

Urinary Biomarkers for Detection of Early and Advanced Chronic Kidney Disease - A Pilot Study

Vlasta Korenkova^{1,2}, Alun Jones³, Wendy E Hoy⁴, Christudas Morais², Matthew A Cooper³ and Glenda C Gobe^{2*}

¹Institute of Biotechnology, Academy of Sciences, Prague, Czech Republic

²Centre for Kidney Disease Research, Translational Research Institute and School of Medicine, University of Queensland, Brisbane, Australia

³Institute for Molecular Biosciences, University of Queensland, Brisbane, Australia

⁴Centre for Chronic Disease, University of Queensland, Brisbane, Australia

Abstract

Chronic kidney disease (CKD) is a significant and costly public health problem with increasing prevalence in most societies and the need for improved diagnosis. The aim of this study was to demonstrate how it is possible to decipher a combination of markers from a compendium of differentially-expressed proteins using available statistical tools when more than two study groups are involved. Screening for potential urinary biomarker profile stratifying early and later stages of CKD were performed. Sixteen patients with defined CKD staging were selected and compared with ten healthy individuals. Urinary proteins were quantified using the iTRAQ method and analysed with ProteinPilot, GenEx and Meta Core software. Four proteins (out of 194) (apolipoprotein D; protein AMBP; zinc-alpha-2-glycoprotein; and kininogen 1) were identified as critical for CKD stage separation. In conclusion, this preliminary work provides evidence that several unique urinary proteins are involved in early and later stages of CKD and suggests that a selected combination of biomarkers could be used to profile patients into different CKD stages. Further validation studies are now needed.

Keywords: Alpha-1-microglobulin; Apolipoprotein D; AZGP1; iTRAQ; KNG1; Proteome; Urine

Abbreviations: ACP: Prostatic acid phosphatase; ALB: Serum albumin; AMBP: Alpha-1-microglobulin; APOD: Apolipoprotein D; AZGP1: Zinc-alpha-2-glycoprotein; CD59: CD59 glycoprotein; CE-MS: Capillary electrophoresis-coupled mass spectrometry analysis; CKD: Chronic kidney disease; E9KL23: Epididymis secretory sperm binding protein Li 44a; ESKD: End stage kidney disease; estimated GFR: eGFR; FDR: False discovery rate; GFR: Glomerular filtration rate; GO: Gene ontology; HP: Haptoglobin; HSPG2: Basement membrane-specific heparansulfate proteoglycan core protein; iTRAQ: Isobaric tag for relative and absolute quantitation; KNG1: Kininogen 1 isoform CRA_a; LC-MS/MS: Liquid chromatography-tandem mass spectrometry; ORM1: Gene that encodes human alpha-1-acid glycoprotein 1; PCA: Principal component analysis; "PUP": Putative uncharacterized protein; SOM: Self organizing map; UA: Urea; UMOD: Uromodulin

Introduction

Chronic kidney disease (CKD) is a complex continuum of structural and functional changes that progress from kidney health to terminal kidney failure. CKD is stratified into 5 stages based on the presence of kidney damage and/or decreased kidney function, typically measured by decreased glomerular filtration rate (GFR) for 3 months or more [1]. Stages 1 and 2 define conditions of kidney damage in the presence of a GFR of at least 90 ml/min/1.73 m² or 60 to 89 ml/min/1.73 m², respectively, and stages 3A, 3B, 4 and 5 define conditions of moderately and severely reduced GFR irrespective of markers of kidney damage. Early stages of the disease could be reversible, but they are often asymptomatic and usually only detected during the assessment of comorbid disorders, for example, diabetes mellitus [2]. The development of CKD usually occurs over decades but in some patients can be fast. In this case, end stage kidney disease (ESKD), needing dialysis or kidney transplantation for patient survival, could develop within several months [3]. The prevalence of CKD continues to increase and is estimated to be 8-16% of populations worldwide [4,5]. Early identification of patients with kidney disease is needed, as measures may be instituted to slow progression of the disease and improve quality of life.

The diagnosis of CKD is typically made on the basis of increased serum creatinine or urinary protein. However, the use of serum creatinine as an indirect GFR and CKD marker is limited. There is considerable individual variability and poor assay specificity affecting creatinine measurement. Confounding contributors are co-morbid non-renal illnesses, medications and nutrition. The estimation of GFR by serum creatinine also differs between healthy people and patients with CKD because of differences in the estimation of GFR range and creatinine production, with a risk of overestimating GFR in the CKD population [6]. Thus proteinuria, or more specifically albuminuria, is used as the earliest marker of kidney damage in patients with diabetes, hypertension and glomerular diseases [7]. However, the importance of albuminuria has been questioned because there have been several reports describing the development of CKD in patients with diabetes despite the absence of albuminuria [8]. It is clear that new and reliable biomarkers or biomarker-based models are needed for the clinical management of CKD.

Use of urine possesses several advantages for new biomarker discovery. The first is that urine may be obtained non-invasively, and in larger volumes compared with plasma. The larger volumes of urine compensate for the lower concentration of proteins and peptides in urine compared with plasma [9]. Secondly, the urinary proteome is a filtrate of the serum proteins, but also contains proteins arising from the shed cells of the kidney and urogenital tract, from tubular secretion and secreted exosomes [10]. Various techniques have been

***Corresponding author:** Glenda C Gobe, Centre for Kidney Disease Research, UQ School of Medicine, Translational Research Institute, 37 Kent Street, Woolloongabba, Brisbane, Australia, Tel: +61 734 438 011; E-mail: g.gobe@uq.edu.au

Received January 21, 2015; **Accepted** March 10, 2015; **Published** March 12, 2015

Citation: Korenkova V, Jones A, Hoy WE, Morais C, Cooper MA, et al. (2015) Urinary Biomarkers for Detection of Early and Advanced Chronic Kidney Disease - A Pilot Study. Med chem 5: 096-103. doi:10.4172/2161-0444.1000249

Copyright: © 2015 Korenkova V, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

used so far in order to reflect a healthy renal physiology [11] that may be distinguished from pathophysiological state caused by defined disease, using differentially expressed proteins or peptides [12]. In most of the large-scaled studies, capillary electrophoresis-coupled mass spectrometry analysis (CE-MS) has been employed for the identification of new biomarkers for CKD [11].

