A comprehensive analysis and annotation of human normal urinary proteome

Mindi Zhao1,2, Menglin Li1,3, Yehong Yang4, Zhengguang Guo4, Ying Sun5, Chen Shao1, Mingxi Li5, Wei Sun4 & Youhe Gao6

Biomarkers are measurable changes associated with the disease. Urine can reflect the changes of the body while blood is under control of the homeostatic mechanisms; thus, urine is considered an important source for early and sensitive disease biomarker discovery. A comprehensive profile of the urinary proteome will provide a basic understanding of urinary proteins. In this paper, we present an in-depth analysis of the urinary proteome based on different separation strategies, including direct one dimensional liquid chromatography–tandem mass spectrometry (LC/MS/MS), two dimensional LC/MS/MS, and gel-eluted liquid fraction entrapment electrophoresis/liquid-phase isoelectric focusing followed by two dimensional LC/MS/MS. A total of 6085 proteins were identified in healthy urine, of which 2001 were not reported in previous studies and the concentrations of 2571 proteins were estimated (spanning a magnitude of 106) with an intensity-based absolute quantification algorithm. The urinary proteins were annotated by their tissue distribution. Detailed information can be accessed at the “Human Urine Proteome Database” (www.urimarker.com/urine).

Urine is associated with glomerular filtration, tubular reabsorption and secretion. Biomarkers are measurable changes associated with the disease. Because urine can accumulate changes from the body, it is considered to be one of the most attractive sources for early and sensitive biomarker discovery. Urinary proteomic studies have identified many candidate biomarkers for various urogenital diseases, such as acute kidney injury, bladder cancer and diabetic nephropathy. As urinary proteins are composed largely of filtered plasma proteins, the urinary proteome is also considered to be valuable for detecting a broad range of complex disorders, such as encephalopathy, heart failure and intestinal ischemia.

In the biomarker discovery process, it is essential to comprehensively profile the normal urinary proteome as a baseline reference. With the rapid development of mass spectrometry (MS), larger numbers of urinary proteins were identified by various strategies. In 2001, Patterson et al. first identified 124 urinary proteins. In 2005, Sun et al. identified 226 proteins in normal urine with quality control LC/MS/MS data. In 2006, Adachi et al. reported the first urinary proteome result (1543 proteins) from high resolution mass spectrometry. In 2011, 1823 urinary proteins were identified by high resolution MS and MS/MS. Many efforts have been made to identify more urinary proteins in recent years. Currently, the human urine PeptideAtlas database contains a total of 23,739 peptides corresponding to 2487 proteins.

In 2014, two large-scale MS-based drafts of the human proteome identified 17,294 and 18,097 human gene products from 30 and 44 tissues and body fluids, respectively. In each study, the number of identified proteins was quite large and even approached the number of protein-coding genes in the complete human genome.
Compared with the depth of the human tissue proteome, the urinary proteome has been relatively less studied. We are curious about how many proteins could be identified in human urinary proteome. Therefore, we performed an in-depth urinary proteomic analysis using normal human urine samples. And to achieve maximal urinary proteome coverage one-, two- and three-dimensional separation strategies (Fig. 1) were employed in this study. By in-depth analysis, a readily obtainable source for the human urinary proteome, “Human Urinary Proteome Database” could be provided. The comparison of three separation strategies could provide detailed information about the potential application of different separation methods. The detailed workflow was as followings: In one-dimensional (1D) separation, digested urinary peptides were directly analyzed by 1D liquid chromatography-tandem mass spectrometry (LC/MS/MS). In two-dimensional (2D) separation, urinary peptides were fractionated by offline high-pH reverse-phase liquid chromatography (RPLC) prior to analysis by 1D LC/MS/MS. In three-dimensional (3D) separation, urinary proteins were first fractionated by gel-eluted liquid fraction entrapment electrophoresis (GELFrEE) or liquid-phase isoelectric focusing (LP-IEF) and urinary peptides digested from GELFrEE/LP-IEF fractions were fractionated by RPLC as performed for 2D separation and finally analyzed by 1D LC/MS/MS. In total, 383 LC/MS/MS runs were analyzed by high-resolution TripleTOF 5600 MS. A urine proteome database was then constructed based on bioinformatics analyses.

Results and Discussion

Comprehensive identification of urinary proteome. In this study, pooled urine samples were used to establish a large database of urinary proteins. The following filters were used to select the final protein identification list (1). The FDR at the protein level was set to < 1%, and (2) each protein should include at least two unique peptides. When identified peptides were shared between two proteins, they were combined and reported as one protein group. The results from 1D LC, 2D LC and 3D LC yielded average FDRs of 0.10%, 0.26% and 1% at the spectrum, peptide and protein levels, respectively (Supplemental File 1). Then the datasets were combined together with Scaffold perSPECtives.

In 1D analysis, 808 protein groups were identified in three technical replicates, and the protein-overlapping rate was 86.3%, indicating the superior reproducibility of LC/MS analysis. In 2D analysis, a total of 3162 protein groups were identified. In 3D analysis, urinary proteins were first separated by GELFrEE/LP-IEF (Fig. 2A,B). GELFrEE enables mass range proteome separations based on molecular weight (MW), and IEF fractionates proteins according to isoelectric point (pI)11,12. The GELFrEE and LP-IEF fractions were then further separated by RPLC, and a total of 6085 protein groups were identified. The overlap among the proteins identified in the 1D, 2D and 3D analyses is displayed in Fig. 2C. Almost all proteins from the 1D and 2D analyses were included in the 3D results except for 9 and 15 proteins from the 1D and 2D results, respectively. The possible reasons why these proteins cannot be identified in the 3D methods were still unknown. Maybe these proteins were lost during 2D or 3D separate by high pH RPLC or IEF/GELFrEE. It may be also as a result of the random sampling of DDA detection modes. Therefore, we are not sure whether these proteins were false positive identification or not. Then these proteins (Supplemental File 2) were removed from the subsequent analysis to ensure data accuracy and reliability. Thus, the whole urine proteome dataset eventually contained 6085 protein groups (Supplemental Table 1).

