. The CE migration time deviation was linearly increased over the entire electropherogram from 2-5%. These clustering parameters showed minimal error rates and considered increased peak widths at higher migration times. Disease-specific protein/peptide patterns were generated using support-vector-machine (SVM) based MosaCluster software[17].

Statistical analysis

Estimates of sensitivity and specificity were calculated based on tabulating the number of correctly classified samples. Confidence intervals (95% CI) were based on exact binomial calculations performed with MedCalc version 8.1.1.0 (MedCalc Software, Belgium, www.medcalc.be). The ROC plot was evaluated, as it provides a single measure of overall accuracy that is not dependent upon a particular threshold[41].

The reported p-values were calculated using the natural logarithm transformed intensities and the Gaussian approximation to the t-distribution. Bonferroni adjustments were obtained by applying the standard Bonferroni criterion to the subset of markers that passed the frequency threshold of 70%. The maxT p-values were calculated using the Westfall and Young maxT-procedure[41]. This function computes permutation-based step-down adjusted p-values. A total of 100,000 permutations were performed. To ensure stability of the results, we verified that the p-values by the minP procedure of Westfall and Young were of similar magnitude[42]. Both procedures were implemented as macros in the commercial statistical package SAS (www.sas.com) and are also part of the multitest R-package of Dudoit et al. (see *e.g.*[30] and references therein) available at www.bioconductor.org.

Sequencing of peptides

Candidate biomarkers and other native peptides from urine were sequenced using CE- or LC-MS/MS analysis as recently described in detail[28].

In addition, MS/MS experiments were performed on an Ultimate 3000 nanoflow system (Dionex/LC Packings, USA) connected to an LTQ Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific, Germany) equipped with a nanoelectrospray ion source. The mass spectrometer was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition. Survey full-scan MS spectra (from *m*/*z* 300–2,000) were acquired in the Orbitrap. Ions were sequentially isolated for fragmentation in the linear ion trap using collisionally induced dissociation. General mass spectrometric conditions were: electrospray voltage, 1.6 kV; no sheath and auxiliary gas flow; ion transfer tube temperature, 225 °C; collision gas pressure, 1.3 mT; normalized collision energy, 32% for MS^2. Ion selection threshold was 500 counts for MS/MS.

Samples were also analyzed using Electron Transfer Dissociation (ETD)[43-45]. Peptides were separated by nRP-HPLC (Agilent 1100; flow split by tee to ~60 nL/min) and introduced into an ETD-capable Finnigan LTQ quadrupole linear ion trap via nESI, using previously described instrumental parameters[46].

All resulting MS/MS data were submitted to MASCOT (www.matrixscience.com) for a search against human entries in the MDSB Protein Database. Accepted parent ion mass deviation was 50 ppm; accepted fragment ion mass deviation was 500 ppm. Only search results with a MASCOT peptide score of 20 or better, which also met ion coverage stipulations as related to the main spectral features were included. Data files from experiments performed on the ETD-enabled LTQ were searched against the NCBI human non-redundant database using the Open Mass Spectrometry Search Algorithm (OMSSA), with an e-value cut-off of 0.01. The number of basic and neutral polar amino acids of the

peptide sequences was utilized to correlate peptide sequencing data to CE-MS data, as described earlier[28].

Table legends:

Table 1: Distribution of native peptides identified with respect to their protein precursor (described by SwissProt protein name and gene symbol). Comparison of the located peptides to other references[12, 33].