In the present study, a state-of-the-art hybrid quadrupole time-off flight (TOF) TripleTOF 5600 MS system (AB SCIEX, Mt Waverley, Australia) was employed. This instrument is able to analyse complex samples in great depth, and allow the whole proteome to be explored without the need of enriching low abundance proteins or removing high abundance proteins [13]. Our aim was to screen the differentially-expressed proteins in different defined stages of CKD and identify potential new diagnostic biomarkers for improving early classification of CKD. Using isobaric tags for relative and absolute quantitation (iTRAQ) labelling coupled with LC-MS/MS quantitative analysis, 8 proteins were identified that could discriminate healthy controls from early and later stages of CKD.

Materials and Methods

Sample collection

This work had ethics approval from the Princess Alexandra Hospital Human Research Ethics Committee and the University of Queensland. All samples were obtained with informed consent. Urine samples from CKD patients were collected at the Princess Alexandra Hospital (Brisbane, Australia) between 2011 and 2013. Urine specimens for proteomic analysis were obtained during the course of regular clinical management. Control samples were collected at the Translational Research Institute (Brisbane, Australia) from healthy volunteers, the effect of age was not considered in this analysis [14]. All samples were mid-stream, second or third morning urine. Urine specimens were stored at 4°C before the transport to laboratories at the Translational Research Institute. Immediately upon arrival, urine was centrifuged at 4,000 RPM at 4°C for 10 minutes to remove debris material and cells, aliquoted and stored at -80°C until further processing.

Sample preparation

Urine creatinine and total protein concentration were measured for each sample in the Chemical Pathology Department at the Princess Alexandra Hospital (Brisbane, Australia) by the Jaffe method using a Beckman DxC800 general chemistry analyser (Beckman Coulter, USA). The ratio of creatinine: total protein was used to determine the volume of urine needed for each individual patient sample to contribute to the pool of each group of samples (cohorts). There were 7 cohorts based on CKD staging: healthy control males, CKD 3A males, CKD 3B males, CKD 4 males, healthy control females, CKD 3A females, CKD 3B + CKD 4 females. At least 3 samples, maximally 5 samples, were pooled for each cohort. Males and females were treated independently. All measured characteristics and clinical parameters for individual samples are described in Supplementary Material 1.

Urine was desalted and protein concentrated using 15 ml 10,000 NMVL Amicon filters (Merck Millipore Pty Ltd, Kilsyth, Australia) according to manufacturer's instructions. Ethanol protein precipitation was performed routinely and the pellet left to dry at room temperature before it was solubilized in 8M urea in 0.1M Tris/HCl pH 8.5 (UA) and the concentration measured using a Quick Start Bradford Protein Assay (Bio-Rad Laboratories Pty Ltd, Gladesville, Australia). UA was used as diluent and gamma-globulin as a standard. The solutions of all cohorts were normalized for protein concentration of 1 mg and applied on the 0.5 ml 10,000 cut-off Microcon Centrifugal device

filters (Merck Millipore Pty Ltd) to implement filter-aided sample preparation (FASP) at room temperature [15]. Protein solutions were reduced by dithiothreitol in UA (DTT reduction solution), centrifuged, then washed with UA and centrifuged again. Protein solutions were alkylated in the dark with iodoacetamide in UA for 30 min. 100 µl DTT reduction solution were added to remove the excess alkylation reagent, followed by centrifugation and washing in UA, then washing in 50 mM ammonium bicarbonate pH 8.0 (ABC) and centrifugation. Following routine methods, peptides were collected, lyophilised and resuspended in 40 µl of 1% formic acid in HPLC graded water.

Label free proteomics

The urine protein tryptic digest samples were analyzed by NanoLC-MS/MS on a Shimadzu Prominence Nano HPLC (Japan) coupled to a Triple TOF 5600 mass spectrometer (AB SCIEX, Canada) equipped with a Nano electrospray ion source. 6 µl of each extract was injected onto a 50 mm x 300 µm C18 trap column (Agilent Technologies, Melbourne, Australia) at 30 µl/min. The samples were de-salted on the trap column for 5 min using 0.1 % formic acid at 30 µl/min. The trap column was then placed in-line with the analytical nano HPLC column, a 150 mm x 75 µm 300SBC18, 3.5 µm (Agilent Technologies, Australia) for mass spectrometry analysis. Linear gradients of 2-40% solvent B over 165 min at 300 nl/min flow rate, followed by a steeper gradient from 40% to 80% solvent B in 15 min were used for peptide elution. Solvent B was held at 80% for 5 min for washing the column and returned to 2% solvent B for equilibration prior to the next sample injection. Solvent A consisted of 0.1% formic acid and solvent B contained 90/10 acetonitrile/0.1 % formic acid. The ion spray voltage was set to 2400 V, declustering potential (DP) 100 V, curtain gas flow 25, nebuliser gas 1 (GS1) 12 and interface heater at 150 °C. The mass spectrometer acquired 100 ms full scan TOF-MS data followed by 20 and by 50 ms full scan product ion data in an information dependant acquisition (IDA) mode. Full scan TOF MS data were acquired over the mass range 300-1400 and for product ion MS/MS 80-1400. Ions observed in the TOF-MS scan exceeding a threshold of 150 counts and a charge state of +2 to +5 were set to trigger the acquisition of product ion, MS/MS spectra of the resultant 20 most intense ions.

iTRAQ labelling and quantitative proteomics

Tryptic peptides (10 µg) were amine labelled with iTRAQ reagents 113, 114, 115, 116, 117, 118, 119 according to the manufacturer's instruction (AB SCIEX). After the labelling reaction, the seven samples were pooled and salts plus excess iTRAQ reagents removed using SCX chromatography according to the manufacturer's instruction (AB SCIEX). The SCX peptide fraction was vacuum centrifuged to dryness. The samples were reconstituted in 30 µl of 1% formic acid (aq) for subsequent Nano LC-MS/MS analysis. The chromatographic plus mass spectrometric conditions were as for the label-free experiment except that the MS/MS product ion experiment scans were enhanced between m/z 112 and m/z 125 to obtain improved sensitivity of the iTRAQ label quantitative ions.