Several studies have been conducted to characterize the normal human urinary proteome. Table 1 summarizes the current largest-scale studies of human urine and urinary exosomes using high-resolution MS12–16,23–25. The protein accessions in each dataset were mapped to the corresponding gene IDs26,27. Total nine large-scale urinary and exosome proteomic analyses were performed in recent years. When all of the data from these nine studies were combined, a total of 8021 gene products were detected in the human urinary proteome (Supplemental...
Table 2). When comparing previous data with our results (Fig. 2D), total 2001 gene products were uniquely identified in this study. The possible reasons of differences in urine proteome between different studies may be genetic factors, individual variations, different separate methods and MS preference.

Urinary proteins, which are considered to represent the protein composition of the output of the kidneys, are primarily composed of proteins derived from plasma filtration and urinary tract system secretion. A comparative analysis of the urine, plasma and kidney proteome would provide a more concrete link to determine how many plasma- and kidney-related proteins could be detected in urine. The PeptideAtlas builds yielded 3553 and 4005 non-redundant proteins at 1% FDR for the plasma and kidney proteomes. In contrast, a total of 2940 (47.7%) and 3238 (52.6%) of the gene products identified in this urinary proteome study were common to the gene products that were reported in the plasma (81.1%) and kidney proteomes, respectively. According to previous report, approximately 30% of urinary proteins originate from the plasma proteins, whereas 70% comes from the kidneys.

Table 1. Recent large-scale proteomic studies of healthy human urine.
from the kidney and the urinary tract. From our study, maybe the difference between plasma and urine is smaller than expected. And it might be possible that more common proteins will be identified with the development of MS in the future. By comparison with kidney proteome, we want to show the overlap between urine and kidney proteome. The large overlap may give evidences that urine can better reflect the functions of kidney.

Quantitative analysis of urinary proteins. Quantitative analysis of urinary proteins will improve our understanding of the urinary proteome and will facilitate the development of urinary biomarkers. Accordingly, we aimed to quantify each protein using the iBAQ algorithm, which provides a rough indication of actual protein levels. In 3D analysis, equal amounts of protein from each GELFrEE/LP-IEF fraction were used for LC/MS/MS analysis, which thus cannot provide an accurate quantitative analysis. Therefore, data from the 2D strategy were used, and a total of 2571 proteins were quantified with the iBAQ algorithm. The dynamic range of relative abundance spanned six orders of magnitude (Fig. 3A), which was consistent with previous observations. Considering that more than 3000 proteins in the 3D analysis were not quantified, the dynamic range is expected to be even bigger. In the present analysis, serum albumin and uromodulin were the most abundant urinary proteins.

The average concentration of urinary albumin, which was one of the most easily detected urinary proteins, was approximately 2.2–3.3 µg/mL in normal human urine. With the iBAQ value ratios, concentrations of the other 2570 proteins could subsequently be estimated (Supplemental Table 3). The estimated concentration of arginine-tRNA ligase (RARS), which had the lowest relative abundance in the 2D analysis, was 0.68–1.02 pg/mL. As the 2D results contained almost all of the proteins from the 1D separation, concentrations of the 753 proteins from the 1D analysis could be inferred. Among them, ribonuclease K6 (RNASE 6) was the least abundant protein in 1D analysis. (2) RARS: arginine-tRNA ligase, the least abundant protein in 2D analysis; (3) RNASE 6: ribonuclease K6, the least abundant protein in 1D analysis. (B) Correlation plot between estimated concentrations and immunoassays results.

Functional annotation of three separation strategies. Functional annotations of urinary proteins based on the degree of analysis depth may be helpful in providing insight into the analysis approach difference in protein composition, reflecting pathophysiological states and determining suitable separation methods for some diseases. To analyze the protein identification data from the three separation strategies, 6085 proteins were divided into three groups as follows: Group 1D, proteins identified in 1D analysis (799 proteins); Group 2D, proteins identified in 2D analysis, excluding those identified in the 1D analysis (2362 proteins); and Group 3D, proteins identified in 3D analysis, excluding those identified in both the 1D and 2D analyses (2924 proteins).

IPA analysis was performed to provide insight into the functions of the three groups (Fig. 4 and Supplemental Table 5). Extracellular proteins and plasma membrane proteins were enriched in Group 1D (56%), as previously reported. The most significant pathways in Group 1D were functionally similar to plasma components, such as inflammatory responses, coagulation and glucose metabolism. Acute phase response signaling, which is one of the top pathways for Group 1D, is a rapid inflammatory response that provides protection against some infections by nonspecific defenses. It consists of an increase in inflammatory factors (such as IL-1) and a change in the levels of several plasma proteins (such as ALB and APOA1/2). For example, alpha-1-acid glycoprotein 1 (ORM1), an extracellular protein, is involved in the acute phase response. Overexpression of ORM1 in urine was associated with acute pediatric appendicitis. As Group 2D demonstrated considerable enrichment of intracellular proteins (58%), most of the pathways were involved in cellular signaling such as EIF2 Signaling and Regulation of eIF4 and p70S6K signaling. Proteins in Group 3D were also over-represented in the cytoplasm and nucleus (63%). Most of the canonical pathways in Group 3D were closely related to interleukin signaling.