Number	Protein name	Gene symbol	Proteins detected by:	
			Adachi et al.	Castagna et al.
142	Collagen alpha-1 (I) chain [H. sapiens]	*COL1A1*	yes	no
57	Collagen alpha-1 (III) chain [H. sapiens]	*COL3A1*	yes	no
21	Collagen alpha-2 (I) chain [H. sapiens]	*COL1A2*	yes	no
18	Uromodulin [H. sapiens]	*UMOD*	yes	yes
12	Alpha-1-antitrypsin [H. sapiens]	*SERPINA1*	yes	yes
12	Fibrinogen alpha chain [H. sapiens]	*FGA*	yes	no
11	Serum albumin [H. sapiens]	*ALB*	yes	yes
7	Hemoglobin beta subunit [H. sapiens]	*HBB*	yes	yes
6	Polymeric-immunoglobulin receptor [H. sapiens]	*PIGR*	yes	yes
5	Hemoglobin alpha subunit [H. sapiens]	*HBA1, HBA2*	yes	no
3	Beta-2-microglobulin [H. sapiens]	*B2M*	yes	yes
3	Collagen alpha-1 (II) chain [H. sapiens]	*COL2A1*	no	no
3	Membrane associated progesterone receptor component 1 [H. sapiens]	*PGRMC1*	yes	no
3	Osteopontin [H. sapiens]	*SPP1*	yes	no
2	Alpha-1-microglobulin [H. sapiens]	*AMBP*	yes	yes
2	Alpha-2-HS-glycoprotein [H. sapiens]	*AHSG*	yes	yes
2	Apolipoprotein A-I [H. sapiens]	*APOA1*	no	yes
2	CD99 antigen [H. sapiens]	*CD99*	no	no
2	Clusterin [H. sapiens]	*CLU*	yes	yes
2	Collagen alpha-1 (XVIII) chain [H. sapiens]	*COL18A1*	yes	no
2	Epithelial-cadherin [H. sapiens]	*CDH1*	yes	yes
2	Insulin; includes C peptide [H. sapiens]	*INS*	no	no
2	ProSAAS [H. sapiens]	*PCSK1N*	yes	no
2	Prostaglandin-H2 D-isomerase [H. sapiens]	*PTGDS*	yes	yes
1	Alpha-1-acid glycoprotein 1 [H. sapiens]	*ORM1*	yes	yes
1	Antithrombin-III [H. sapiens]	*SERPINC1*	yes	no
1	Basement membrane-specific heparan sulfate proteoglycan core protein [H. sapiens]	*HSPG2*	yes	yes
1	Collagen alpha-1 (XIX) chain [H. sapiens]	*COL19A1*	no	no
1	Collagen alpha-1 (XV) chain [H. sapiens]	*COL15A1*	yes	no
1	Collagen alpha-1 (XVII) chain [H. sapiens]	*COL17A1*	no	no
1	Collagen alpha-1 (XXII) chain [H. sapiens]	*COL22A1*	no	no
1	Collagen alpha-2 (VIII) chain [H. sapiens]	*COL8A2*	no	no
1	Collagen alpha-3 (IX) chain [H. sapiens]	*COL9A3*	no	no
1	Cystatin-B [H. sapiens]	*CSTB*	yes	no
1	Fibrinogen beta chain [H. sapiens]	*FGB*	no	no
1	Fillagrin [H. sapiens]	*FLG*	yes	no
1	Gelsolin [H. sapiens]	*GSN*	yes	yes
1	Ig kappa chain C region [H. sapiens]	*IGKC*	yes	yes
1	Ig kappa chain V-III region [H. sapiens]	none	no	yes
1	Ig lambda chain C regions protein [H. sapiens]	*IGLC1*	yes	yes
1	Insulin-like growth factor II precursor [H. sapiens]	*IGF2*	yes	no
1	Josephin-1 [H. sapiens]	*JOSD1*	no	no
1	Liprin-beta-2 [H. sapiens]	*PPFIBP2*	no	no
1	Microfibrillar-associated protein 5 [H. sapiens]	*MFAP5*	no	no
1	Neurosecretory protein VGF [H. sapiens]	*VGF*	yes	no
1	Peptidoglycan recognition protein [H. sapiens]	*PGLYRP1*	yes	yes
1	Psoriasis susceptibility 1 candidate gene 2 protein [H. sapiens]	*PSORS1C2*	yes	no
1	PX domain-containing protein kinase-like protein [H. sapiens]	*PXK*	no	no
1	Secreted and transmembrane protein 1 [H. sapiens]	*SECTM1*	yes	yes
1	Sodium/potassium-transporting ATPase gamma chain [H. sapiens]	*FXYD2*	yes	no
1	Transthyretin precursor (Prealbumin) [H. sapiens]	*TTR*	yes	yes
1	Zinc finger CCHC domain-containing protein 3 [H. sapiens]	*ZCCHC3*	no	no
1	Zinc finger protein 653 [H. sapiens]	*ZNF653*	no	no

Table 2: Compendium of all patients and healthy controls of the human urinary proteome database, which were used for the establishment of training- and test-sets. Additionally, all sensitivities and specificities including their confidence interval are shown. (Abbreviations: HC=healthy control; CRD= chronic renal disease; MNGN=membranous glomerulonephritis; FSGS=focal segmental glomerulosclerosis; MCD=minimal change disease; SLE=systemic lupus erythematosus; IgAN=IgA nephropathy, DN=diabetic nephropathy; CAD=coronary artery disease; NTx=renal transplantation; LTx=liver transplantation; HSCT=hematopoietic stem cell transplantation; PTLD=post-transplant lymphoproliferative disorders; Fanconi=Fanconi's syndrome; AD=Alzheimer's disease; DM=diabetes mellitus.