Database search and iTRAQ quantification

Proteins were identified by database searching using Protein Pilot v4.5 (AB SCIEX) against the UniProt_Sprot_20130205 database (Approx 106,000 entries searched, false discovery rate/FDR of 1%). Search parameters were defined as a thorough search using trypsin digestion enzyme iodoacetamide plus other cysteine alkylation and all entries in the database. Proteins were considered positive if there were 2 or above peptides identified with a 95% confidence and a 1% global FDR. Proteins identified by the unlabelled approach, with their characteristic functions, are listed in Supplementary Material 2, and proteins identified

by iTRAQ labelling with their characteristic functions are listed in Supplementary Material 3. Tryptic peptides (10 µg) were amine labelled with iTRAQ reagents 113, 114, 115, 116, 117, 118, 119.

Analysis of proteomic data

The basic statistics, as the expression ratios and uncorrected p value calculated for contributing peptide ratios within a single protein, were calculated using ProteinPilot v4.5 (AB SCIEX). The data from three iTRAQ measurements were checked for outliers (removed) and averaged. Principal component analysis (PCA), Kohonen self-organising maps (SOM) and hierarchical cluster analysis using the unweighted pair method and the Euclidean distance were calculated with autoscaled data using GenEx Enterprise software (Multi D Analyses AB, Sweden). The SOM of size 3 × 1 and 5 × 1, dividing the samples into 3 and 5 groups, respectively, were trained using GenEx with the following parameters: 0.1 learning rate, 3 neighbours and 5,000 iterations. The SOM analysis was repeated eight times with identical classification. The 6 proteins with differential expression and known protein codes, which were selected as biomarker candidates in CKD were subjected to a GO Database search and further analyses using Meta Core software (Gene Go, Inc., USA). The protein codes were not found for undescribed putative protein (Q8NEJ1) and E9KL23 that is why they were not included in the analysis. GO enrichment analysis was performed to identify GO terms and already associated diseases in differentially expressed proteins.

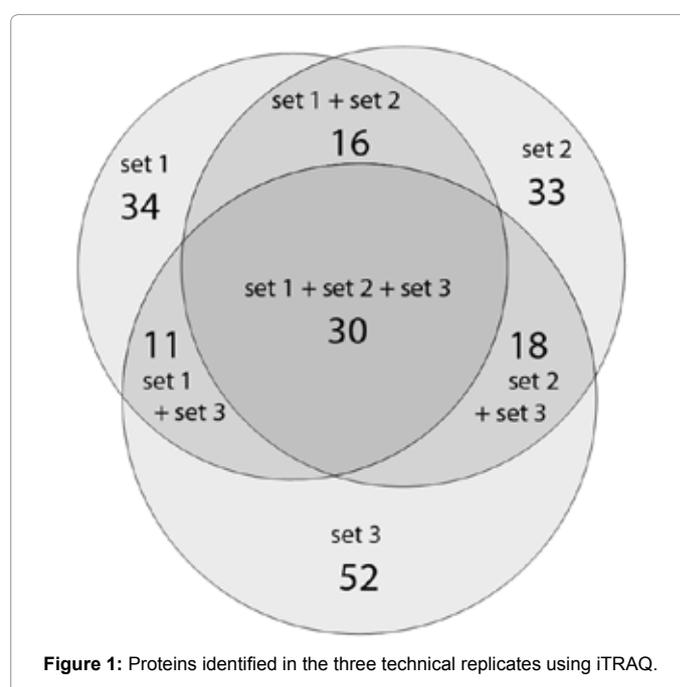
Results

Due to the pilot nature of this study and a need for lowering biological variation of individual samples and screening for prominent changes, pooled urine samples were used. As has been reported previously, significant differences in the urinary protein profiles can be found across individuals of the same group and the variation of the urinary protein across and within individuals seems to be inevitable [16].

Proteins were firstly identified as described in methods by an unlabelled approach in control groups and CKD 3A groups. The number of the proteins shared between healthy controls and CKD 3A stages was similar in both male and female groups measured: 104 proteins in male group, 113 proteins in female group. However, when shared proteins were compared, 41 proteins out of 104 proteins (39.4%), were unique for male CKD 3A and 50 proteins out of 113 proteins (44.3%)

were unique for female CKD 3A group only. The high percentage of proteins that are gender unique justifies our decision to separate male and female groups for biomarker identification in order to describe the biomarkers that would not be specific for one gender in pooled samples only.

Using the iTRAQ-labelled approach, three technical replicates were used for complementation of proteins. From all analysed technical replicates, 194 proteins were identified. 91 unique proteins (656 distinct peptides), 97 unique proteins (854 distinct peptides) and 111 unique proteins (1462 distinct peptides) were identified in set 1, set 2 and set 3, respectively. 30 proteins were found in all three sets. For biomarker identification, we selected proteins that were present at least twice in 3 technical replicate datasets: set 1+2+3, set 1+2, set 2+3 and set 1+3 (Figure 1), and which were differentially expressed with fold change ratio ≥ 1.5 or ≤ 0.7 and p value <0.05. Uncorrected p value is calculated for contributing peptide ratios within a single protein. Table 1 lists 13



