Ashley Di Meo, Ihor Batruch, Arsani G. Yousef, Maria D. Pasic, Eleftherios P. Diamandis and George M. Yousef*

An integrated proteomic and peptidomic assessment of the normal human urinome

DOI 10.1515/cclm-2016-0390
Received May 3, 2016; accepted June 9, 2016; previously published online July 9, 2016

Abstract

Background: Urine represents an ideal source of clinically relevant biomarkers as it contains a large number of proteins and low molecular weight peptides. The comprehensive characterization of the normal urinary proteome and peptidome can serve as a reference for future biomarker discovery. Proteomic and peptidomic analysis of urine can also provide insight into normal physiology and disease pathology, especially for urogenital diseases.
Methods: We developed an integrated proteomic and peptidomic analytical protocol in normal urine. We employed ultrafiltration to separate protein and peptide fractions, which were analyzed separately using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) on the Q-Exactive mass spectrometer.
Results: By analyzing six urines from healthy individuals with advanced age, we identified 1754 proteins by proteomic analysis and 4543 endogenous peptides, arising from 566 proteins by peptidomic analysis. Overall, we identified 2091 non-redundant proteins by this integrated approach. In silico protease activity analysis indicated that metalloproteases are predominantly involved in the generation of the endogenous peptide signature. In addition, a number of proteins that were detected in normal urine have previously been implicated in various urological malignancies, including bladder cancer and renal cell carcinoma (RCC).
Conclusions: We utilized a highly sensitive proteomics approach that enabled us to identify one of the largest sets of protein identifications documented in normal human urine. The raw proteomics and peptidomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD003595.

Keywords: mass spectrometry; peptide sequence alignment; protease activity analysis; urinary peptidome; urinary proteome.

***Corresponding author: George M. Yousef,** MD, PhD, FRCPC (Path), Department of Laboratory Medicine, and the Keenan Research Centre for Biomedical Science at the Li Ka Shing Knowledge Institute, St. Michael's Hospital, 30 Bond Street, Toronto, ON, M5B 1W8, Canada, Phone: +1-416-864-6060, Ext: 77605, Fax: +416-864-5648, E-mail: yousefg@smh.ca; and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
Ashley Di Meo: Department of Laboratory Medicine, and the Keenan Research Centre for Biomedical Science at the Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada; Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada; and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
Ihor Batruch and Arsani G. Yousef: Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada
Maria D. Pasic: Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; and Department of Laboratory Medicine, St. Joseph's Health Centre, Toronto, ON, Canada
Eleftherios P. Diamandis: Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada; Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; and Department of Clinical Biochemistry, University Health Network, Toronto, ON, Canada

Introduction

The kidney plays an important role in the homeostatic regulation of electrolytes, bone/mineral metabolism, erythropoiesis, and blood pressure as well as the removal of metabolic by-products. Abundant serum proteins including albumin, immunoglobulin light chains, transferrin, vitamin D binding protein and myoglobin are filtered at the glomerulus and reabsorbed by the proximal tubules [1, 2].

Urine is a valuable biological fluid. Urine is composed of a diverse set of proteins and peptides that are primarily derived from the kidney, bladder and prostate as well as systemic circulation [3]. Urinary proteins and peptides are likely to reflect normal kidney tubular physiology as well as systemic physiology. Alterations in the urinary proteome and peptidome may be an indicator of disease. Proteomics is the large-scale study of proteins within a biological system. Peptidomics is an emerging field branching from proteomics that enables the comprehensive, qualitative and quantitative study of endogenous peptides. Unlike proteomics,

enzymatic digestion is not required for analysis [4–6]. Moreover, peptidomic profiling offers additional value to standard proteomic analysis through detection of bioactive peptides and insight into proteolytic events. Various groups have applied proteomic and peptidomic analyses for diagnostic, prognostic and predictive biomarker identification in several kidney-related disorders such as acute allograft rejection [7], RCC [8] and IgA nephropathy [9]. This has also been shown for systemic conditions including prostate cancer [10] and ovarian cancer [11]. Urine is an ideal fluid for such applications as it can be collected in large volumes non-invasively [12]. This is contrary to other biological fluids such as blood, where collection is more invasive and often leads to collection artifacts due to the activation of proteases [13, 14]. Moreover, the stability of proteins in urine is relatively high as proteolytic degradation is thought to be completed by the time of voiding [15].

A number of studies have attempted to characterize the healthy human urinary proteome. An early study by Thongboonkerd et al. reported 150 spots by two-dimensional polyacrylamide gel electrophoresis resulting in the identification of 47 unique proteins in normal human urine. The identified proteins included transporters, adhesion molecules, complement, chaperones, receptors and matrix proteins, in addition to others [16]. Another study observed 1400 spots by two-dimensional electrophoresis which lead to the identification of 150 unique urinary proteins [17]. In addition, Sun et al. identified 226 unique proteins in normal urine by liquid chromatography tandem mass spectrometry (LC-MS/MS) [18]. The introduction of high resolution MS technologies had a major impact, as the catalog of characterized urine proteins was significantly increased. A study by Li et al. identified 1310 non-redundant proteins in urine by high resolution MS [19]. In addition, Adachi et al. identified 1543 proteins in urine that provided a useful reference for comparing datasets obtained using different methodologies [20]. Another study identified 1823 proteins, of which 671 had not been previously characterized in urine [21]. A recent study by Santucci et al. identified 3429 proteins in urine from 12 healthy volunteers. This group found that 1615 proteins were contained within urinary vesicles while the remaining 1794 proteins were identified using combinatorial solid-phase peptide ligand enrichment [22, 23].

Several studies have also attempted to characterize the normal urinary peptidome. One study identified 193 unique endogenous peptides in normal urine by μRPLC-MS/MS analysis [24]. A peptidomic study by Fiedler et al. identified 427 unique mass signals in normal urine by magnetic bead (MB) separation followed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry [25]. Moreover, Mischak et al. characterized the healthy urine proteome and peptidome using several MS platforms across multiple centers. The authors identified a total of 361 unique endogenous peptides that originated from 41 proteins by capillary electrophoresis mass spectrometry (CE-MS) and LC-MS. For proteomic analysis, the authors employed 1D gel analysis coupled to nano-LC-MS/MS and 2DE, which allowed for the identification of 230 unique proteins [26]. Another study assessed the proteome and peptidome of urinary extracellular vesicles (UEVs). Analysis by LC-ESI-MS identified a total of 942 proteins by proteomic analysis. In addition, the authors identified 3115 unique endogenous peptides that were derived from 973 different protein isoforms [27].