Figure 3. Quantitative analysis of urinary proteins by the iBAQ method. (A) The relative expression and concentrations of 2,571 proteins in the 2D analysis were estimated by iBAQ. The left y-axis represents relative abundance, and the right y-axis represents estimated concentration (µg/mL). (1) ALB: albumin; UROM: uromodulin, the two most abundant proteins. (2) RARS: arginine-tRNA ligase, the least abundant protein in 2D analysis; (3) RNASE 6: ribonuclease K6, the least abundant protein in 1D analysis. (B) Correlation plot between estimated concentrations and immunoassays results.
Considering the above results, we assumed that proteins in the three groups were functionally different. If the purpose of research was to study basic physiological activities, such as cell movement and proliferation, maybe the urinary proteome can be analyzed without further separation in most cases. If aim at intracellular activities and functions of organs, maybe the in-depth analysis is necessary.

**Landscape of proteins detected in urine.** Previous studies reported that urine might reflect kidney function and identified some potential biomarkers of kidney disease. According to previous analyses, the Human Urinary Proteome Database contains proteins localized in the glomeruli of nephron segments (Table 2) and molecules to detect injuries to specific tubules of nephron segments (Table 3). Extracellular macromolecular laminin, type IV collagen α3α4α5, heparan sulfate proteoglycan agrin, and nidogen were the main components of the glomerular basement membrane (GBM) and could all be identified by the one-dimensional method without fractionation. Nephrin and podocin are both specifically expressed in the slit diaphragm, which is pivotal in maintaining the selective permeability of podocytes in the glomerular filtration barrier. The cytoplasmic protein CD2-associated protein (CD2AP) localizes to the podocyte slit diaphragm where it has been shown to bind to nephrin and podocin. The above three podocyte-related proteins could be identified in Group 2D. The fatty acid-binding proteins (FABPs) in Group 2D are a class of small intracellular proteins that bind long chain fatty acids. Liver-type FABP is mainly present in the cells of the proximal tubules, while heart-type FABP is predominantly localized in the distal tubules. The above results showed that both glomerulus and tubules-related proteins could be found in the urine, which indicated that the urine proteome might reflect changes of kidney function.

Some tissue or serum biomarkers of kidney diseases could also be detected in our urine proteome database. For example, the phospholipase A2 receptor (PLA,R), a plasma membrane glycoprotein located on normal podocytes, was a major target antigen in idiopathic membranous nephropathy. PLA,R could be detected in Groups 2D and 3D. Urokinase plasminogen activator surface receptor (uPAR) is a glycosylphosphatidylinositol-anchored three-domain protein and is expressed in human glomerular cells. Serum concentrations of soluble uPAR are significantly elevated in most subjects with primary focal segmental glomerulosclerosis (FSGS). If these tissue or serum biomarkers could be confirmed as urinary biomarkers, the human urinary proteome database would provide a convenient way to discover noninvasive urinary candidate biomarkers. In addition to kidney diseases, previous studies also reported that some other diseases, such as acute pancreatitis, might possess urinary biomarkers. The human urinary proteome database provides brief information on known biomarkers for predicting various types of organ injury (Supplemental Table 6).

Moreover, these proteins detected in urine were annotated by their tissue distribution based on an integrated omics approach that involves quantitative transcriptomics and tissue microarray–based immunohistochemistry in previous studies. The detailed annotation data of each protein were shown in the following database. The tissue with maximum numbers of highly expressed proteins detected in urine both at protein and mRNA levels was brain (Fig. 5A, Supplemental Figure 1). Other tissues with more highly expressed proteins were mostly digestive.
organ such as colon and stomach. As expected, more tissue-related proteins could be detected in Group 2D and 3D than in Group 1D (Fig. 5B).

The Human Urinary Proteome Database. To provide a readily obtainable source for the human urinary proteome, the "Human Urinary Proteome Database" was constructed (Fig. 6) based on the above analyses. The database was constructed using open source technologies and is freely available at www.urimarker.com/urine. A total of 3048 spectra, 68151 unique peptides and 6085 proteins are included, along with detailed information such as the protein name, accession number, peptide sequence, sequence coverage and unique peptide number. Each protein is featured with annotated data, including relative quantitative information, estimated concentrations, theoretical and experimental MW and pI. Remarkably, some high-abundance proteins were observed spanning multiple fractions in both the GELFrEE and LP-IEF separations. It is generally accepted that mass/pI deviation may occur due to the presence of fragments, protein polymers, isoforms, protein degradation, post-translational modifications and low focusing quality in the basic region of the immobilized pH-gradient strips, as well as due to the pI prediction algorithm used.