UP Ident.	Protein code	Av Pep	Males 3A	Males 3B	Males 4	Females 3A	Females 3B+4	UnLab	BC
P02763	ORM1	25	6.0* ± 0.5	1.9* ± 0.1	3.6* ± 0.2	8.3* ± 0.1	6.9* ± 0.5	Yes	No
C9JF17	APOD	12	1.4 ± 0.2	0.2* ± 0.1	0.6 ± 0.1	1.2 ± 0.2	0.4* ± 0.1	No	Yes
P98160	HSPG2	14	1.3 ± 0.2	0.5* ± 0.1	0.8 ± 0.1	0.9 ± 0.1	0.5 ± 0.1	No	Yes
Q6FHM9	CD59	12	0.7 ± 0.1	0.3* ± 0.1	0.6* ± 0.1	1.2 ± 0.1	0.7 ± 0.1	No	Yes
E9KL23		12	1.4 ± 0.0	2.2* ± 0.3	0.5* ± 0.1	4.2* ± 0.4	0.7 ± 0.0	Yes	Yes
P00738	HP	5	3.3* ± 0.1	1.0 ± 0.3	2.3* ± 0.2	8.7 ± 0.0	2.3* ± 0.7	No	No
D3DNU8	KNG1	19	0.4* ± 0.1	0.2* ± 0.1	0.2* ± 0.1	0.6 ± 0.1	0.2* ± 0.0	Yes	Yes
P15309	ACPP	6	0.2* ± 0.1	0.2* ± 0.1	0.8 ± 0.0	0.9 ± 0.2	0.5 ± 0.2	No	No
P02760	AMB	74	2.0* ± 0.1	0.4* ± 0.1	2.3* ± 0.0	2.1* ± 0.3	1.2 ± 0.1	Yes	Yes
Q8TCD0	"PUP"	7	1.6 ± 0.4	0.6 ± 0.0	1.7 ± 0.5	1.9* ± 0.2	1.5* ± 0.3	Yes	Yes
P02768	ALB	302	3.1* ± 0.2	4.6* ± 0.0	2.3* ± 0.2	4.2* ± 0.2	2.8* ± 0.3	Yes	N/A
E9PEA4	UMOD	39	2.2* ± 0.1	0.2* ± 0.1	0.2* ± 0.1	0.7 ± 0.0	0.1* ± 0.0	No	No
P25311	AZGP1	21	3.7* ± 0.5	0.6* ± 0.1	2.7* ± 0.2	3.2* ± 0.4	2.4* ± 0.4	Yes	Yes

UP Ident. = UniProt identifier; Av Pep = average number of peptides; Males 3A = CKD stage 3A ratio CKD males:control males; similarly, for Males 3B and Males 4; Females 3A = CKD stage 3A ratio CKD females:control females; similarly, for Females 3B+4; UnLab = identified by unlabelled approach in all tested groups; BC = biomarker candidate.

* = p<0.05

Table 1: Significantly differentially regulated proteins identified in all CKD subjects by iTRAQ labelling.

proteins that were significantly differentially expressed at least twice in the same tested CKD stage when compared to controls. Note that the CKD 3B male group demonstrated dissimilar behaviour when compared to other groups.

To identify diagnostic biomarker candidates that would stratify healthy individuals from early (CKD 3A) and later or progressive stages of CKD (CKD 3B or CKD 4), different bioinformatics methods were applied. The multivariate data analysis method principal component analysis (PCA) was used to reduce the dimensionality of a data set, which consisted of 13 variables (differentially expressed proteins) and 5 different CKD groups in ratios with controls: male CKD 3A, CKD 3B and CKD 4 reported as ratios with the male controls; and female

CKD 3A and CKD 3B4 reported as ratios with the female controls. All expression values (peptide ratios) were auto-scaled to remove the influence of both the expression level and the magnitudes of the changes and gave rise to classification based on the relative changes in expression. When all 13 differentially-expressed proteins were used for PCA analysis, the separation of the early and late CKD stages was not satisfactory (Figure 2A). Thus, we further searched for the proteins responsible for separation of CKD groups by inspecting their loadings and by performing dynamic PCA in GenEx (MultiD). Figure 2B shows the best separation of groups with 8 proteins in 2-dimensional PCA: apolipoprotein D (APOD); basement membrane-specific heparansulfate proteoglycan core protein (HSPG2); CD59 antigen, complement regulatory protein, isoform CRA_b (CD59); epididymis