In this study, we characterized the normal urinary proteome and peptidome of six healthy donors with advanced age by LC-MS/MS on the high mass accuracy Q-Exactive mass spectrometer. We identified 1754 proteins by proteomic analysis in addition to 4543 endogenous peptides arising from 566 proteins by peptidomic analysis. To our knowledge this is the largest number of normal urine peptides published to date. Together our integrated proteomic and peptidomic analytical approach identified a total of 2091 non-redundant proteins. The dataset presented here is one of the largest sets of protein identifications document in normal urine, successfully widening the existing human urinome. Furthermore, our collection of proteins and endogenous peptides may prove useful for future biomarker discovery.

Materials and methods

Sample collection

The study was approved by the Research Ethics Board of St. Michael's Hospital. Approximately 20–50 mL of second morning, mid-stream urine was collected from three healthy female volunteers (65–79 years of age) and three healthy male volunteers (62–75 years of age) using a self-sampling method (Supplemental Table 1). All subjects included in the study had no history of kidney-related or systemic disease. All urine specimens were individually processed in parallel using the same reagents and columns to minimize bias in sample preparation. Following the addition of a protease inhibitor cocktail tablet (Complete™; Roche Diagnostics, Mannheim Germany) the samples were centrifuged at 2000 *g* for 10 min at 4 °C and immediately frozen at −80 °C until further processing.

Proteomic and peptidomic sample preparation

Urine samples were thawed on ice and vortexed to resuspend any precipitate. The samples were centrifuged at 2000× *g*. Urine volumes

were normalized with respect to 90 µmol of creatinine per sample (~3–10 mL per individual). The pH of the urine samples was adjusted to 8.0 by dropwise addition of 0.5 mol/L ammonium bicarbonate (ABC). The urine was ultrafiltered using Vivaspin 20 ultracentrifugation filter devices (10 kDa MWCO; Sartorius Stedim Biotech), according to the manufacturer's protocol. The retentate (proteomic sample) was recovered following ultracentrifugation and stored at –20 °C overnight for processing the next day. The filtrate (peptidomic sample) was recovered following ultracentrifugation and dithiothreitol (DTT) (Sigma-Aldrich) was added to a final concentration of 2 mM. The filtrate was then incubated at room temperature for 30 min, followed by alkylation with 4 mM iodoacetamide (Sigma-Aldrich) at room temperature in the dark for 45 min. Filtrate samples were then acidified by dropwise addition of formic acid to pH 4.0. Acidified samples were passed through a hydrophilic-lipophilic-balanced reversed-phase cartridge (Oasis HLB). The cartridge [1 mL (30 mg); Waters cat# WAT094225] was pre-equilibrated with 1 mL 90% acetonitrile (ACN), 0.1% formic acid, and 0.02% trifluoroacetic acid (TFA). The cartridge was washed with 15 mL buffer A (5% ACN, 0.1% formic acid and 0.02% TFA). The acidified sample was then loaded and washed with 15 mL of buffer A. Peptides were eluted by adding 700 µL of 60% ACN, 0.1% formic acid, 0.02% TFA. Eluted samples were then mixed with equal volumes of ethyl acetate and centrifuged at 17,000 *g* for 5 min. The upper layer was discarded and the sample was reduced to a volume of 200 µL using the Speedvac system. Peptide samples were then stored at –20 °C until further processing.

The next day retentate samples were thawed on ice and vortexed. Retentate samples were assayed using the bicinchoninic acid assay (BCA) to determine the protein concentration. An equal mass of protein (200 µg) was collected from each sample and brought to a total volume of 100 µL using 50 mM ammonium bicarbonate. Powdered urea was added to each tube to a final concentration of 8 M for protein denaturation. A final concentration of 20 mM DTT (Sigma-Aldrich) was added to the retentate samples and incubated at room temperature for 30 min, which was then followed by alkylation with 40 mM iodoacetemide (Sigma-Aldrich) at room temperature in the dark for 45 min. Samples were diluted 4-fold with 50 mM ABC to reduce the urea concentration to 2 M. The samples were then digested with trypsin (50:1 protein/trypsin) overnight at 37 °C. Trypsin digestion was stopped the next day with formic acid to a final concentration of 1%. The samples were subsequently reduced to approximately 200 µL using the Speedvac system. Trypsin-digested protein samples were then stored at –20 °C until further processing.

Strong cation exchange chromatography

Both proteomic and peptidomic samples were re-suspended in 300 µL mobile phase A (0.26 M formic acid in 5% ACN). The samples were injected into a PolySUFLOETHYL A column with a 200-Å pore size and diameter of 5 µm (The Nest Group) containing a hydrophilic, anionic polymer (poly-2-sulfoethyl aspartamide). A 60 min linear gradient separation was performed on an HPLC system (Agilent 1100) using mobile phase B (0.26 M formic acid in 5% ACN and 1 M ammonium formate). The eluate was monitored at a wavelength of 280 nm. In total, 26 fractions per sample were collected for proteomics. Fractions 11–21 were pooled (Frs. 11–13, 14–17, and 18–21). In addition, 12 fractions per sample were collected for peptidomics. Fractions 3–9 were pooled (Frs. 3–5, 6, and 7–9). Fractions for both proteomic and peptidomic analysis were collected at a flow rate of 0.2 mL/min.

LC-MS/MS analysis

The proteomic and peptidomic fractions were desalted using an Omix C18 pipette tip (Agilent) and eluted in 5 µL buffer B (65% ACN, 0.1% formic acid in H_2O). After elution, 60 µL of buffer A (0.1% formic acid in H_2O) were added to each sample and 18 µL were loaded onto a 2 cm C18 trap column, packed with Varian Pursuit (5 µm C18) with an 8 µm tip (New Objective). This LC setup was coupled online to a Q Exactive (Thermo Fisher Scientific) mass spectrometer with a nanoelectrospray ionization source (Proxeon Biosystems). Each fraction underwent a 60 min gradient. The eluted peptides were analyzed by tandem mass spectrometry in positive ion mode. Individual urine specimens were run in duplicate on the Q-Exactive mass spectrometer.

Data analysis

Both the proteomic and peptidomic raw data were analyzed in MaxQuant 1.5.2.8 with the integrated Andromeda search engine [28]. Proteins were identified by searching MS and MS/MS data against a reverted version of the SwissProt human protein database (version January 2015) for proteomic analysis and a randomized version for peptidomic analysis. Label-free quantification was carried out as described by Luber et al. [29]. Protein abundance was calculated on the basis of the normalized spectral protein intensity.