The Human Urinary Proteome Database serves as a reference repository for urinary proteins, as it offers the largest number of such proteins to date. All of the data retrieved from three separations not only detail the normal human urinary proteome but also categorize all proteins by different separation methods. Moreover, the database can be used for targeted proteomics that rely on the proper selection of peptides and transitions to guide the selection of proteotypic peptides for candidate proteins.

| Protein Name | Uniprot ID | Protein in Group | Nephron segment | Location | Molecular Function | Biomarker Application | Reference |
|--------------|------------|------------------|-----------------|----------|--------------------|-----------------------|----------|
| Podocin      | Q9N985     | 2D               | Podocyte & slit diaphragm | Plasma Membrane | other             | IgA nephropathy, membranous nephropathy | 57, 58   |
| Alpha-actinin-4 | O43707     | 1D               | Podocyte | Cytoplasm | other             | Diabetic nephropathy, focal segmental glomerulosclerosis | 59, 60   |
| Nephrin      | P08473     | 1D               | Podocyte | Plasma Membrane | peptidase | Glomerulonephritis | 61       |
| Myosin-9     | P33579     | 1D               | Podocyte & mesangial cells | Cytoplasm | enzyme             | Glomerulopathy | 62       |
| Agrin        | P00468     | 1D               | Glomerular basement membrane | Plasma Membrane | other             | Diabetic nephropathy, transplant glomerulopathy | 63, 64   |
| Collagen alpha-3(VI) chain | P12111    | 1D               | Glomerular basement membrane | Extracellular Space | other             | Alport syndrome, diabetic nephropathy | 65, 66   |
| Nidogen      | P14543, Q14112 | 1D           | Glomerular basement membrane | Extracellular Space | other             | Membranous nephropathy | 67       |
| Laminin      | Multiple Ma | 1D               | Glomerular basement membrane | Extracellular Space | other             | Diabetic nephropathy | 68       |
| Nephrin      | O60500     | 2D               | Podocyte | Plasma Membrane | other             | Diabetic nephropathy | 69       |
| CD2-associated protein | Q9Y5K6     | 2D               | Podocyte | Cytoplasm | other             | Focal segmental glomerulosclerosis | 70       |
| Podocalyxin  | O00592     | 1D               | Podocyte & parietal epithelial cells | Plasma Membrane | kinase            | Diabetic nephropathy | 71       |
| Vascular endothelial growth factor | P13692, P49767, P49765 | 3D         | Podocyte | Extracellular Space | growth factor | Diabetic nephropathy | 72       |
| Proliferating cell nuclear antigen | P12004 | 3D               | Parietal epithelial cells & podocyte | Nucleus | enzyme            | Schistosomal nephropathy | 73       |
| Secretory phospholipase A2 receptor | Q13018 | 2D               | Glomerulus | Plasma Membrane | transmembrane receptor | Idiopathic membranous nephropathy | 41       |
| Complement C3 | P01024 | 1D               | Glomerular basement membrane, mesangium, capillary loops | Extracellular Space | peptidase         | Lupus nephritis | 74       |
| Apolipoprotein E | P06499 | 1D               | Mesangial cells | Extracellular Space | transporter | Diabetic nephropathy | 75       |
| CD151 antigen | P48509 | 2D               | Podocyte, glomerular basement membrane | Plasma Membrane | other             | Type 1 diabetic nephropathy | 76       |
| Cofilin-1    | P23528     | 1D               | Podocyte | Nucleus | other             | Hypertension-induced renal damage | 77       |
| Fibronecin   | P02751     | 1D               | Mesangial and subendothelial cells | Extracellular Space | enzyme             | Glomerulopathy with fibronecin deposits | 78       |
| Myeloperoxidase | P05164 | 1D               | Glomerular capillary | Cytoplasm | enzyme             | Anti-neutrophil cytoplasmic antibody-associated glomerulonephritis | 79       |

Table 2. Urinary candidate biomarkers of glomerular injury.
Materials and Methods

Ethics statement. Prior to study enrollment, all of the healthy volunteers were given a verbal explanation of the study and each participant signed an informed consent document. The consent procedure and the research protocol were approved by the Medical Ethics Committee of Peking Union Medical College. All methods in this study were performed in accordance with the guidelines and regulations.

Experimental design and statistical rationale. Twenty-four healthy volunteers (38 ± 11 years old), including twelve males and twelve females, were enrolled. Exclusion criteria included the following conditions: menstrual bleeding, any prescription drug use and acute or chronic medical illness. The age, sex and smoking habits of the healthy subjects were recorded (Supplemental File 3).

After random urine collection, all of the samples were immediately centrifuged for 30 min at 3,500 g. Next, the samples underwent SDS removal using Pierce Detergent Removal Spin Columns (Pierce, Rockford, IL, USA). For LP-IEF fractionation, urinary proteins were desalted and cleaned using Amicon Ultrafiltration devices with a 10-kDa molecular weight cutoff (Merck Millipore Inc., Billerica, MA, USA). Then, the desalted urinary proteins were focused (approximately 2.5 h at 1 W) using a ten-chamber Microrotofor LP-IEF system (Bio-Rad, Hemel Hempstead, UK). Ten IEF fractions were collected; few protein bands appeared in fractions 7–10. Then fractions 6–10 were pooled into one sample. 

Table 3. Urinary candidate biomarkers of tubular injury.