Disease condition	Training set				Test set			
	CASE	CONTROL	Sensitivity [%]	Specificity [%]	CASE	CONTROL	Sensitivity [%]	Specificity [%]
CRD	**Σ: 226** 18 vasculitis, 21 SLE, 31 MNGN, 24 MCD, 44 IgAN, 27 FSGS, 61 DN	**Σ: 386** 386 HC	**92.0** [95% CI: 87.7-95.2]	**99.5** [95% CI: 98.1-99.9]	**Σ: 96** 3 FSGS, 4 MNGN, 8 CAD, 5 SLE, 39 Vasculitis, 15 IgAN, 10 NTx, 1 DN, 11 other CRD	**Σ: 35** 35 HC	**89.6** [95% CI: 81.7-94.9]	**94.3** [95% CI: 80.8-99.1]
FSGS	**Σ: 27** 27 FSGS	**Σ: 199** 31 MNGN 21 SLE 24 MCD 44 IgAN 61 DN 18 Vasculitis	**100.0** [95% CI: 87.1-100.0]	**95.5** [95% CI: 91.6-97.9]	**Σ: 3** 3 FSGS	**Σ: 128** 35 HC, 4 MNGN, 8 CAD, 5 SLE, 39 Vasculitis, 15 IgAN, 10 NTx, 1 DN, 11 other CRD	**100.0** [95% CI: 30.5-100.0]	**91.4** [95% CI: 85.1-95.6]
MNGN	**Σ: 31** 31 MNGN	**Σ: 195** 27 FSGS 21 SLE 24 MCD 44 IgAN 61 DN 18 Vasculitis	**100.0** [95% CI: 88.7-100.0]	**90.3** [95% CI: 89.2-94.0]	**Σ: 4** 4 MNGN	**Σ: 127** 35 HC, 3 FSGS, 8 CAD, 5 SLE, 39 Vasculitis, 15 IgAN, 10 NTx, 1 DN, 11 other CRD	**100.0** [95% CI: 40.2-100.0]	**93.7** [95% CI: 88.0-97.2]
Transplantation	**Σ: 135** 77 NTx, 17 HSCT, 24 LTx, 17 PTLD	**Σ: 260** 66 HC, 7 Fanconi, 7 MCD, 10 FSGS, 8 MNGN, 2 CAD, 4 SLE, 17 Vasculitis, 16 IgAN, 43 DM, 17 DN, 31 AD, 28 blinded CRD 4 other CRD	**90.4** [95% CI: 84.1-94-8]	**87.7** [95% CI: 83.1-91.4]	**Σ: 400** 251 NTx, 49 HSCT, 64 LTx, 36 PTLD	**Σ: 904** 298 HC, 21 Fanconi, 20 MCD, 19 FSGS, 28 MNGN, 7 CAD, 21 SLE, 46 Vasculitis, 41 IgAN, 136 DM, 43 DN, 95 AD, 124 blinded CRD 5 other CRD	**80.5** [95% CI: 76.3-84.3]	**82.5** [95% CI: 79.9-84.9]

Figures legends:

Figure 1: Disease conditions which are to date represented in the human urinary proteome database.

Figure 2: Peptide patterns distinguishing patients with chronic renal disease (CRD) from healthy controls (HC). This figure shows the compiled data sets of 226 CRD samples (upper left panel) and 386 healthy control subjects (upper right panel) of the training set. Normalized molecular weight (y axis) is plotted against normalized CE-migration time (x-axis). The mean signal intensity is given in 3D-depiction. The lower panel depicts the 35 indicative peptides defining the specific pattern for CRD (lower left panel) and controls (lower right panel). Red arrows indicate decreased signal intensities and green arrows show increased signal intensities in urine of patients with CRD.

Figure 3: ROC curve of the 'Chronic Renal Disease' specific peptide panel.

ROC analysis of **A**: the training set (bold line; AUC=0.98; $p<0.0001$) and **B**: the test set after unblinding (bold line, AUC=0.92; $p<0.0001$). 95 % confidence intervals (95% CI) are indicated by thin lines.

Figure 4: Flow chart describing differential analysis of spot urine samples from 131 patients in the blinded test-set by CE-MS. In the first step, samples were assessed for renal injury using the 35 biomarkers comprising the 'Chronic Renal Disease' pattern. In the second step, samples positive for 'Chronic Renal Disease' pattern were analyzed for the FSGS pattern and for the MNGN pattern, respectively.