For proteomic analysis carbamidomethyl cysteine was set as a fixed modification and oxidized methionine and protein N-acetylation were set as variable modifications. Trypsin was selected as the digestion method. A maximum of two missed cleavages was allowed for proteomic analysis. For peptidomic analysis carbamidomethyl cysteine was set as a fixed modification and oxidized methionine and oxidized proline were set as variable modifications. An unspecific enzyme search was selected as the digestion method. A maximum of 50 missed cleavages were allowed for peptidomic analysis. For both proteomic and peptidomic analyses, first search peptide tolerance was set to 20 ppm against a small 'human-first-search' database (part of MaxQuant) for the purpose of mass recalibration and main search was performed at 4.5 ppm against the SwissProt human protein database (version January 2015) [30]. Both the peptide-spectrum match and protein false discovery rate was specified at 0.01.

Bioinformatic analysis

In silico protease prediction analysis was done using Proteasix (http://www.proteasix.org/), a publically available prediction tool that allows for the identification of proteases involved in generating the observed peptide signature [8, 31]. The observed naturally occurring urinary peptide sequences were aligned with the full-length SWISS-PROT sequence for cleavage site retrieval. Up to three amino acid substitutions in the eight amino acid-long cleavage site were allowed. The percentage frequency plot was generated by calculating the number of observed cleavage sites per protease related to the total number of input cleavage sites. In addition, peptide sequence

alignment was done using a previously described computational program, peptide extractor (PepEx) [32]. Proteolytic maps display the position of endogenous peptides in their respective precursor protein in addition to their abundance.

The Human Protein Atlas (http://www.proteinatlas.org/) was mined to obtain a list of genes that display elevated transcript abundance in the kidney compared to other tissues. Three hundred and ninety-seven genes were found to have elevated expression in the kidney.

Gene ontology (GO) data was retrieved by running gene names through The Database for Annotation, Visualization, and Integrated Discovery (DAVID 6.7) (https://david.ncifcrf.gov/) search engine for biological process, molecular function, and cellular component annotations. Terms with a p value <0.05 were considered statistically significant.

Pathway enrichment analysis was carried out using DAVID 6.7 search engine that matches gene names to significant pathways within the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Terms with a p value <0.05 were considered statistically significant.

Statistical analysis

Proteomic and peptidomic inter-individual variances were calculated from the median coefficient of variation (CV), which is the standard deviation divided by the mean measurement. Only proteins that had a label-free quantification intensity in at least 3 (50%) of the six urine samples for included for analysis.

Results

Identification of the normal urinary proteome and peptidome

We developed an integrated proteomic and peptidomic protocol which is a multi-level assessment that incorporates both proteomic and peptidomic profiling of urine samples, as illustrated in Figure 1. Using a label-free LC-MS/MS approach we identified 1754 proteins from six urine samples by proteomic analysis (Supplemental Table 2A). Supplemental Table 2B displays a complete list of tryptic peptides identified by proteomic analysis. We also identified 4543 endogenous peptides arising from 566 proteins by peptidomic analysis as shown in Supplemental Table 3A and B. In total, we identified 2091 non-redundant proteins by combining proteomic and peptidomic approaches (Figure 2). Inter-individual variation was calculated for our integrated proteomic and peptidomic analysis. The inter-individual CV value was 0.65 for our integrated approach. This is consistent with another study that assessed variability of the normal urinary proteome, reporting an inter-individual CV value of 0.66 [33].

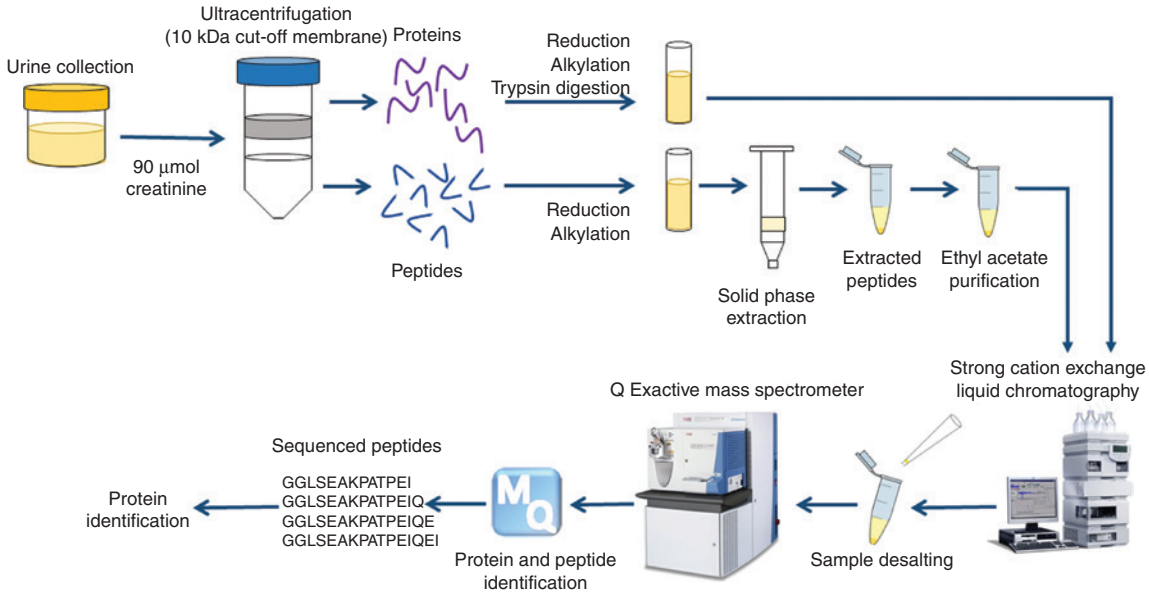

Figure 1: An overview of the experimental workflow for urinary proteomic and peptidomic analysis.
Urine is collected and a volume containing 90 µmol creatinine is processed further. Samples are then concentrated using a 10 kDa ultrafiltration unit. Retentate (protein) and filtrate (peptide) samples are reduced and alkylated. Retentate samples are further digested using trypsin. Peptides are extracted by solid phase extraction and purified using ethyl acetate. Strong cation exchange is then performed on both proteomic and peptidomic samples. Fractions are desalted and loaded onto the Q Exactive mass spectrometer. Proteins and peptides are identified using MaxQuant 1.5.2.8.

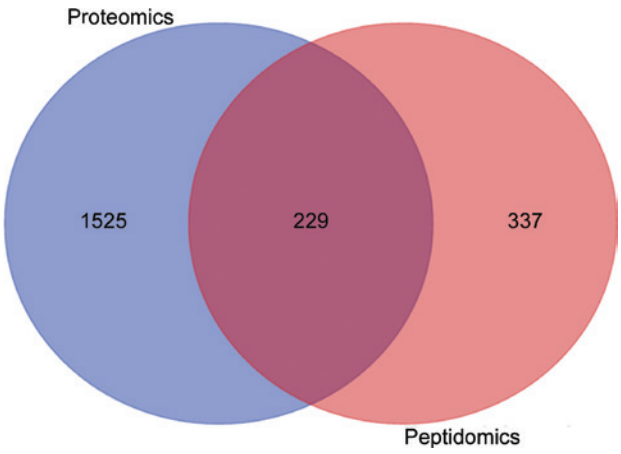

Figure 2: Venn diagram of the 2091 non-redundant proteins detected in normal urine by proteomic and peptidomic analysis. The total number of proteins identified was 1754 (proteomics) and 566 (peptidomics).