| Protein Name | Uniprot ID | Protein Group | Nephron segment | Location | Molecular Function | Biomarker Application                                                                 | Ref. |
|--------------|------------|---------------|----------------|----------|--------------------|---------------------------------------------------------------------------------------|------|
| Beta-2-microglobulin | P61769 | 1D | Proximal tubule | Plasma Membrane | transmembrane receptor | Acute renal allograft rejection, acute kidney injury, diabetic nephropathy | 81, 82 |
| GST-alpha | P09210 | 1D | Proximal tubule | Cytoplasm | enzyme | Acute kidney injury | 83 |
| GSTP1 | P09211 | 1D | Distal tubule | Cytoplasm | enzyme | Acute renal failure | 81 |
| Clusterin | P19090 | 1D | Proximal tubule & distal tubule | Cytoplasm | other | Renal-cell carcinoma, acute kidney injury | 84 |
| Cabilin | O66494 | 1D | Proximal tubule | Plasma Membrane | transmembrane receptor | Type 1 diabetes | 85 |
| Liver-type fatty acid-binding protein | P07148 | 2D | Proximal tubule | Cytoplasm | transporter | Diabetic nephropathy, contrast nephropathy, IgA nephropathy | 40 |
| Heart-type fatty acid-binding protein | P05413 | 2D | Distal tubule | Cytoplasm | transporter | Acute kidney injury after cardiac surgery | 86 |
| Cystatin-C | P01034 | 1D | Glomerulus & proximal tubule | Extracellular Space | other | Acute kidney injury, acute renal dysfunction | 87, 88 |
| Calbindin | P05937 | 1D | Distal tubule & collecting duct | Cytoplasm | other | Distal nephron segment injuries | 89 |
| CYR61 | O60622 | 2D | Proximal tubule | Extracellular Space | other | Glomerular disease | 90 |
| Alkaline phosphatase, tissue-nonspecific isozyme | P09923 | 2D | Proximal tubule | Plasma Membrane | phosphatase | Acute renal failure | 91 |
| Intestinal-type alkaline phosphatase | P05186 | 2D | Proximal tubule | Plasma Membrane | phosphatase | Diabetic nephropathy, acute renal failure | 92 |
| Alpha-N-acetylglycosaminidase | P54802 | 1D | Proximal tubule | Cytoplasm | enzyme | Acute kidney injury | 93 |
| Neprilysin | O95631 | 3D | Proximal tubule | Extracellular Space | growth factor | Acute kidney injury, diabetic nephropathy | 94 |
| Neutrophil gelatinase-associated lipocalin | P80188 | 1D | Proximal tubule & distal tubule | Extracellular Space | transporter | Acute kidney injury, chronic kidney disease | 95 |
| Osteopontin | P10451 | 1D | Proximal tubule & loop of Henle & distal tubule | Extracellular Space | cytokine | Progressive renal injury | 96 |
| Interleukin-18 | Q14116 | 2D | Proximal tubule | Extracellular Space | cytokine | Acute kidney injury | 97 |
| Retinol-binding protein | P02753, P82880, P50120, P09455 | 1D | Proximal tubule | Extracellular Space, Cytoplasm | transporter | Acute kidney injury, renal failure | 98 |
Protein digestion. Urinary proteins were digested with trypsin (Trypsin Gold, mass spec grade, Promega, WI, USA) using filter-aided sample preparation methods. Proteins were loaded onto 10-kDa filter devices (Pall, Port Washington, NY, USA), and 8 M urea in 0.1 M Tris-HCl (pH 8.5) was added to wash the samples. The proteins were denatured by incubation with 50 mM dithiothreitol at 56 °C for 1 h and then alkylated in the dark for 45 min in 55 mM iodoacetamide. Trypsin was added (enzyme to protein ratio of 1:50), and the samples
were incubated at 37 °C overnight. After digestion, the peptide mixtures were desalted on Oasis HLB cartridges (Waters, Milford, USA) and lyophilized for high-performance liquid chromatography separation.

**Offline high-pH RPLC separation.** In total, nineteen samples, including eighteen fractions that were separated by GELFEE and LP-IEF and a pooled urine sample, were fractionated by offline high-pH RPLC columns (4.6 mm × 250 mm, C18, 3 μm; Waters Corp, Milford, USA). The samples were loaded onto the column in buffer A1 (10 mM NH₄FA in H₂O, pH = 10). The elution gradient was 5–30% buffer B1 (10 mM NH₄FA in 90% acetonitrile, pH = 10; flow rate = 1 mL/min) for 60 min. The eluted peptides were collected at one fraction per minute.

After lyophilization, the 60 fractions were re-suspended in 0.1% formic acid and concatenated into 20 fractions by combining fractions 1, 21, 41 and so on51.

**Online LC-MS/MS analysis.** Each sample was analyzed on a reverse-phase C18 self-packed capillary LC column (75 μm × 100 mm, 3 μm). The elution gradient was 5–30% buffer B2 (0.1% formic acid, 99.9% acetonitrile; flow rate = 0.3 μL/min) for 100 min. A TripleTOF 5600 coupled with UPLC system was used to analyze the sample, and the MS data were acquired in a high-sensitivity mode using the following parameters: 30 data-dependent MS/MS scans per full scan; full scans were acquired at a resolution of 40,000 and MS/MS scans were acquired at 20,000; rolling collision energy; charge state screening (including precursors with +2 to +4 charge state);
dynamic exclusion (exclusion duration 15 s); MS/MS scan range of 250–1800 m/z; and scan time of 50 ms. For 1D separation, the pooled urine sample was analyzed with three technical replicates.

**Data processing.** The MS/MS data were processed using Mascot software (version 2.3.02, Matrix Science, London, UK) and searched against the SwissProt database (Homo sapiens, 20,267 sequences, 2013_07 version). The search allowed two missed cleavage sites in the trypsin digestion, cysteine carbamidomethylation was set as a fixed modification and both parent and fragment ion mass tolerances were set to 0.05 Da. Mascot search results were filtered using the decoy database method in Scaffold (version 4.3.2, Proteome Software Inc., Portland, OR). Peptide identifications were accepted if they could be shown to achieve a false discovery rate (FDR) of less than 1.0% by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be shown to achieve a FDR of less than 1.0% and contained at least 2 unique identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Total 20 results from 1DLC, 2DLC and 3DLC (12 GELFReE fractions and 6 LP-IeF fractions) were filtered by Scaffold with the above parameters and yielded average FDRs of 0.10%, 0.26% and 1% at the spectrum, peptide and protein levels, respectively. Then, the 20 datasets were combined together with Scaffold perSPECTives (version 2.0.4, Proteome Software Inc., Portland, OR).