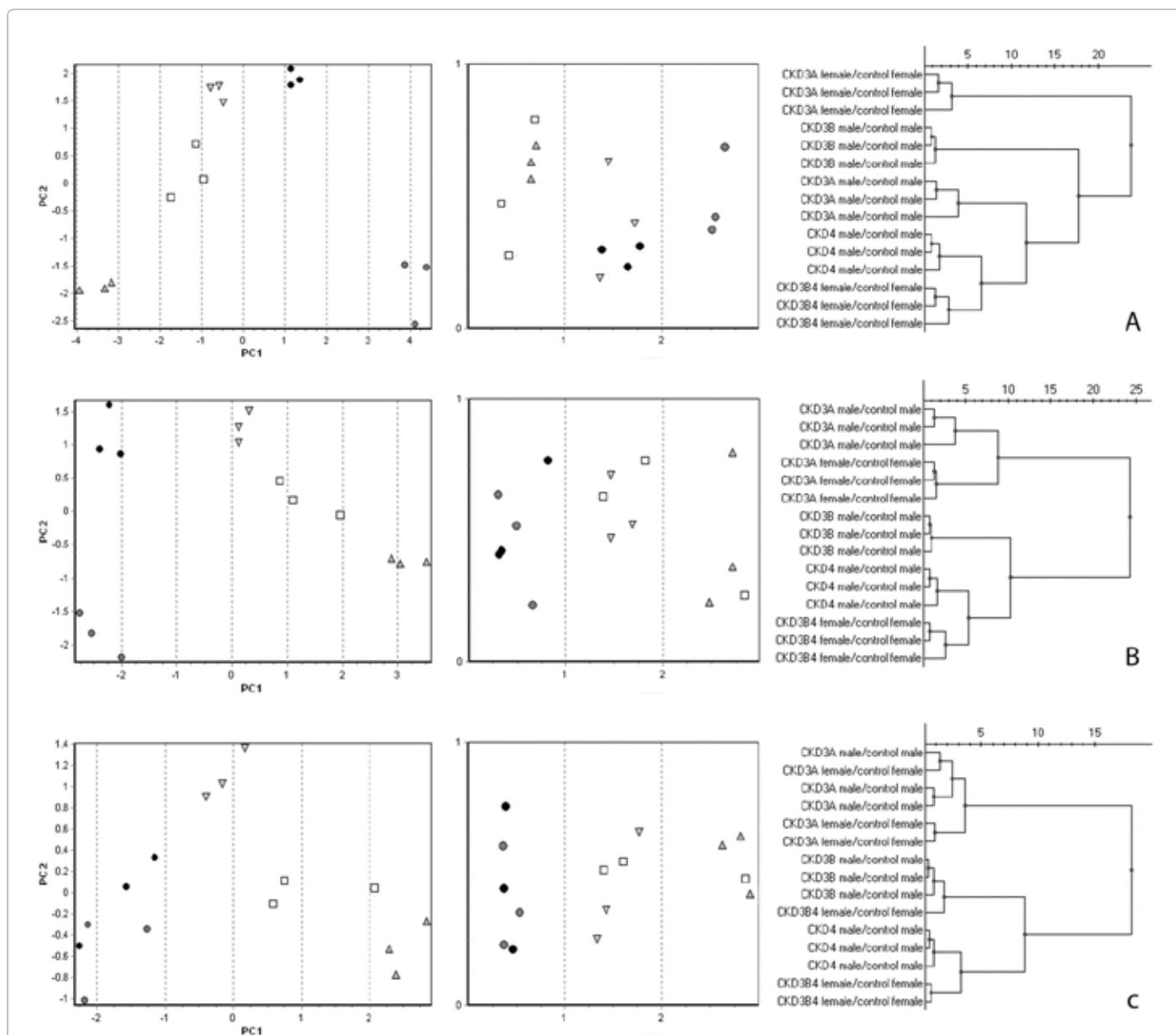


Figure 2: Multivariate analysis. Grey circle: CKD 3A female ratios, black circle: CKD 3A male ratios, grey reversed triangle: CKD 4 male ratios, white square: CKD 3B4 female ratios, grey triangle: CKD 3B male ratios. All data are auto-scaled. 2A: Principal component analysis (PCA) based on all 13 differentially expressed proteins followed by Kohonen's self-organising map (SOM) and hierarchical cluster analysis (HCA) using the unweighted pair method and the Euclidean distance. 2B: PCA, SOM and HCA based on expression of 8 differentially expressed proteins. 2C: PCA, SOM and HCA based on expression of AMBP, APOD, AZGP1 and KNG1.

secretory sperm binding protein Li 44a (E9KL23); kininogen 1; isoform CRA_a (KNG1); protein AMBP (AMBP); putative uncharacterized protein (Q8NEJ1); and zinc-alpha-2-glycoprotein (AZGP1). The first two principal components explained 80.5% variance in the data. The results from PCA were validated with two other methods using auto-scaled data. The first was hierarchical cluster analysis using the unweighted pair method and the Euclidean distance. The dendrogram had the same main features as PCA (Figure 2B). Two main groups were identified; one formed from both male and female CKD3A stages, the second one consisting of the later male and female CKD stages 3B – 4 in two sub clusters. The second method that helped to confirm the outcome from the PCA was Kohonen's self-organizing feature maps (SOM), which are based on a branch of mathematical techniques that do not require formal equations, but use rules to organise the data through a series of random events. The samples were divided with SOM into 3 groups or 5 groups based on their expression profiles (Figure 2B). One group consisted of male and female early stages CKD 3A, and the other two groups contained the male and female later stages. Dividing the samples into 5 groups with SOM revealed spatial separation of all 5 groups. SOM classification was fully reproducible – repeated

independent classification predicted the same groups. Further analysis, using the same methods: 2-D PCA analysis (explaining 91.9% of variance), unsupervised clustering and SOM; revealed 4 final proteins that were critical for the described cluster separation, these being AMBP, APOD, AZGP1 and KNG1 (Figure 2C).

Protein codes were not available for E9KL23 and putative protein Q8NEJ1. Blast analysis of the obtained sequence of putative protein Q8NEJ1 ranked it among proteins from immunoglobulin superfamily (Supplementary Material 4). The remaining six biomarker candidates were analysed using Meta Core software (Gene GO, Inc., Carlsbad, USA) to determine their molecular function, participation in biological processes and their known biomarker association to diseases. The results of the search are listed in Table 2. We also performed pathway analysis on 6 dysregulated proteins. GeneGO analysis demonstrated that the selected biomarker candidates are involved in different pathways. KNG1 is involved in protein folding and maturation of bradykinin/kallidin and in blood coagulation, AMBP participates in cell adhesion of cell matrix glycoconjugates, CD59 is an inhibitor of complement system and plays role in immune response alternative