Comparison with other proteomic and peptidomic studies

We compared the 1754 proteins identified exclusively by our proteomics approach to three independent large-scale studies [20–22]. The overlap between our proteomic approach and the three other studies is illustrated in Figure 3A. As shown in Supplemental Table 4, 289 proteins were uniquely identified by our proteomic approach. We also compared the 2091 proteins identified by our integrated approach to the same three large-scale proteomic studies. Figure 3B illustrates the overlap between our integrated method and the three independent studies. As shown in Supplemental Table 5, we identified 445 proteins that were unique to our integrated analytical approach. Among the unique proteins detected by this study, 10 (SOST, FXYD4, CYP4A11, TDGF1, MT1H, PTH1R, GLYAT, HAVCR1, PAH and PDZK1) were found to display elevated transcript expression in the kidney according to The human protein atlas (data not shown).

We also compared the 566 precursor proteins identified by our peptidomic approach to an independent large-scale study by Liu et al. [27]. The overlap between our peptidomic approach and the study by Liu et al. is shown in Figure 4. Moreover, 494 precursor proteins were uniquely identified by our peptidomic approach (Supplemental Table 6).

In addition to protein identifications, we were able to determine the relative abundance of identified proteins using our integrated analytical approach. To determine whether abundant proteins are consistently present in the urine we assessed the overlap between one hundred of the most abundant proteins identified in our study to three independent large-scale studies [20–22]. As shown in Supplemental Figure 1, there was a 96% overlap between the 100 most abundant proteins identified in our study and the three independent studies.

Characterization of the normal urinary proteome by gene ontology analysis

Identified proteins were categorized based on GO annotation including biological process, molecular function and cellular component using DAVID 6.7 search engine. Figure 5A shows the most significantly enriched terms in the biological process category. Proteins (protein numbers in brackets) were found to map to cell adhesion (241), proteolysis (174)

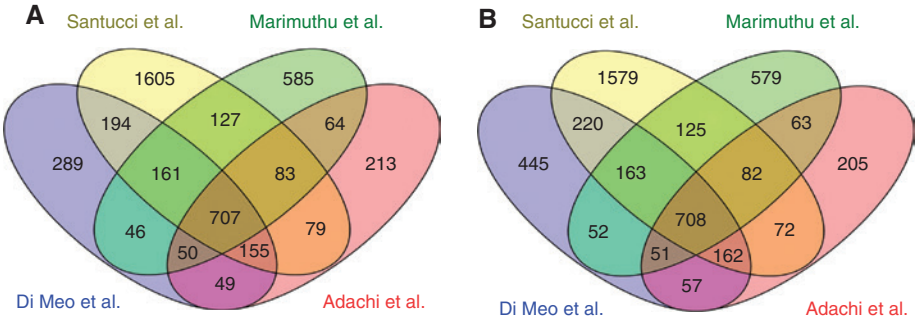

Figure 3: Comparison between our proteomic and integrated approach and three independent large scale proteomic studies. (A) Comparison between proteins identified by our proteomic analysis and three independent studies by Santucci et al. [22], Marimuthu et al. [21] and Adachi et al. [20]. Two hundred and eighty-nine proteins were uniquely identified by our proteomic approach alone. (B) Comparison between our integrated proteomic and peptidomic approach and three independent studies by Santucci et al. [22], Marimuthu et al. [21] and Adachi et al. [20]. A total of 445 proteins were uniquely identified by our integrated analytical approach.

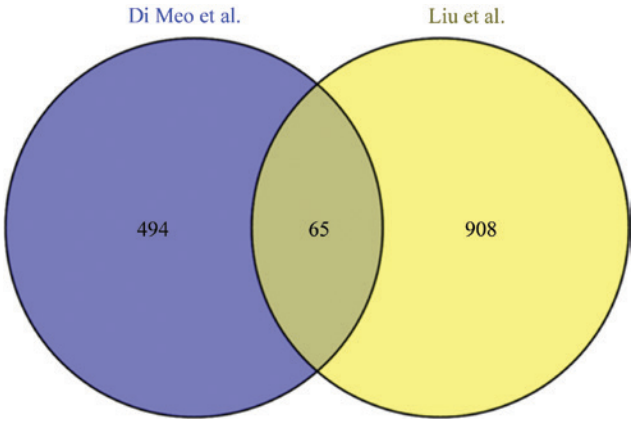

Figure 4: Comparison between our peptidomic approach and an independent study by Liu et al. [27].
A total of 494 proteins were uniquely identified by our peptidomic approach.

and response to wounding (166). Classification of proteins based on molecular function shows that a large majority of proteins are involved in calcium ion binding (236), peptidase activity (138) and carbohydrate binding (133). Figure 5B shows the most significantly enriched terms. Classification of proteins based on cellular component revealed that a large number of proteins are derived from the plasma membrane (610) and extracellular region (603). Figure 5C shows the most significantly enriched terms.

In addition, we compared our significant GO annotations ($p<0.001$) to over-represented GO terms reported by Adachi et al. [20]. As shown in Supplemental Figure 2, the overlap between the two studies was approximately 30%. Overlapping annotations included calcium ion binding, growth factor binding, extracellular matrix (ECM) and cell surface.

Characterization of the normal urinary peptidome by gene ontology analysis

Identified proteins were mapped to GO categories including biological process, molecular function and cellular component using the DAVID 6.7 search engine. Supplemental Figure 3A shows the most significantly enriched terms in the biological process category. Proteins mapped to cell adhesion (79), response to wounding (70) and proteolysis (64) in the biological process category. As shown in Supplemental Figure 3B, proteins mapped to calcium ion binding (76), structural molecular activity (70), and peptidase activity (46) in the molecular function category. Moreover, Supplemental Figure 3C shows the most