Supplementary material:

Supplementary table 1:

Table consists of 19 different spreadsheets called (**1**) polypeptides, (**2**) disease conditions, and (**3-19**) patients raw data part1 to 17.

(**1**): Polypeptides. Table listing all 5,010 different peptides/proteins (Protein ID) detected, their calibrated molecular mass [Da], and normalized CE migration time [min].

(**2**): Disease conditions. Table includes all 3,687 patients. Sample ID correlated to their specific indication of diseases.

(**3-16**): Patients raw data part 1 to 17. Tables in pivot format show the CE-MS data of the 3,687 samples in the database. The protein IDs of all peptides are given in the first column named “Protein ID”; the unique Sample IDs constitute the first row. The MS data from each sample are reflected in one column. The number in each cell represents the calibrated amplitude of the mass spectrometric signal of each peptide/protein detected in the sample. The table is divided into seventeen spreadsheets, since Microsoft Excel limits the maximal number of columns to 256.

Supplementary table 2:

352 peptide sequences obtained with MS/MS sequencing.

The table contains (from left to right) the number of the data base entry (Protein ID), the associated mass, the CE migration time, and the amplitude of the peptide signal. The additional information after sequence analysis is the peptide sequence, the name of the protein fragment, the SwissProt entry, the accession number, the calculated monoisotopic mass, and the deviation between observed and expected mass. In addition, FT-ICR masses are shown, including their deviation to the calculated mass.

Supplementary table 3:

Disease peptide panels.

Peptide markers derived from trainings sets of different disease conditions (see table 2) (CRD=chronic renal disease; MNGN=membranous glomerulonephritis; FSGS=focal segmental glomerulosclerosis; immunosuppression after transplantation; Kaposi's sarcoma). The peptide identification results from protein ID; molecular weight [Da] and CE-migration time [min] of the potential biomarkers.

Acknowledgement

This work was supported in part by Eurotransbio grant #ETB-2006-016 to the Urosysteomics consortium (www.urosysteomics.com) to ES and HM. and by grant #203.19-32329-5-461 from the Lower Saxony Ministry of Economy to AG, EMW and HM. HM and LT were supported by grant #LSHM-CT-2005-018733 to PREDICTIONS from the European Union, and HM and AFD were supported by grant #LSHM-CT-2006-037093 from InGenious HyperCare. Furthermore, AFD was supported by grant #RG/02/012 from the British Heart Foundation (BHF). WK was supported by grant #505520 from the FP6 EU project 'Interaction Proteome' and the Wellcome Trust JIF proteomics grant #29240. JPS was supported by Inserm, the "Direction Régional de la Recherche" (CHU de Toulouse, France) under the Interface program and by the Fondation pour la Recherche Médicale. DMG gratefully acknowledges support from an NIH pre-doctoral fellowship – the Biotechnology Training Program (NIH 5T32GM08349). BAJ and JN acknowledge the support from NIH (DK78244).

REFERENCES

1. Fliser, D. *et al.* Advances in urinary proteome analysis and biomarker discovery. *J. Am. Soc. Nephrol.* **18**, 1057-1071 (2007).

2. Gonzalez-Buitrago, J.M., Ferreira, L., & Lorenzo, I. Urinary proteomics. *Clin. Chim. Acta* **375**, 49-56 (2007).

3. Vidal, B.C., Bonventre, J.V., & Hong, H.S. Towards the application of proteomics in renal disease diagnosis. *Clin. Sci. (Lond)* **109**, 421-430 (2005).

4. Pieper, R. *et al.* Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics* **4**, 1159-1174 (2004).

5. Thongboonkerd, V., McLeish, K.R., Arthur, J.M., & Klein, J.B. Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. *Kidney Int.* **62**, 1461-1469 (2002).

6. Pisitkun, T., Shen, R.F., & Knepper, M.A. Identification and proteomic profiling of exosomes in human urine. *Proc. Natl. Acad. Sci. U. S. A* **101**, 13368-13373 (2004).

7. Marshall, T. & Williams, K. Two-dimensional electrophoresis of human urinary proteins following concentration by dye precipitation. *Electrophoresis* **17**, 1265-1272 (1996).

8. Shihabi, Z.K., Konen, J.C., & O'Connor, M.L. Albuminuria vs urinary total protein for detecting chronic renal disorders. *Clin. Sci.* **37**, 621-624 (1991).