To rank the relative abundance of different proteins, an intensity-based absolute quantification (iBAQ) algorithm was used. The protein intensities were first computed by Progenesis LC–MS (version 2.6, Nonlinear Dynamics, UK) as the sum of all identified peptide intensities (maximum peak intensities of the peptide elution profile, including all peaks in the isotope cluster). The iBAQ result was obtained as the peptide intensities divided by the number of theoretically observable peptides of the protein (calculated in silico protein digestion; all fully tryptic peptides between 6 and 30 amino acids were counted).

For functional analysis, ingenuity pathway analysis (IPA) software (Ingenuity Systems, www.ingenuity.com) was used to analyze cellular components, canonical gene pathways, functions and candidate biomarkers.

**References**

1. Casey, H. W., Ayers, K. N. & Robinson, F. In *Pathology of laboratory animals* 115–173 (Springer, 1978).
2. Gao, Y. Urine-an unappetized goldmine for biomarker discovery! *Science China. Life sciences* 56, 1145–1146, doi:10.1007/s11427-013-4574-1 (2013).
3. Sun, W. et al. Human urine proteome analysis by three separation approaches. *Proteomics* 5, 4994–5001, doi:10.1002/pmic.200401334 (2005).
4. Thuijls, G. et al. Subfractionation, characterization, and in-depth proteomic analysis of glomerular membrane vesicles in human urine. *Nephrology: JASN* 20, 363–379, doi:10.1681/ASN.20080800622 (2009).
5. Davis, M. T. & Doucette, A. A. Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Analytical chemistry* 80, 1568–1573, doi:10.1021/ac702197v (2008).
6. Ahmed, F. E. Sample preparation and fractionation for proteome analysis and cancer biomarker discovery by mass spectrometry. *Journal of separation science* 32, 771–798, doi:10.1002/soic.200800622 (2009).
7. Gonzales, P. A. et al. Large-scale proteomics and phosphoproteomics of urinary exosomes. *Journal of the American Society of Nephrology: JASN* 20, 363–379, doi:10.1681/ASN.200808040406 (2009).
8. Wang, Z., Hill, S., Luther, J. M., Hacht, D. L. & Schey, K. L. Proteomic analysis of urine exosomes by multidimensional protein identification technology (MudPIT). *Proteomics* 12, 329–338, doi:10.1002/pmic.2011100477 (2012).
9. Hogan, M. C. et al. Subfractionation, characterization, and in-depth proteomic analysis of glomerular membrane vesicles in human urine. *Kidney international* 85, 1225–1237, doi:10.1038/ki.2013.422 (2014).
63. Yard, B. A.
62. Johnstone, D. B.
60. Kaplan, J. M.
53. Schwanhausser, B.
58. Mansour, H.
59. Kimura, M.
57. Fukuda, H.
56. Skoberne, A., Konieczny, A. & Schiffer, M. Glomerular epithelial cells in the urine: what has to be done to make them worthwhile?
51. Wang, Y.
50. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis.
47. Rabilloud, T.
45. Tran, J. C.
28. Jia, L.
41. Beck, L. H. Jr.
38. Welsh, G. I. & Saleem, M. A. The podocyte cytoskeleton—key to a functioning glomerulus in health and disease.
37. Suh, J. H. & Miner, J. H. The glomerular basement membrane as a barrier to albumin.
36. Shao, C., Wang, Y. & Gao, Y. Applications of urinary proteomics in biomarker discovery.
35. Kentsis, A.
34. Nolen, B. M.
33. Dyer, A. R.
32. Nagaraj, N. & Mann, M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome.
31. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Molecular & cellular proteomics: MCP 11, M111 014050, doi:10.1074/mcp.M111.014050 (2012).
Nagaraj, N. & Mann, M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. Journal of proteome research 10, 637–645, doi:10.1021/pr100833s (2011).
Dyer, A. R. et al. Evaluation of measures of urinary albumin excretion in epidemiologic studies. American journal of epidemiology 160, 1122–1131, doi:10.1093/aje/kwh3326 (2004).
Nolen, B. M. et al. An extensive targeted proteomic analysis of disease-related protein biomarkers in urine from healthy donors. PloS one 8, e63368, doi:10.1371/journal.pone.0063368 (2013).
Kentisi, A. et al. Discovery and validation of urine markers of acute pediatric appendicitis using high-accuracy mass spectrometry. Annals of emergency medicine 55, 62–70 e64, doi:10.1016/j.annemergmed.2009.04.020 (2010).
Shao, C., Wang, Y. & Gao, Y. Applications of urinary proteomics in biomarker discovery. Science China Life sciences 54, 409–417, doi:10.1007/s11427-011-1462-1 (2011).
Suh, J. H. & Miner, J. H. The glomerular basement membrane as a barrier to albumin. Nature reviews. Nephrology 9, 470–477, doi:10.1038/nrneph.2013.109 (2013).
Welsh, G. I. & Saleem, M. A. The podocyte cytoskeleton—key to a functioning glomerulus in health and disease. Nature reviews. Nephrology 8, 14–21, doi:10.1038/nrneph.2011.151 (2012).
Schwarz, K. et al. Podocin, a raft-associated component of the glomerular slit diaphragm, interacts with CD2AP and nephrin. The journal of clinical investigation 108, 1621–1629, doi:10.1172/JCI21289 (2001).
Vaidya, V. S., Ferguson, M. A. & Bonventre, J. V. Biomarkers of acute kidney injury. Annual review of pharmacology and toxicology 48, 463–493, doi:10.1146/annurev.pharmaco.48.113306.094615 (2008).
Beck, L. H. Jr. et al. M-type phospholipase A2 receptor as target antigen in idiopathic membranous nephropathy. The New England journal of medicine 361, 11–21, doi:10.1056/NEJMoa0781045 (2007).
Wei, C. et al. Circulating urokinase receptor as a cause of focal segmental glomerulosclerosis. Nature medicine 17, 952–960, doi:10.1038/nrn2341 (2011).
Flint, R. S. et al. Probing the urinary proteome of severe acute pancreatitis. HPB: the official journal of the International Hepato Pancreato Biliary Association 9, 447–455, doi:10.1111/j.1365-2273.2007.02744.x (2007).
Ulhen, M. et al. Proteomics. Tissue-based map of the human proteome. Science 347, 1260419, doi:10.1126/science.1260419 (2015).
Tran, J. C. et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature 480, 254–258, doi:10.1038/ nature10375 (2011).
Cargle, B. L., Sevinsky, J. R., Essader, A. S., Eau, J. P. & Stephenson, J. L. Jr. Calculation of the isoelectric point of tryptic peptides in the pH 3.5–4.5 range based on adjacent amino acid effects. Electrophoresis 29, 2768–2778, doi:10.1002/elps.200700701 (2008).
Rabilloot, T. et al. Power and limitations of electrophoretic separations in proteomics strategies. Mass spectrometry reviews 28, 816–845, doi:10.1002/mas.20204 (2009).
Vaezadeh, A. R., Briscoe, A. C., Steen, H. & Lee, R. S. One-step sample concentration, purification, and albumin depletion method for urinary proteomics. Journal of proteome research 9, 6082–6089, doi:10.1021/pr09024as (2010).
Guldbrandsen, A. et al. In-depth characterization of the cerebrospinal fluid (CSF) proteome displayed through the CSF proteome resource (CSF-PR). Molecular & cellular proteomics: MCP 13, 3152–3163, doi:10.1074/mcp.M113.038554 (2014).
Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. Nature methods 6, 359–362, doi:10.1038/nmeth.1322 (2009).
Wang, Y. et al. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteins 111, 2019–2026, doi:10.1002/prot.201000722 (2011).
Nevzhitski, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. Analytical chemistry 75, 4646–4658 (2003).
Schwanhausser, B. et al. Global quantification of mammalian gene expression control. Nature 473, 337–342, doi:10.1038/nature10098 (2011).
Hauck, S. M. et al. Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. Molecular & cellular proteomics: MCP 9, 2292–2305, doi:10.1074/mcp.M1001073 (2010).
Satoskar, A. A. et al. Characterization of glomerular diseases using proteomic analysis of laser capture microdissected glomeruli. Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc 25, 709–721, doi:10.1038/modpathol.2011.265 (2012).
Skoberne, A., Konieczny, A. & Schiffer, M. Glomerular epithelial cells in the urine: what has to be done to make them worthwhile? American journal of physiology. Renal physiology 296, F230–241, doi:10.1152/ajprenal.90507.2008 (2009).
Pukuda, H. et al. Podocin is translocated to cytoplasm in pyruvycin aminonucleoside nephrosis rats and in poor-prognosis patients with IgA nephropathy. Cell and tissue research 360, 391–400, doi:10.1007/s00441-014-2100-9 (2015).
Manour, H. et al. T-cell transcriptome analysis points to a thymic disorder in idiopathic nephrotic syndrome. Kidney international 67, 2168–2177, doi:10.1038/s41598-017-03226-6 (2019).
64. Joosteen, S. A. et al. Antibody response against the glomerular basement membrane protein agrin in patients with transplant glomerulopathy. American journal of transplantation: official journal of the American Society of Transplantation and the American Society of Transplant Surgeons 5, 383–393, doi:10.1111/j.1600-6143.2005.00690.x (2005).