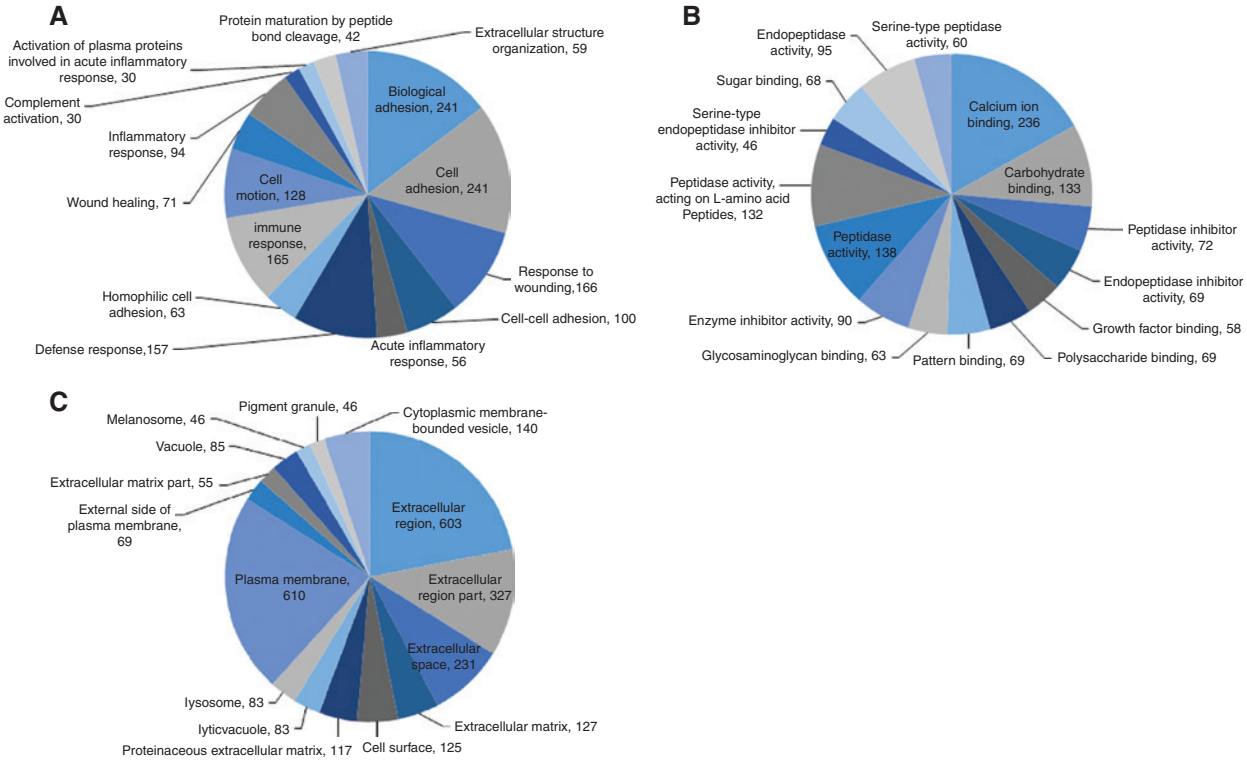

Figure 5: Gene ontology annotations of proteins identified by proteomic analysis.
(A) Biological process, (B) molecular function, and (C) cellular component annotations of identified proteins.

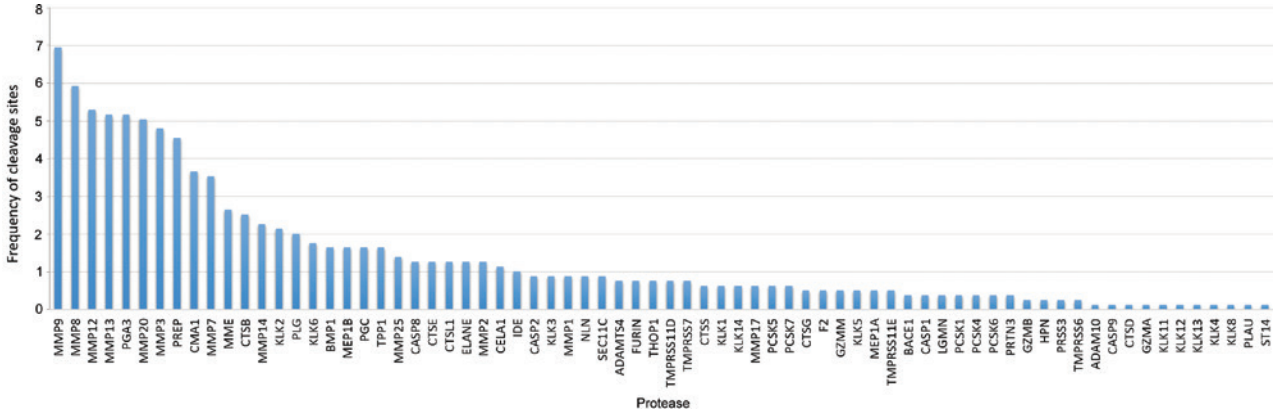

Figure 6: Predicted protease activity on the endogenous urinary peptide signature.
The percentage of predicted protease/cleavage site combinations was calculated in comparison to the total number of potential cleavage sites. Up to three amino acid substitutions were allowed.

significantly enriched terms in the cellular component category. Proteins mapped to extracellular region (260) and extracellular region part (161).

KEGG enrichment analysis

Significantly enriched pathways were identified. Overlap was observed between the two approaches. Supplemental Tables 7 and 8 show a list of the significantly enriched pathways by proteomic and peptidomic analysis, respectively. For proteomic analysis, complement and coagulation cascades (52), cell adhesion molecules (57), and systemic lupus erythematosus (42) were significantly enriched. For peptidomic analysis, complement and coagulation cascades (23), focal adhesion (28) and systemic lupus erythematosus (17) were significantly enriched.

In silico protease prediction analysis

Using Proteasix, the native urinary peptide signature was screened against a database that contains 3500 protease/cleavage site combinations of 191 proteases. As shown in Figure 6, matrix metalloproteases were found to be predominantly involved in urinary peptide generation with MMP-9, -8, -12, and -13 displaying the highest predicted involvement in the generation of the native peptide signature. Kallikreins showed a more limited contribution to the generation of the peptide signature.

Peptide sequence alignment analysis

A computation program called PepEx was used to map the position of endogenous peptides within their precursor proteins. APOA1, APOA2, MMP9, FXYD2, and MUC1 were selected for peptide sequence alignment. As shown in Supplemental Figure 4A, endogenous peptides from APOA1 are derived from the N- and C-terminus, with peptides aligning at the C-terminus being the most abundant. As shown in Supplemental Figure 4B, peptides from APOA2 were found to be most abundantly derived from the C-terminus. Supplemental Figure 4C illustrates that endogenous peptides from MMP-9 are most abundantly derived between amino acids 341–392. Supplemental Figure 4D shows that peptides are most abundantly derived from the N-terminus of FXYD2. Peptides are most abundantly derived from the C-terminus of MUC1 (data not shown).