Name of protein	Biological GO process	Molecular GO function	Biomarkers in illnesses by GeneGO
Apo-lipoprotein D	Aging, angiogenesis, brain development, glucose metabolic process, lipid metabolic process, lipid transport, negative regulation: of T cell migration, of cytokine production involved in inflammatory response, of focal adhesion assembly, of lipoprotein lipid oxidation, of monocyte chemotactic protein-1 production, of platelet-derived growth factor receptor signaling pathway, of protein import into nucleus, of smooth muscle cell proliferation, of smooth muscle cell-matrix adhesion; peripheral nervous system axon regeneration, response to axon injury, response to drug, response to reactive oxygen species, tissue regeneration, transport	Small molecules binding, lipid transporter, lipid binding, cholesterol binding	Alzheimer disease, Cardiovascular diseases, Dermatofibrosarcoma, Diabetes mellitus type 2, Encephalitis, Endometrial stromal tumors, Fibroma, Heart failure, Inflammation, Parkinson disease, Schizophrenia, Breast, endometrial, lung, mouth, ovarian, pancreatic, prostatic, rectal and skin neoplasms
Basement membrane-specific heparansulfate proteoglycan core protein	Angiogenesis, brain development, carbohydrate metabolic process, cardiac muscle tissue development, cartilage development involved in endochondral bone morphogenesis, chondrocyte differentiation, chondroitin sulfate metabolic process, embryonic skeletal system morphogenesis, endochondral ossification, extracellular matrix organization, glycosaminoglycan metabolic process, lipoprotein metabolic process, phototransduction, protein localization, retinoid metabolic process, small molecule metabolic process	Collagen V binding, protein C-terminus binding	Alzheimer disease, Ameloblastoma, Arteriosclerosis, Asthma, Hepatocellular carcinoma, Non-Small-Cell Lung, Diabetes mellitus type 1, Gingival overgrowth, Glioblastoma, HIV infections, Chronic B-Cell leukemia, Melanoma, Mitral valve prolapse, Precursor T-Cell lymphoblastic leukemia-lymphoma, Schwartz-Jampel syndrome, Lymphocytic Duchenne muscular dystrophies, Inclusion body myositis, Odontogenic tumors, Periodontitis, Polymyositis, Breast, colorectal, endometrial, lung, kidney, mouth, oropharyngeal, ovarian, prostatic, rectal, skin, urinary bladder and uterine cervical neoplasms
CD59 glycoprotein	Blood coagulation, cell surface receptor signaling pathway, innate immune response, negative regulation: of activation of membrane attack complex, of apoptotic process, of complement activation, of cytolysis; positive regulation of T cell proliferation, regulation of complement activation	Complement binding	Anemia, Anoxia, Arteriosclerosis, Ductal carcinoma, Cerebral infarction, Hereditary nonpolyposis, Diabetes mellitus type 2, HIV infections, Hemoglobinuria, Lupus erythematosus, Paroxysmal, Infarction, Sjogren's Syndrome, Teratocarcinoma, Thrombosis, Vascular diseases, wounds and injuries, Colorectal, endometrial, esophageal, lung, systemic, pancreatic, prostatic and stomach neoplasms
Kininogen 1, isoform CRA_a	Elevation of cytosolic calcium ion concentration, inflammatory response, negative regulation: of blood coagulation, of cell adhesion, of peptidase activity; platelet degranulation, positive regulation: of apoptotic process, of renal sodium excretion, of urine volume; regulation of blood vessel size, smooth muscle contraction, vasodilation	Cysteine-type endopeptidase inhibitor, protein transmembrane transporter, ribonuclease, antigen binding	Arthritis, Rheumatoid, Carcinoma, Hepatocellular, Diabetes mellitus type 2, Diabetic nephropathies, Hypertension, Hypertrophy, Inflammation, Sepsis, Neoplasm metastasis, Lung, prostatic and skin neoplasms
Protein AMBP	Cell adhesion, female pregnancy, heme catabolic process, modulation by virus of host morphology or physiology. Negative regulation of JNK cascade, of immune response, of peptidase activity. Protein catabolic process, protein maturation, protein-chromophore linkage	Calcium channel inhibitor, serine-type endopeptidase inhibitor, heme binding, transmembrane transporter, ribonuclease	Acute kidney injury, Hepatocellular carcinoma, Heart arrest, Hepatitis, Kidney tubular necrosis, Leukemia, Myeloid, Non-Hodgkin lymphoma, Renal insufficiency, Multiple myeloma, Myelodysplastic Syndromes, Neoplasm metastasis, Endometrial, lung and ovarian neoplasms
Zinc-alpha-2-glycoprotein	Antigen processing and presentation, cell adhesion, immune response. Negative regulation, proliferation. Positive regulation, T cell mediated cytotoxicity, and of multicellular organism growth	Antigen binding, protein transmembrane transporter, ribonuclease	Adenocarcinoma, Asthma, Glioblastoma, Sinusitis, Obesity, Breast, Endometrial, ovarian, prostatic, rectal and stomach neoplasms

Table 2: Gene ontology (GO) and GeneGO annotation of differentially-expressed proteins.

complement pathway, and HSPG2 is involved in the Fc gamma receptor signalling pathway.

Discussion

A “proteomic biomarker” can be described as a specific peptide or protein that reflects normal or abnormal physiologic processes of a condition or disease. Similarly, a “biomarker profile” is defined as a combination of distinct proteomic biomarkers, using a clearly defined algorithm giving a readout parameter associated with the specific condition [17]. A combination of biomarkers is preferred, because a single biomarker of a disease would be always hampered with the problem of high variability, especially in urine, where biological reproducibility of the results within the individual during the day drops to 70% [12]. It is very unlikely that an “ideal biomarker”, which should display unique characteristics (for example, presence/absence or signal intensity) in each sample from patients with a certain disease, does exist [18]. However, a combination of biomarkers could lead to generation of distinct diagnostic patterns. The biomarker pattern method would be more robust and specific [19]. The new, fast evolving high-throughput proteomic techniques enable us to screen the whole proteome of the sample and enable us to create such proteomic profiles.

In our pilot experiment using iTRAQ labelling of pooled samples from CKD patients in different stages, we employed the later approach to analyse differentially expressed proteins. We divided our tested groups not only to different CKD stages but also according to a gender, as it was described previously that gender can influence the protein profiles [20]. Multivariate data analysis methods PCA and SOM proved that our 8 final protein candidates or 4 critical final candidates (out of 13 significantly differentially expressed proteins) were able to distinguish well-separated groups based on their CKD stage. In addition, hierarchical cluster analysis resulted in two main distinct groups. One group brought together early stage CKD patients: both female and male CKD 3A groups that were normalized with appropriate gender controls. The second cluster contained the later CKD stages, which were possible to distinguish further. Of note, the expression of the CKD 3B group was very different when compared to the CKD 3A group, using data available for male groups (Table 2). This confirmed the premise for distinguishing CKD stage 3 into 2 separated groups 3A and 3B, as described earlier [21].

Most of the biomarker candidates identified by us are present in high abundance in urine and had been reported previously either in direct connection with CKD, for example AZGP1, KNG1, APOD and E9KL23 were described in CKD 3A stage [22], or were linked to CKD-associated diseases such as diabetes mellitus (AZGP1, KNG1, AMBP, HSPG2 and CD59) [23] or hypertension (KNG1) [24]. We also identified one differentially-expressed undescribed putative protein (Q8NEJ1) which appears to improve discrimination between early and later CKD stages. A BLAST search in human database (NCBI/protein-protein BLAST) revealed this protein to have 82% homology for protein immunoglobulin lambda-like polypeptide 5 isoform 1 (NP 001171597.1).

In more detail, AZGP1 glycoprotein, one of the crucial players and an adipocytokine, is involved in immune response. This differentially-regulated protein was recently associated not only with kidney malfunction, for example normo-albuminuric diabetic nephropathy (reported differentially-abundant in non-albuminuric patients) [25], but also with other diseases such as diabetes mellitus type 1 (significantly up-regulated in diabetic patients) [23], diabetes mellitus type 2 (increased levels of AZGP1 in diabetic patients) [26], urinary bladder cancer (increased levels in cancer patients) [27], prostate

cancer [28] and others, suggesting that this protein cannot be used as a single biomarker for a single disease but it might be useful in combination with other biomarkers, as suggested above, in a specific diagnostic pattern. The same lack of specificity for disease biomarkers can be found for all other biomarker candidates, such as KNG1, AMBP, APOD, HSPG2 and CD59 (Table 2).