9. Yudkin, J.S., Forrest, R.D., & Jackson, C.A. Microalbuminuria as predictor of vascular disease in non-diabetic subjects. Islington Diabetes Survey. *Lancet* **2**, 530-533 (1988).

10. Schaub, S. *et al.* Proteomic-based detection of urine proteins associated with acute renal allograft rejection. *J. Am. Soc. Nephrol.* **15**, 219-227 (2004).

11. Bueler, M.R., Wiederkehr, F., & Vonderschmitt, D.J. Electrophoretic, chromatographic and immunological studies of human urinary proteins. *Electrophoresis* **16**, 124-134 (1995).

12. Adachi, J. *et al.* The human urinary proteome contains more than 1500 proteins including a large proportion of membranes proteins. *Genome Biol.* **7**, R80.1-R80.16 (2006).

13. Ngai, H.H. *et al.* Serial changes in urinary proteome profile of membranous nephropathy: implications for pathophysiology and biomarker discovery. *J. Proteome Res.* **5**, 3038-3047 (2006).

14. Oh, J. *et al.* Establishment of a near-standard two-dimensional human urine proteomic map. *Proteomics* **4**, 3485-3497 (2004).

15. Spahr, C.S. *et al.* Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. *Proteomics* **1**, 93-107 (2001).

16. Cadieux, P.A. *et al.* Surface-enhanced laser desorption/ionization-time of flight-mass spectrometry (SELDI-TOF-MS): a new proteomic urinary test for patients with urolithiasis. *J. Clin. Lab Anal.* **18**, 170-175 (2004).

17. Decramer, S. *et al.* Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat. Med.* **12**, 398-400 (2006).

18. Haubitz, M. *et al.* Urine protein patterns can serve as diagnostic tools in patients with IgA nephropathy. *Kidney Int.* **67**, 2313-2320 (2005).

19. Rossing, K. *et al.* Impact of diabetic nephropathy and angiotensin II receptor blockade on urinary polypeptide patterns. *Kidney Int.* **68**, 193-205 (2005).

20. Theodorescu, D. *et al.* Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis* **26**, 2797-2808 (2005).

21. Theodorescu, D. *et al.* Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol.* **7**, 230-240 (2006).

22. Weissinger, E.M. *et al.* Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes. *Kidney Int.* **65**, 2426-2434 (2004).

23. Weissinger, E.M. *et al.* Proteomic patterns predict acute graft-versus-host disease after allogeneic hematopoietic stem cell transplantation. *Blood* **109**, 5511-5519 (2007).

24. Wittke, S. *et al.* Detection of acute tubulointerstitial rejection by proteomic analysis of urinary samples in renal transplant recipients. *Am. J. Transplant.* **5**, 2479-2488 (2005).

25. Taylor, C.F. *et al.* The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **25**, 887-893 (2007).

26. Kolch, W., Neususs, C., Pelzing, M., & Mischak, H. Capillary electrophoresis-mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. *Mass Spectrom. Rev.* **24**, 959-977 (2005).

27. Mischak,H., Julian,B.A., & Novak,J. High-resolution proteome/peptidome analysis of peptides and low-molecular-weight proteins in urine. Proteomics Clin. Appl. 1, 792-804. 2007.

28. Zurbig, P. *et al.* Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation. *Electrophoresis* **27**, 2111-2125 (2006).

29. Williams, B.J., Russell, W.K., & Russell, D.H. Utility of CE-MS data in protein identification. *Anal. Chem.* **79**, 3850-3855 (2007).

30. Dudoit, S. & van der Laan, M.J. *Multiple Testing Procedures and Applications to Genomics.* (Springer, Berlin, 2007).

31. Mischak, H. *et al.* Clinical Proteomics: a need to define the field and to begin to set adequate standards. *Proteomics Clin. Appl.* **1**, 148-156 (2007).

32. Woodcock, J. The prospects for "personalized medicine" in drug development and drug therapy. *Clin. Pharmacol. Ther.* **81**, 164-169 (2007).

33. Castagna, A. *et al.* Exploring the hidden human urinary proteome via ligand library beads. *J. Proteome Res.* **4**, 1917-1930 (2005).

34. Thornalley, P.J. *et al.* Mass spectrometric monitoring of albumin in uremia. *Kidney Int.* **58**, 2228-2234 (2000).

35. Chalmers, M.J. *et al.* Combined top-down and bottom-up mass spectrometric approach to characterization of biomarkers for renal disease. *Anal. Chem.* **77**, 7163-7171 (2005).