65. Kashtan, C. E. & Kim, Y. Distribution of the alpha 1 and alpha 2 chains of collagen IV and of collagens V and VI in Alport syndrome. Kidney international 42, 115–126 (1992).

66. Kim, Y. et al. Differential expression of basement membrane collagen chains in diabetic nephropathy. The American journal of pathology 138, 413–420 (1991).

67. Kim, Y. et al. Differential expression of basement membrane collagen chains in membranous nephropathy. The American journal of pathology 139, 1381–1388 (1991).

68. Setty, S. et al. Differential expression of laminin isoforms in diabetic nephropathy and other renal diseases. Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc 25, 859–868, doi:10.1038/modpathol.2011.216 (2012).

69. Kim, B. et al. Dysregulated nephron in diabetic nephropathy of type 2 diabetes: a cross-sectional study. PloS one 7, e36041, doi:10.1371/journal.pone.0036041 (2012).

70. Kim, J. M. et al. CD2-associated protein haploinsufficiency is linked to glomerular disease susceptibility. Science 300, 1298–1304 (2003).

71. Ye, H. et al. Urinary podocalyxin positive-element occurs in the early stage of diabetic nephropathy and is correlated with a clinical diagnosis of diabetic nephropathy. Journal of diabetes and its complications 28, 96–100, doi:10.1016/j.jdiatcomp.2013.08.006 (2014).