KNG1, another crucial glycoprotein, is a protein involved in the renal kallikrein-kinin system [29]. Its reduction was described in ESKD and end-stage liver disease [30]. Very low levels of KNG1 were also observed in the urine of IgA nephropathy patients [31]. Of note, KNG1 was up-regulated in a study by Vivekanandan-Giri [22] in urine of CKD 3A stage patients when compared with healthy controls in a small dataset (6 controls and 6 CKD 3A patients). This is in direct contrast with our findings, where we observed lowered levels of KNG1 in all tested CKD stages when compared to controls. KNG1, together with CD59 were also suggested as diagnostic biomarkers for patients with ovarian carcinoma [32].

Another key biomarker candidate, AMBP protein, belongs to the superfamily of lipocalin transport proteins [33] and may play a role in the regulation of inflammatory processes. Normal urine contains a very small amount of AMBP. AMBP was found to be over-represented in patients with type 1 diabetes mellitus in comparison with the control group [23]. Increased excretion of this protein was also linked to uropathy in a large study in 1998 [34]. Urinary AMBP expression was shown to significantly and steadily increase from CKD stage 1 to CKD stage 3 in 106 CKD patients [35]. In our study, a continuation of this trend in the later stages was not observed; our data showed higher fluctuation, however we were testing a much smaller dataset of patients. AMBP along with CD59 and APOD were also identified as prognostic candidates for acute kidney injury, all candidates showed increased expression [36].

APOD glycoprotein, the fourth crucial biomarker candidate, also belongs to the superfamily of lipocalin transport proteins [33]. APOD is a typical apolipoprotein with broad tissue distribution. Its plasma levels were reduced in mice with visceral obesity and altered lipid metabolism [37]. APOD is also involved in the mechanisms regulating protection from oxidative stress, where loss of mouse APOD function increases the sensitivity to oxidative stress [38]. We also observed reduced expression of APOD in urine in the later stages of CKD (CKD 3B - 4). On the other hand, we have observed a non-significantly increased level of APOD in patients with CKD stage 3A, which was in agreement with the observation of Vivekanandan-Giri's group [22]. Increased levels of APOD were also reported in critically-ill patients with non-recovery acute kidney injury [36]. The higher levels of APOD protein were reported in the failing heart and have been suggested as a biomarker in human end-stage heart failure [39].

The aim of this study was to demonstrate how it is possible to decipher a combination of markers from a compendium of differentially-expressed proteins using available statistical tools when more than two study groups are involved. The multivariate approach is able to distinguish more populations based on certain set of markers (diagnostic profile) where each marker plays its own partial role [40]. What is more, usage of more defined diseased groups is less prone to suffer of potential mistakes of pilot experiments comparing only two experimental groups (e.g. healthy controls with one stage of a selected disease). For example, if the control group is of different age than one diseased group and we are looking for the markers of disease not age, the risk is quite high that we also detect “age markers” that would be undistinguishable from other “disease markers”. On the other hand if we have more diseased groups, we have higher probability to recognize

such a false marker and to remove it from analysis. It should be noted that such evaluations (multivariate data analysis) are more powerful on larger sample sets [41], and because our own preliminary assumptions and calculations are based on only small number of samples, they should be verified on bigger, un-pooled datasets.

Conclusion

This preliminary work provides evidence that several unique urinary proteins are involved in early and later stages of CKD. The analyses are costly in work hours and resources, but we believe the cost-benefit is strong, in that discovery of new urinary biomarkers for CKD may eventually allow identification of people who are predisposed to develop CKD and also the progress with the disease. We propose that a selected combination of biomarkers could give a profile that leads to the desired separation of patients into different CKD stages. The suggested profile comprises of four proteins AMBP, APOD, AZGP1 and KNG1. Our preliminary assumptions and calculations are aimed to be verified on bigger, un-pooled datasets.

Acknowledgement

We certify that there is no conflict of interest with any financial organization regarding the material discussed in manuscript. We certify that this study involving human subjects is in accordance with the Helsinki declaration of 1975 as revised in 2000 and that it has been approved by the relevant institutional Ethical Committee. This study was supported by Go8 European Fellowship and BIOCEV CZ.1.05/1.1.00/02.0109 from the ERDF. We thank all volunteers who contributed to this study.