Discussion

A number of studies have characterized the urine proteome from healthy individuals. Here, we aimed to expand the normal urine proteome by combining both proteomic and peptidomic analysis of urine. Using a label-free LC-MS/MS approach, we identified 1754 proteins by proteomic analysis in addition to 4543 endogenous peptides arising from 566 proteins by peptidomic analysis. To our knowledge this is the largest number of normal urine peptides published to date. Together our integrated proteomic and peptidomic analytical approach identified a total of 2091 non-redundant proteins. We compared our proteomic and integrated approach to three independent large-scale studies [20–22]. We were able to identify 289 unique proteins by our proteomic approach alone. Moreover, 445 proteins were uniquely identified by our integrated analytical approach. We also compared our peptidomic approach to the study by Liu et al. [27]. We were able to identify 494

unique precursor proteins by our peptidomic approach alone. Furthermore, we assessed the overlap between one hundred of the most abundant proteins identified in our study to three independent large-scale studies. There was a 96% overlap between the 100 most abundant proteins identified in our study and three independent large-scale studies [20–22]. This suggests that abundant proteins are a stable and recurrent component of the normal urine proteome. Moreover, inter-individual variation was 0.65 for our integrated proteomic and peptidomic approach. Our result is consistent with another study that assessed variability of the normal urinary proteome, reporting an inter-individual variation of 0.66 [33].

GO is a valuable annotated database providing a functional description of gene products [34]. Proteins identified by proteomic and peptidomic analysis were separately characterized based on GO annotation terms. We observed an overlap between proteomic and peptidomic characterization for biological process, molecular function, and cellular component annotations. Proteins were also subjected to pathway enrichment analysis. Again, we observed an overlap between significantly enriched pathways. Of interest was the unexpected enrichment of systemic lupus erythematosus. Systemic lupus erythematosus has previously been shown to correlate with age. In fact, systemic lupus erythematosus is thought to be a result of reduced clearance of nucleosomal antigens by DNase-1 [35]. Interestingly, DNase-1 was found to be reduced in the urine proteome of the elderly [36]. Thus, we speculate that the identification of systemic lupus erythematosus is due to the advanced age of our subjects. In addition, we compared our significant GO annotations to over-represented GO terms reported by Adachi et al. The overlap between the two studies was approximately 30%. We speculate that differences in significant GO terms may have resulted from evaluating the urinary proteome of different age groups. Here, we focused on a homogeneous elderly population ranging between 62 and 79 years of age whereas the study by Adachi et al. focused nine healthy individuals between 26 and 61 years of age [20]. Both pathway enrichment and GO analyses provided valuable insight into normal physiological processes occurring within our sample population. In addition, pathway enrichment analysis can also be utilized to identify dysregulated pathways in disease, which can help discriminate disease from control groups. In fact, one study found that dysregulated pathways can serve as better biomarkers for disease compared with single gene markers [37]. Thus, GO and pathway enrichment analysis of the normal urine proteome and peptidome can serve as a basis for assessing biological variations in disease.

Proteases have a critical role in physiological and pathophysiological processes by irreversibly modifying the function of their substrate protein [38, 39]. Naturally occurring urinary peptides are likely to be the byproduct of upstream proteolytic processing. Using N- and C-terminal cleavage sequences of endogenous peptides, protease activity can be predicted. Thus, we performed protease prediction analysis to assess which endogenous proteases are responsible for the observed peptide signature. Consistent with previous reports, our data revealed that metalloproteases are predominantly involved in the generation of the urinary peptide signature. A study by Klein et al. reported a high predicted contribution of MMP-2, -3, -8, -9, -12 and -13 in urinary peptidome generation. Moreover, this group found ADAMTS4 and ADAMTS5 to be over-represented. The authors also reported a limited contribution of plasminogen and kallikreins [31]. Similarly, our study revealed that MMP-9, -8, -12, -13 and -3 significantly contributed to the urinary peptide signature. Moreover, KLK-4, -8, -11 and -12 had a limited contribution whereas KLK-2 and -6 displayed a larger contribution than what was reported by Klein et al. The ability to predict protease activity on the basis of N- and C-terminal cleavage sites allows for a clearer understanding of biological processes occurring during normal physiology. As shown by our study, matrix metalloproteases contribute most to the generation of the observed peptide profile. Matrix metalloproteases regulate physiologic events by remodeling ECM molecules, altering the ECM microenvironment and modulating the activity of biologically active molecules by direct cleavage, release from bound stores or modulating the activity of their inhibitors [40–42]. Furthermore, in silico protease prediction analysis can predict disease-specific changes in protease activity by comparing peptide profiles in health and disease [31, 43]. For instance, although matrix metalloproteases play a crucial role in normal physiology, overexpression of active matrix metalloproteases has been associated with tumor growth, migration, invasion and metastasis [44]. Moreover, protease prediction analysis can help to define new biomarkers, assess therapeutic efficiency of protease inhibitor treatments and guide focused validation experiments [38, 45].

Our peptidomic analysis allowed detection of endogenous peptides derived from proteins implicated in various urological malignancies, including APOA1, APOA2, MMP9, FXYD2, and MUC1. Several reports have demonstrated that urinary levels of APOA1 and APOA2 are elevated in bladder cancer [46, 47]. In addition, elevated levels of MMP9 were found to correlate with poor prognosis in RCC [48, 49]. A study by Gaut et al. was first to report the potential utility of FXYD2 in the diagnosis of chromophobe RCC

[50]. Another study found that MUC1 is overexpressed in clear cell RCC [51]. Furthermore, differential expression of MUC1 in RCC was found to correlate with metastatic disease and poor prognosis [52]. Thus, the characterization of the normal urine proteome and peptidome may prove useful in identifying variations in the abundance of disease biomarkers.