36. Mischak, H. *et al.* Proteomic analysis for the assessment of diabetic renal damage in humans. *Clin. Sci. (Lond)* **107**, 485-495 (2004).

37. Candiano, G. *et al.* Repetitive fragmentation products of albumin and alpha1-antitrypsin in glomerular diseases associated with nephrotic syndrome. *J. Am. Soc. Nephrol.* **17**, 3139-3148 (2006).

38. Villanueva, J. *et al.* Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J. Clin. Invest.* **116**, 271-284 (2006).

39. Cheng, S. *et al.* Matrix metalloproteinase 2 and basement membrane integrity: a unifying mechanism for progressive renal injury. *FASEB J.* **20**, 1898-1900 (2006).

40. Neuhoff, N. *et al.* Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry. *Rapid Commun. Mass Spectrom* **18**, 149-156 (2004).

41. DeLeo, J. M. *Receiver operating characteristic laboratory (ROCLAB): Software for developing decision strategies that account for uncertainty.* 318-325 (Second International Symposium on Uncertainty Modeling and Analysis, College Park, MD, USA, 1993).

42. Westfall, P.H. & Young, S.S. *Resampling-based Multiple Testing: Examples and Methods for P-Value Adjustment.* (John Wiley & Sons, New York, 1993).

43. Coon, J.J., Shabanowitz, J., Hunt, D.F., & Syka, J.E. Electron transfer dissociation of peptide anions. *J. Am. Soc. Mass Spectrom.* **16**, 880-882 (2005).

44. Syka, J.E. *et al.* Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A* **101**, 9528-9533 (2004).

45. Good, D.M. & Coon, J.J. Advancing proteomics with ion/ion chemistry. *Biotechniques* **40**, 783-789 (2006).

46. Good, D.M., Wirtala, M., McAlister, G.C., and Coon, J.J. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* in press, (2007).