References

1. National Kidney Foundation (2002) K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis* 39: S1-266.
2. McLaughlin MJ, Courtney AE (2013) Early recognition of CKD can delay progression. *Practitioner* 257: 13-17.
3. Levey AS, Coresh J (2012) Chronic kidney disease. *Lancet* 379: 165-180.
4. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, et al. (2013) Chronic kidney disease: global dimension and perspectives. *Lancet* 382: 260-272.
5. Gu YM, Thijs L, Liu YP, Zhang Z, Jacobs L, et al. (2014) The urinary proteome as correlate and predictor of renal function in a population study. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 29: 2260-2268.
6. Nair S, Mishra V, Hayden K, Lisboa PJ, Pandya B, et al. (2011) The four-variable modification of diet in renal disease formula underestimates glomerular filtration rate in obese type 2 diabetic individuals with chronic kidney disease. *Diabetologia* 54: 1304-1307.
7. Keane WF (2000) Proteinuria: its clinical importance and role in progressive renal disease. *Am J Kidney Dis* 35: S97-105.
8. Macisaac RJ, Jerums G (2011) Diabetic kidney disease with and without albuminuria. *Curr Opin Nephrol Hypertens* 20: 246-257.
9. Brunzel NA (2013) *Fundamentals of Urine and Body Fluid Analysis*. 3rd ed.; Elsevier - Health Sciences Division: 464.
10. Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome biology* 7: R80.
11. Good DM, Zürgbilg P, Argilés A, Bauer HW, Behrens G, et al. (2010) Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Mol Cell Proteomics* 9: 2424-2437.
12. Rodríguez-Suárez E, Siwy J, Zürgbilg P, Mischak H (2014) Urine as a source for clinical proteome analysis: from discovery to clinical application. *Biochim Biophys Acta* 1844: 884-898.
13. Andrews GL, Simons BL, Young JB, Hawkrigde AM, Muddiman DC (2011) Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal Chem* 83: 5442-5446.
14. Faresjo T, Faresjo A (2010) To match or not to match in epidemiological studies - same outcome but less power. *Int J Environ Res Public Health* 7: 325-332.
15. Wisniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6: 359-362.
16. Calvano CD, Aresta A, Iacovone M, De Benedetto GE, Zamboni CG, et al. (2010) Optimization of analytical and pre-analytical conditions for MALDI-TOF-MS human urine protein profiles. *J Pharm Biomed Anal* 51: 907-914.
17. Mischak H, Allmaier G, Apweiler R, Attwood T, Baumann M, et al. (2010) Recommendations for biomarker identification and qualification in clinical proteomics. *Science translational medicine* 2: 46ps42.
18. Kolch W, Neuss C, Pelzing M, Mischak H (2005) Capillary electrophoresis-mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. *Mass Spectrom Rev* 24: 959-977.
19. Argiles Á, Siwy J, Duranton F, Gayraud N, Dakna M, et al. (2013) CKD273, a new proteomics classifier assessing CKD and its prognosis. *PLoS One* 8: e62837.
20. Thongboonkerd V (2007) Practical points in urinary proteomics. *J Proteome Res* 6: 3881-3890.
21. Baek SD, Baek CH, Kim JS, Kim SM, Kim JH, et al. (2012) Does stage III chronic kidney disease always progress to end-stage renal disease? A ten-year follow-up study. *Scand J UrolNephrol* 46: 232-238.
22. Vivekanandan-Giri A, Slocum JL, Buller CL, Basrur V, Ju W, et al. (2011) Urine glycoprotein profile reveals novel markers for chronic kidney disease. *Int J Proteomics* 2011: 214715.
23. Soggiu A, Piras C, Bonizzi L, Hussein HA, Pisanu S, et al. (2012) A discovery-phase urine proteomics investigation in type 1 diabetes. *Acta Diabetol* 49: 453-464.
24. Jaffa AA, Durazo-Arvizu R, Zheng D, Lackland DT, Srikanth S, et al. (2003) Plasma prekallikrein: a risk marker for hypertension and nephropathy in type 1 diabetes. *Diabetes* 52: 1215-1221.
25. Lim SC, Liying DQ, Toy WC, Wong M, Yeoh LY, et al. (2012) Adipocytokine zinc alpha2 glycoprotein (ZAG) as a novel urinary biomarker for normo-albuminuric diabetic nephropathy. *Diabet Med* 29: 945-949.
26. Riaz S, Alam SS, Srari SK, Skinner V, Riaz A, et al. (2010) Proteomic identification of human urinary biomarkers in diabetes mellitus type 2. *Diabetes Technol Ther* 12: 979-988.
27. Irmak S, Tilki D, Heukeshoven J, Oliveira-Ferrer L, Friedrich M, et al. (2005) Stage-dependent increase of orosomucoid and zinc-alpha2-glycoprotein in urinary bladder cancer. *Proteomics* 5: 4296-4304.
28. Katafigiotis I, Tyrizis SI, Stravodimos KG, Alamanis C, Pavlakis K, et al. (2012) Zinc α 2-glycoprotein as a potential novel urine biomarker for the early diagnosis of prostate cancer. *BJU Int* 110: E688-693.
29. Tomita H, Sanford RB, Smithies O, Kakoki M (2012) The kallikrein-kinin system in diabetic nephropathy. *Kidney Int* 81: 733-744.
30. Weinberg MS, Trebbin WM, Solomon RJ (1986) Urinary kininogen: a possible regulator of kinin formation in normal individuals and subjects with essential hypertension, end-stage renal and liver disease. *Advances in experimental medicine and biology* 1986: 119-125.
31. Rocchetti MT, Centra M, Papale M, Bortone G, Palermo C, et al. (2008) Urine protein profile of IgA nephropathy patients may predict the response to ACE-inhibitor therapy. *Proteomics* 8: 206-216.
32. Abdullah-Soheimi SS, Lim BK, Hashim OH, Shuib AS (2010) Patients with ovarian carcinoma excrete different altered levels of urine CD59, kininogen-1 and fragments of inter-alpha-trypsin inhibitor heavy chain H4 and albumin. *Proteome science* 8: 58.
33. Xu S, Venge P (2000) Lipocalins as biochemical markers of disease. *Biochim Biophys Acta* 1482: 298-307.
34. Everaert K, Delanghe J, Vande Wiele C, Hoebeke P, Dierckx RA, et al. (1998) Urinary alpha 1-microglobulin detects uropathy. A prospective study in 483 urological patients. *Clinical chemistry and laboratory medicine: CCLM / FESCC* 36: 309-315.
35. Nanayakkara S, Senevirathna ST, Karunaratne U, Chandrajith R, Harada KH, et al. (2012) Evidence of tubular damage in the very early stage of chronic kidney disease of uncertain etiology in the North Central Province of Sri Lanka: a cross-sectional study. *Environmental health and preventive medicine* 17: 109-117.
36. Aregger F, Uehlinger DE, Witowski J, Brunisholz RA, Hunziker P, et al. (2014)

-
- Identification of IGFBP-7 by urinary proteomics as a novel prognostic marker in early acute kidney injury. *Kidney Int* 85: 909-919.
37. Perdomo G, Kim DH, Zhang T, Qu S, Thomas EA, et al. (2010) A role of apolipoprotein D in triglyceride metabolism. *J Lipid Res* 51: 1298-1311.
38. Ganfornina MD, Do Carmo S, Lora JM, Torres-Schumann S, Vogel M, et al. (2008) Apolipoprotein D is involved in the mechanisms regulating protection from oxidative stress. *Aging Cell* 7: 506-515.
39. Wei YJ, Huang YX, Zhang XL, Li J, Huang J, et al. (2008) Apolipoprotein D as a novel marker in human end-stage heart failure: a preliminary study. *Biomarkers: biochemical indicators of exposure, response, and susceptibility to chemicals* 13: 535-548.
40. Jolliffe IT (2002) *Principal Component Analysis* 2nd ed.; Springer.
41. Esbensen Kim H (2010) *Multivariate Data Analysis - in practice*. 5th ed.; CAMO Software
42. Zerefos PG, Vlahou A (2008) Urine sample preparation and protein profiling by two-dimensional electrophoresis and matrix-assisted laser desorption ionization time of flight mass spectroscopy. *Methods in molecular biology* 428: 141-57. *Voltus cupimor tertatiliaci sentemuntrae iu venatiliica; ne factum rebusum*