While an extensive list of urinary proteins based on the detection of tryptic peptides has been presented, these studies usually lack measurements of abundance. A clear benefit of our integrated analysis is the robust relative quantification of proteins and endogenous peptides. Moreover, our label-free method has several technical advantages. A label-free strategy allows for direct comparison of MS signals between any number of samples in contrast to label-based methods that have a finite number of "plexes" [53]. It can also capture peptides with N-terminal post-translational modifications, which are omitted in labeling methods that rely on free amine groups [54]. Label-free quantification also eliminates the cost and complexity associated with labeling, minimizing variations in sample preparation. Label-free methods are also applicable to any kind of sample, including many clinical samples that cannot be directly labeled [53, 55]. As a result, this is an ideal strategy for shotgun proteomic and peptidomic analysis. We do acknowledge some limitations of our study. For instance, a small number of subjects were assessed, which likely had an effect on the number of protein identifications. However, our integrated analysis still provided one of the largest sets of protein identifications as well as the largest number of peptides documented in normal human urine to date. In addition, the reported proteome and peptidome profiles are more representative of an older age group since protein signatures were generated by analyzing urine specimens from healthy donors between the ages of 62 and 79 years. This may raise concerns on how representative the data are of the healthy population. However, we speculate that our data is relevant for genitourinary diseases that occur frequently in later life. In fact, the median age for RCC diagnosis is between 60 and 65 years of age [56]. Bladder cancer also occurs most frequently in later life, with a median age of diagnosis at 69 years for men and 71 years for women [57]. Moreover, several studies have assessed the impact of age on the urinary proteome and peptidome [36, 58]. A study by Bakun et al. identified 22 proteins that were differentially expressed in the urine proteome of younger subjects compared to elderly subjects. The protein profiles that differed in the elderly included those involved in tissue remodeling, coagulation and fibrinolysis, low-grade inflammation and immune and metabolic dysregulation [36]. Nkuipou-Kenfack et al. identified age-correlated peptides that predominately originated from collagen, uromodulin and fibrinogen. Interestingly, fibrillary and basement membrane collagen fragments showed decreased age-related excretion whereas uromodulin and fibrinogen fragments showed increased age-associated excretion. In our study, the most abundant endogenous peptides were derived from uromodulin which is consistent with the report by Nkuipou-Kenfack et al. [58]. Thus, the assessment of the urinary proteome and peptidome in healthy individuals with advanced age may prove useful in discriminating physiological alterations that occur as a result of aging from pathophysiological alterations that occur frequently in later life.

Conclusions

Characterization of the normal urinary proteome and peptidome can provide insight into normal physiology and can be the basis for non-invasive biomarker discovery. Employing an integrated label-free LC-MS/MS approach allowed us to characterize a total of 2091 proteins with high-confidence. Since cancer is a disease that occurs more frequently in later life, the data presented here represent a group of healthy individuals within a clinically relevant age range. Our analysis provides one of the largest sets of protein identifications documented in normal human urine and may prove a useful reference for future biomarker discovery.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: This work was supported by grants from the Canadian Institute of Health Research (MOP 119606), Kidney Foundation of Canada (KFOC130030), Kidney Cancer Research Network of Canada and Prostate Cancer Canada Movember Discovery Grants (D2013-39).
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Blaine J, Chonchol M, Levi M. Renal control of calcium, phosphate, and magnesium homeostasis. Clin J Am Soc Nephrol 2015;10:1257–72.

2. Tojo A. The role of the kidney in protein metabolism: the capacity of tubular lysosomal proteolysis in nephrotic syndrome. Kidney Int 2013;84:861–3.
3. Bauca JM, Martinez-Morillo E, Diamandis EP. Peptidomics of urine and other biofluids for cancer diagnostics. Clin Chem 2014;60:1052–61.
4. Dallas DC, Guerrero A, Parker EA, Robinson RC, Gan J, German JB, et al. Current peptidomics: applications, purification, identification, quantification, and functional analysis. Proteomics 2015;15:1026–38.
5. Jurgens M, Appel A, Heine G, Neitz S, Menzel C, Tammen H, et al. Towards characterization of the human urinary peptidome. Comb Chem High Throughput Screen 2005;8:757–65.
6. Di Meo A, Pasic MD, Yousef GM. Proteomics and peptidomics: moving toward precision medicine in urological malignancies. Oncotarget 2016. DOI: 10.18632/oncotarget.8931.
7. Sigdel TK, Salomonis N, Nicora CD, Ryu S, He J, Dinh V, et al. The identification of novel potential injury mechanisms and candidate biomarkers in renal allograft rejection by quantitative proteomics. Mol Cell Proteomics 2014;13:621–31.
8. Frantzi M, Metzger J, Banks RE, Husi H, Klein J, Dakna M, et al. Discovery and validation of urinary biomarkers for detection of renal cell carcinoma. J Proteomics 2014;98:44–58.
9. Rocchetti MT, Centra M, Papale M, Bortone G, Palermo C, Centonze D, et al. Urine protein profile of IgA nephropathy patients may predict the response to ACE-inhibitor therapy. Proteomics 2008;8:206–16.
10. Davalieva K, Kiprijanovska S, Komina S, Petrusevska G, Zografska NC, Polenakovic M. Proteomics analysis of urine reveals acute phase response proteins as candidate diagnostic biomarkers for prostate cancer. Proteome Sci 2015;13:2.
11. Smith CR, Batruch I, Bauca JM, Kosanam H, Ridley J, Bernardini MQ, et al. Deciphering the peptidome of urine from ovarian cancer patients and healthy controls. Clin Proteomics 2014;11:23.
12. Frantzi M, Latosinska A, Fluhe L, Hupe MC, Critselis E, Kramer MW, et al. Developing proteomic biomarkers for bladder cancer: towards clinical application. Nat Rev Urol 2015;12:317–30.
13. Davis MT, Auger PL, Patterson SD. Cancer biomarker discovery via low molecular weight serum profiling--are we following circular paths? Clin Chem 2010;56:244–7.
14. Good DM, Thongboonkerd V, Novak J, Bascands JL, Schanstra JP, Coon JJ, et al. Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future. J Proteome Res 2007;6:4549–55.
15. Mischak H, Julian BA, Novak J. High-resolution proteome/peptidome analysis of peptides and low-molecular-weight proteins in urine. Proteomics Clin Appl 2007;1:792.
16. Thongboonkerd V, McLeish KR, Arthur JM, Klein JB. Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. Kidney Int 2002;62: 1461–9.
17. Pieper R, Gatlin CL, McGrath AM, Makusky AJ, Mondal M, Seonarain M, 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 2004;4:1159–74.
18. Sun W, Li F, Wu S, Wang X, Zheng D, Wang J, et al. Human urine proteome analysis by three separation approaches. Proteomics 2005;5:4994–5001.
19. Li QR, Fan KX, Li RX, Dai J, Wu CC, Zhao SL, et al. A comprehensive and non-prefractionation on the protein level approach for the human urinary proteome: touching phosphorylation in urine. Rapid Commun Mass Spectrom 2010;24:823–32.
20. Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. Genome Biol 2006;7:R80.
21. Marimuthu A, O'Meally RN, Chaerkady R, Subbannayya Y, Nanjappa V, Kumar P, et al. A comprehensive map of the human urinary proteome. J Proteome Res 2011;10:2734–43.
22. Santucci L, Candiano G, Petretto A, Bruschi M, Lavarello C, Inglese E, et al. From hundreds to thousands: Widening the normal human Urinome (1). J Proteomics 2015;112:53–62.
23. Santucci L, Candiano G, Petretto A, Bruschi M, Lavarello C, Inglese E, et al. From hundreds to thousands: Widening the normal human Urinome. Data Brief 2014;1:25–8.
24. Yang X, Hu L, Ye M, Zou H. Analysis of the human urine endogenous peptides by nanoparticle extraction and mass spectrometry identification. Anal Chim Acta 2014;829:40–7.
25. Fiedler GM, Baumann S, Leichtle A, Oltmann A, Kase J, Thiery J, et al. Standardized peptidome profiling of human urine by magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Clin Chem 2007;53:421–8.
26. Mischak H, Kolch W, Aivaliotis M, Bouyssie D, Court M, Dihazi H, et al. Comprehensive human urine standards for comparability and standardization in clinical proteome analysis. Proteomics Clin Appl 2010;4:464–78.
27. Liu X, Chinello C, Musante L, Cazzaniga M, Tataruch D, Calzaferri G, et al. Intraluminal proteome and peptidome of human urinary extracellular vesicles. Proteomics Clin Appl 2015;9:568–73.
28. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 2011;10:1794–805.
29. Luber CA, Cox J, Lauterbach H, Fancke B, Selbach M, Tschopp J, et al. Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. Immunity 2010;32:279–89.
30. Cox J, Hubner NC, Mann M. How much peptide sequence information is contained in ion trap tandem mass spectra? J Am Soc Mass Spectrom 2008;19:1813–20.
31. Klein J, Eales J, Zurbig P, Vlahou A, Mischak H, Stevens R. Proteasix: a tool for automated and large-scale prediction of proteases involved in naturally occurring peptide generation. Proteomics 2013;13:1077–82.
32. Guerrero A, Dallas DC, Contreras S, Chee S, Parker EA, Sun X, et al. Mechanistic peptidomics: factors that dictate specificity in the formation of endogenous peptides in human milk. Mol Cell Proteomics 2014;13:3343–51.
33. Nagaraj N, Mann M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. J Proteome Res 2011;10:637–45.
34. Mutowo-Meullenet P, Huntley RP, Dimmer EC, Alam-Faruque Y, Sawford T, Jesus Martin M, et al. Use of Gene Ontology Annotation to understand the peroxisome proteome in humans. Database (Oxford) 2013;2013:bas062.
35. Lachmann PJ. Lupus and desoxyribonuclease. Lupus 2003;12:202–6.
36. Bakun M, Senatorski G, Rubel T, Lukasik A, Zielenkiewicz P, Dadlez M, et al. Urine proteomes of healthy aging humans