Fig. 1

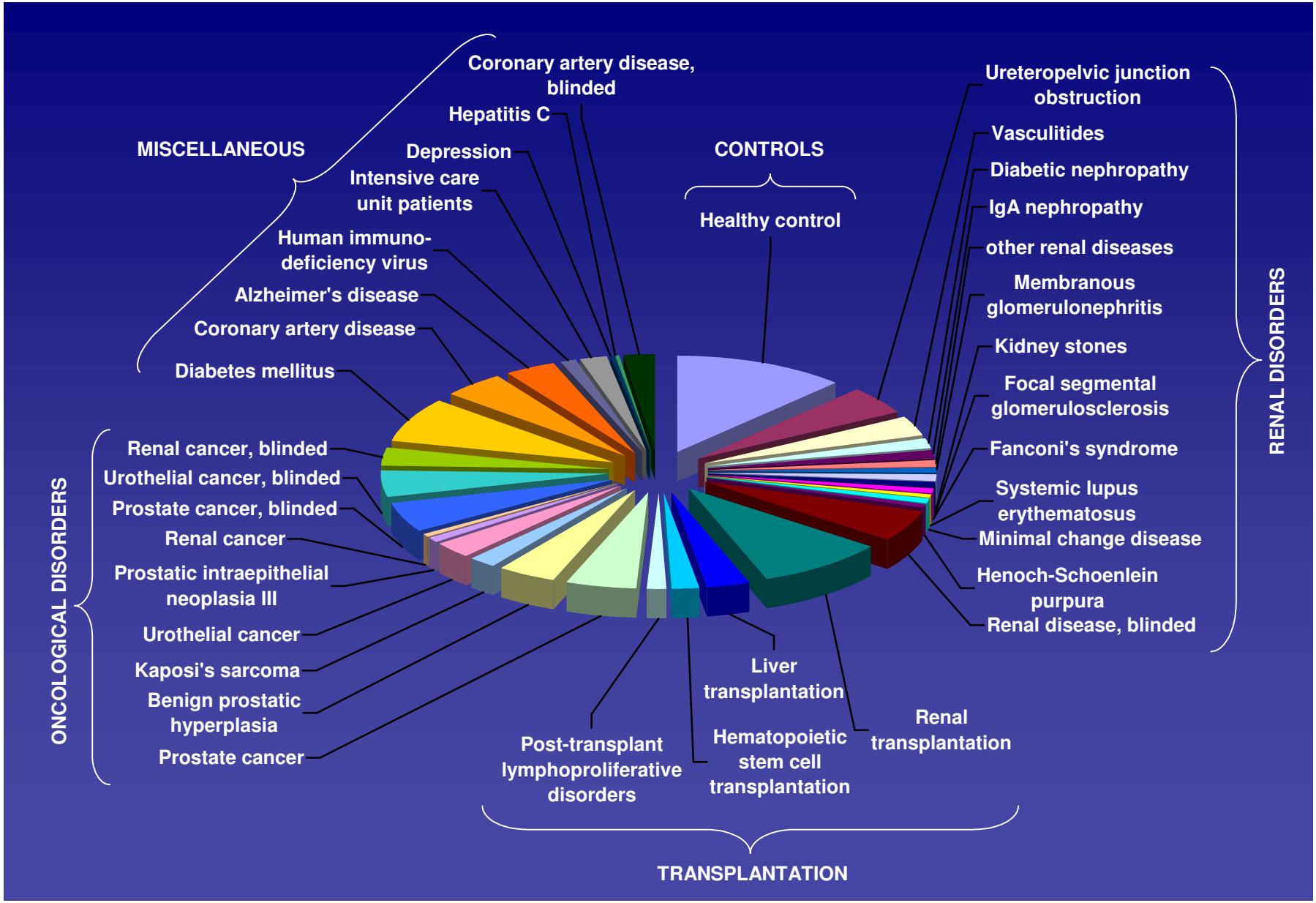

Fig. 2