72. Tufo, A. & Veron, D. VEGF and podocytes in diabetic nephropathy. Seminars in nephrology 32, 383–393, doi:10.1016/j.smnephrol.2012.06.010 (2012).

73. El-Korai, A. F. et al. Cytoskeletal protein expression and regenerative markers in schistosomal nephropathy. Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association 17, 803–812 (2002).

74. Birmingham, D. J. et al. Relationship of circulating Anti-C3b and Anti-C1q IgG to Lupus Nephritis and Its Flare. Clinical journal of the American Society of Nephrology: CJASN 11, 47–53, doi:10.2215/CJN.03990415 (2016).

75. Iida, T. & Ronco, C. Year in review: Critical Care 2004 - nephrology. Critical care 8, 328–334, doi:10.1186/cc3791 (2005).

76. Jim, B. et al. Cofilin1 is involved in hypertension-induced renal damage via the regulation of NF-kappaB in renal tubular epithelial cells. Journal of translational medicine 13, 325, doi:10.1186/s12967-015-0685-8 (2015).

77. Kashtan, C. E. & Kim, Y. Distribution of the alpha 1 and alpha 2 chains of collagen IV and of collagens V and VI in Alport syndrome. Kidney international 42, 115–126 (1992).

78. Kwon, O. & Ahn, K. Netrin-1: a novel universal biomarker of human kidney injury. Experimental nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association 17, 803–812 (2002).

79. Setty, S. et al. Differential expression of laminin isoforms in diabetic nephropathy and other renal diseases. Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc 25, 859–868, doi:10.1038/modpathol.2011.216 (2012).

80. Bonventre, J. V. et al. Biomarkers in Schistosomiasis: a Review of the Literature. PloS one 11, doi:10.1371/journal.pone.0196738 (2016).

81. Kashtan, C. E. & Kim, Y. Distribution of the alpha 1 and alpha 2 chains of collagen IV and of collagens V and VI in Alport syndrome. Kidney international 42, 115–126 (1992).

82. Tufo, A. & Veron, D. VEGF and podocytes in diabetic nephropathy. Seminars in nephrology 32, 383–393, doi:10.1016/j.smnephrol.2012.06.010 (2012).

83. El-Korai, A. F. et al. Cytoskeletal protein expression and regenerative markers in schistosomal nephropathy. Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association 17, 803–812 (2002).

84. Birmingham, D. J. et al. Relationship of circulating Anti-C3b and Anti-C1q IgG to Lupus Nephritis and Its Flare. Clinical journal of the American Society of Nephrology: CJASN 11, 47–53, doi:10.2215/CJN.03990415 (2016).

85. Thrall, J. M. et al. Microalbuminuria in type 1 diabetes is associated with enhanced excretion of the endocytic multiligand receptors megalin and cubilin. Diabetes care 32, 1266–1268, doi:10.2337/dc09-0112 (2009).

86. Schaub, J. A. et al. Perioperative heart-type fatty acid binding protein is associated with acute kidney injury after cardiac surgery. Kidney international 88, 576–583, doi:10.1016/j.kint.2015.10.014 (2015).

87. Wu, I. & Parikh, C. R. Screening for kidney diseases: older measures versus novel biomarkers. Clinical journal of the American Society of Nephrology: CJASN 3, 1895–1901, doi:10.2215/CJN.02303008 (2008).

88. Rusci, Z. & Ronco, C. Year in review: Critical Care 2004 - nephrology. Critical care 9, 523–527, doi:10.1186/cc3791 (2005).

89. Iida, T. et al. Decreased urinary calbindin 1 levels in proteinuric rats and humans with distal nephron segment injuries. Clinical and experimental nephrology 18, 432–443, doi:10.1007/s10157-013-0835-3 (2014).

90. Sawai, K. et al. Expression of CC1N1 (CryRS1) in developing, normal, and diseased human kidney. American journal of physiology. Renal physiology 293, F1363–1372, doi:10.1152/ajprenal.00205.2007 (2007).

91. Lameire, N. Van, Biesen, W. & Vanholder, R. Acute renal failure. Lancet 365, 417–430, doi:10.1016/S0140-6736(05)17831-3 (2005).

92. Gubin, A. D. et al. Human kidney injury: a meta-analytic approach. British journal of medicine & medical research 42, 1058–1067, doi:10.1053/j.kid.2013.05.014 (2013).

93. Varghese, S. A. et al. Identification of diagnostic urinary biomarkers for acute kidney injury. Journal of investigative medicine: the official publication of the American Federation for Clinical Research 58, 612–620, doi:10.231/jim.0013e3181473c7e (2010).

Acknowledgements
This work was supported by the National Basic Research Program of China (No. 2013CB530805, No. 2014CB902005), National Key Research and Development Program of China (No. 2016 YFC 1306300), the National Natural Science Foundation of China (No. 31202064, No. 31400669), the Beijing Natural Science Foundation (No. 5132028, 7173264, 7172076), the Fundamental Research Funds for the Central Universities (2015KJJCB21), Beijing Normal University (No. 11100704) and Biologic Medicine Information Center of China, National Scientific Data Sharing Platform for Population and Health.
Author Contributions
M.Z., W.S. and Y.G. prepared the first draft. M.L. (Menglin Li) and W.S. conceived and designed the experiments. M.L. (Menglin Li), and Z.G. performed the experiments. M.Z., Y.Y. and C.S. analyzed the data. Y.S. and M.L. (Mingxi Li) contributed to collect clinical samples. All authors approved the final manuscript.

Additional Information
Supplementary information accompanies this paper at doi:10.1038/s41598-017-03226-6

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017