reveal extracellular matrix (ECM) alterations and immune system dysfunction. Age 2014;36:299–311.
37. Liu KQ, Liu ZP, Hao JK, Chen L, Zhao XM. Identifying dysregulated pathways in cancers from pathway interaction networks. BMC Bioinformatics 2012;13:126.
38. Overall CM, Blobel CP. In search of partners: linking extracellular proteases to substrates. Nat Rev Mol Cell Biol 2007;8:245–57.
39. Lyons PJ, Fricker LD. Peptidomic approaches to study proteolytic activity. Curr Protoc Protein Sci 2011;Chapter 18:Unit 18.13.
40. Page-McCaw A, Ewald AJ, Werb Z. Matrix metalloproteinases and the regulation of tissue remodelling. Nat Rev Mol Cell Biol 2007;8:221–33.
41. Vu TH, Werb Z. Matrix metalloproteinases: effectors of development and normal physiology. Genes Dev 2000;14:2123–33.
42. Klein T, Bischoff R. Physiology and pathophysiology of matrix metalloproteases. Amino Acids 2011;41:271–90.
43. Turk B. Targeting proteases: successes, failures and future prospects. Nat Rev Drug Discov 2006;5:785–99.
44. Szarvas T, vom Dorp F, Ergun S, Rubben H. Matrix metalloproteinases and their clinical relevance in urinary bladder cancer. Nat Rev Urol 2011;8:241–54.
45. Barkan DT, Hostetter DR, Mahrus S, Pieper U, Wells JA, Craik CS, et al. Prediction of protease substrates using sequence and structure features. Bioinformatics 2010;26:1714–22.
46. Chen YT, Chen CL, Chen HW, Chung T, Wu CC, Chen CD, et al. Discovery of novel bladder cancer biomarkers by comparative urine proteomics using iTRAQ technology. J Proteome Res 2010;9:5803–15.
47. Chen CL, Lin TS, Tsai CH, Wu CC, Chung T, Chien KY, et al. Identification of potential bladder cancer markers in urine by abundant-protein depletion coupled with quantitative proteomics. J Proteomics 2013;85:28–43.
48. Kawata N, Nagane Y, Hirakata H, Ichinose T, Okada Y, Yamaguchi K, et al. Significant relationship of matrix metalloproteinase 9 with nuclear grade and prognostic impact of tissue inhibitor of metalloproteinase 2 for incidental clear cell renal cell carcinoma. Urology 2007;69:1049–53.
49. Sato A, Nagase H, Obinata D, Fujiwara K, Fukuda N, Soma M, et al. Inhibition of MMP-9 using a pyrrole-imidazole polyamide reduces cell invasion in renal cell carcinoma. Int J Oncol 2013;43:1441–6.
50. Gaut JP, Crimmins DL, Lockwood CM, McQuillan JJ, Ladenson JH. Expression of the Na+/K+-transporting ATPase gamma subunit FXYD2 in renal tumors. Mod Pathol 2013;26:716–24.
51. Langner C, Ratschek M, Rehak P, Schips L, Zigeuner R. Expression of MUC1 (EMA) and E-cadherin in renal cell carcinoma: a systematic immunohistochemical analysis of 188 cases. Mod Pathol 2004;17:180–8.
52. Aubert S, Fauquette V, Hemon B, Lepoivre R, Briez N, Bernard D, et al. MUC1, a new hypoxia inducible factor target gene, is an actor in clear renal cell carcinoma tumor progression. Cancer Res 2009;69:5707–15.
53. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 2014;13:2513–26.
54. Romanova EV, Dowd SE, Sweedler JV. Quantitation of endogenous peptides using mass spectrometry based methods. Curr Opin Chem Biol 2013;17:801–8.
55. Olsen JV, Mann M. Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol Cell Proteomics 2013;12:3444–52.
56. Thompson RH, Ordonez MA, Iasonos A, Secin FP, Guillonneau B, Russo P, et al. Renal cell carcinoma in young and old patients–is there a difference? J Urol 2008;180:1262–6; discussion 6.
57. Taylor JA, 3rd, Kuchel GA. Bladder cancer in the elderly: clinical outcomes, basic mechanisms, and future research direction. Nat Clin Pract Urol 2009;6:135–44.
58. Nkuipou-Kenfack E, Bhat A, Klein J, Jankowski V, Mullen W, Vlahou A, et al. Identification of ageing-associated naturally occurring peptides in human urine. Oncotarget 2015;6:34106–17.

Supplemental Material: The online version of this article (DOI: 10.1515/cclm-2016-0390) offers supplementary material, available to authorized users.