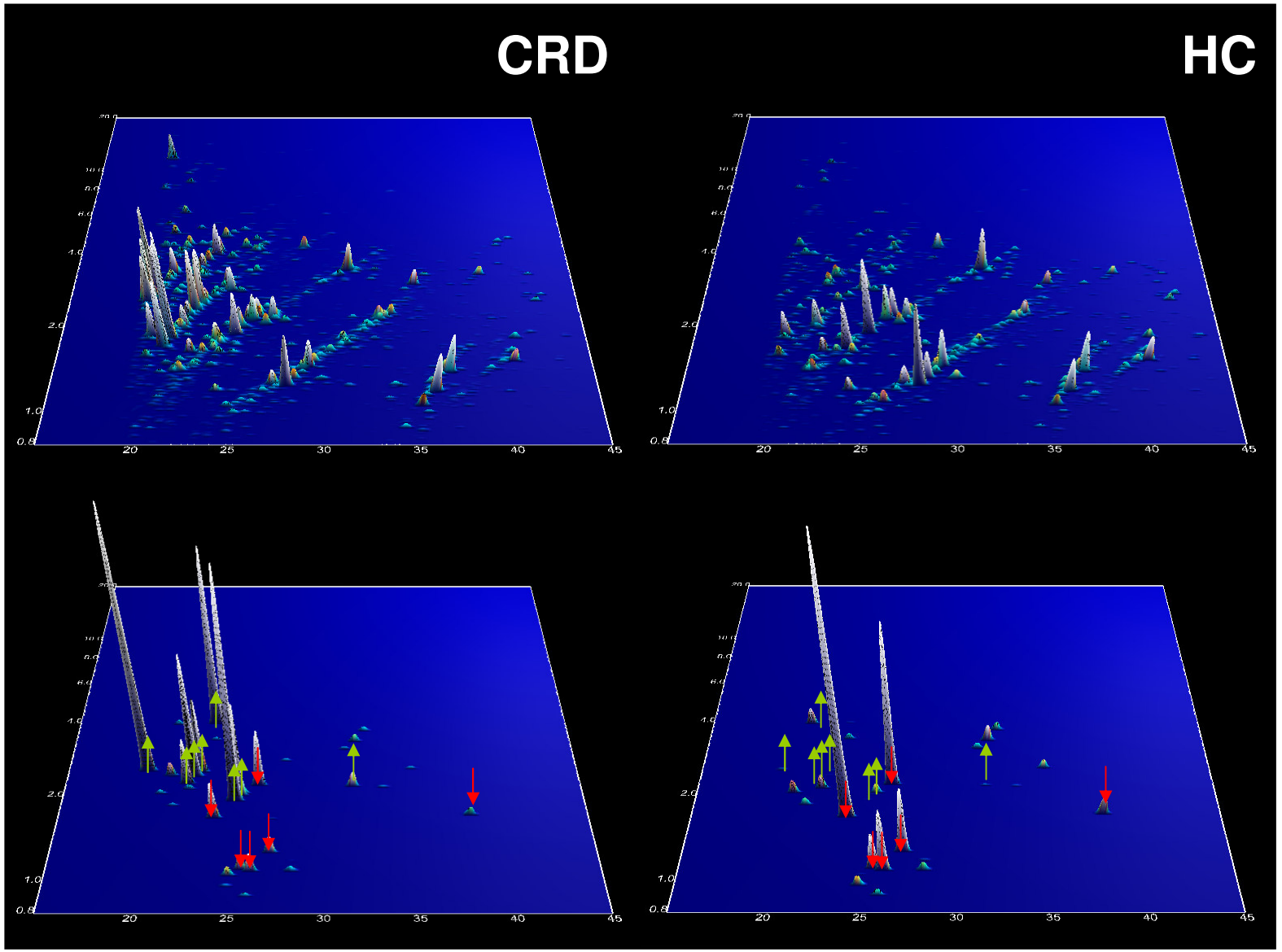

Fig. 3

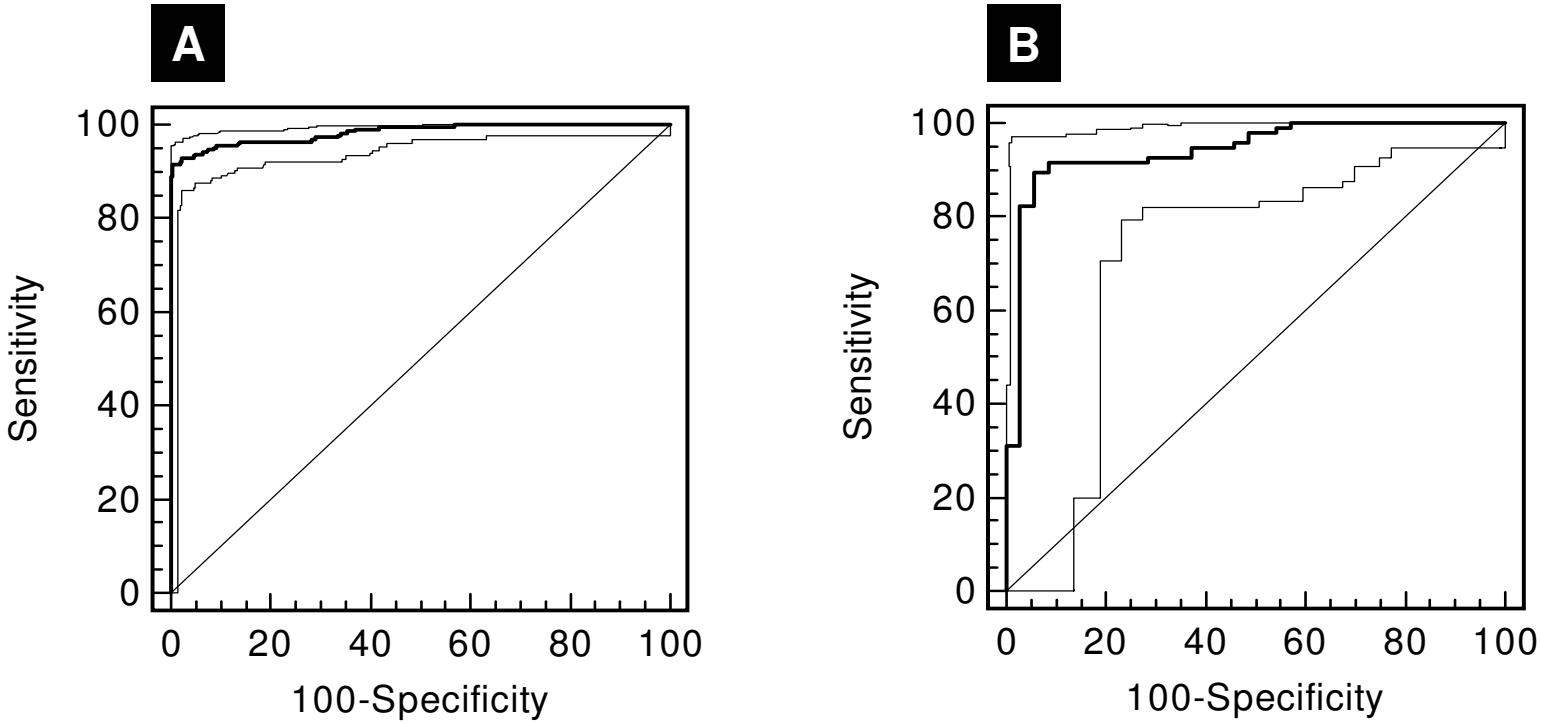

Fig. 4