Clinical Kidney Journal, 2023, vol. 16, no. 11, 1965–1973

https://doi.org/10.1093/ckj/sfad211
Advance Access Publication Date: 2 September 2023
Original Article

ORIGINAL ARTICLE

Urinary peptidome analysis in CKD and IgA nephropathy

Zewen Li[1,*], Nianyi Zeng[1,*], Xin Zhao[1,*], Xuedong Chen[1], Guangqing Liang[1], Haiyue Liu[1,2], Jinyan Lin[3], Peizhuang Zheng[4], Xingtao Lin[5], Hongwei Zhou[1] and Daowen Zheng[1,6]

[1]Division of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, China, [2]Division of Laboratory Medicine, The First Affiliated Hospital of Xiamen University, Xiamen, China, [3]Department of Health Management, Nanfang Hospital, Southern Medical University, Guangzhou, China, [4]Department of Medical Imaging, Nanfang Hospital, Southern Medical University, Guangzhou, China, [5]Department of Pathology, Guangdong Provincial People's Hospital, Guangzhou, China and [6]Department of Gerontology, Zhujiang Hospital, Southern Medical University, Guangzhou, China

*These authors contributed equally.
Correspondence to: Hongwei Zhou; E-mail: hzhou@smu.edu.cn; Daowen Zheng; E-mail: 35156629@qq.com

ABSTRACT

Background. Chronic kidney disease (CKD) has emerged as a significant challenge to human health and economic stability in aging societies worldwide. Current clinical practice strategies remain insufficient for the early identification of kidney dysfunction, and the differential diagnosis of immunoglobulin A nephropathy (IgAN) predominantly relies on invasive kidney biopsy procedures.

Methods. First, we assessed a case–control cohort to obtain urine samples from healthy controls and biopsy-confirmed CKD patients. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) was applied to detect urinary peptide and then these urinary peptide profiles were used to construct diagnostic models to distinguish CKD patients from controls and identify IgAN patients from other nephropathy patients. Furthermore, we assessed the robustness of the diagnostic models and their reproducibility by applying different algorithms.

Results. A rapid and accurate working platform for detecting CKD and its IgAN subtype based on urinary peptide pattern detected by MALDI-TOF MS was established. Naturally occurring urinary peptide profiles were used to construct a diagnostic model to distinguish CKD patients from controls and identify IgAN patients from other nephropathy patients. The performance of several algorithms was assessed and demonstrated that the robustness of the diagnostic models as well as their reproducibility were satisfactory.

Conclusions. The present findings suggest that the CKD-related and IgAN-specific urinary peptides discovered facilitate precise identification of CKD and its IgAN subtype, offering a dependable framework for screening conditions linked to renal dysfunction. This will aid in comprehending the pathogenesis of nephropathy and identifying potential protein targets for the clinical management of nephropathy.

Keywords: chronic kidney disease, diagnostic models, MALDI-TOF, urinary peptide

Received: 9.5.2023; Editorial decision: 22.8.2023

INTRODUCTION

Chronic kidney disease (CKD) is often characterized by progressive loss of renal function over 3 months, and it may eventually lead to end-stage renal disease (ESRD) [1]. The prevalence of CKD has risen dramatically in aging societies and CKD has become a burden on human health and economies [2, 3]. Immunoglobulin A nephropathy (IgAN), the most common subtype of CKD, is pathologically a primary glomerulonephritis due to mesangial deposition of IgA-related immune complexes [4, 5].

In current clinical practice, renal function is generally evaluated based on serum chemistry profiles and/or urine analyses, such as the levels of serum creatinine and urinary microalbumin and the urinary albumin to creatinine ratio (ACR) or estimated glomerular filtration rate (eGFR) [1, 6]. However, these parameters have major limitations [7], as they are nonspecific to kidney diseases and are often late manifestations of renal damage; additionally, the differential diagnosis of subtypes of CKD relies on tissue biopsy, an invasive approach that may cause deadly complications. Therefore, novel biomarkers for renal function are essential to meet the needs for noninvasive testing, early detection, differential diagnosis and even predicting prognosis.

Driven by such clinical needs, new diagnostic methods and technologies based on mass spectrometry have emerged. However, at present, they have not been well promoted or applied in clinical practice due to the barriers associated with advanced equipment and the lack of standardized protocols, including bioinformatic analysis processes and the quality control of such processes [8–12].

Urine has recently emerged as a potential source of noninvasive diagnostic biomarkers for human disease. In addition to urinary protein and microalbumin, urine contains a large number of endogenous peptides [13] that are accurately associated with complex diseases, including kidney diseases [13–16], cardiovascular diseases [17], graft-versus-host diseases [18] and cholangiocarcinoma [19].

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) can be used to acquire peptide profiles from various types of samples in a much simpler way than other mass spectrometry techniques. Previous studies have shown its potential in searching for novel biomarkers for various diseases and in biosamples [13, 20–23], providing a rapid and sensitive analysis of biomolecules [13, 20, 21], including urinary peptides. Currently, most diagnostic labs are equipped to perform MALDI-TOF MS, making its widespread clinical application feasible.

To determine the probability of naturally occurring urinary peptides in the diagnosis of CKDs, a case–control cohort of 194 biopsies from CKD patients and 48 healthy volunteers with normal kidney function was assessed in the present study. We performed a MALDI-TOF MS–based urinary peptidomic analysis to identify potential biomarkers for early detection, differential diagnosis and prognosis prediction.

MATERIALS AND METHODS

Study subjects and ethical approval

Participants with diagnosed CKD and healthy volunteers were recruited from April 2019 to January 2020 at Zhujiang Hospital, Nanfang Hospital and Guangdong Provincial Hospital. This study was approved by the Ethics Committee of Zhujiang Hospital of Southern Medical University (2020-KY-016-01), and the clinical trial registration number was ChiCTR2000033229 (Chinese Clinical Trial Registry, www.chictr.org.cn). We enrolled 194 volunteers with abnormal kidney function diagnosed pathologically by renal biopsy at nephrology clinics as CKD group, which included 82 cases of IgAN, 35 cases of membranous nephropathy, 33 cases of minimal change nephrosis, 15 cases of focal segmental glomerulosclerosis, 17 cases of lupus nephritis, 7 cases of diabetic nephropathy and 5 cases of membranoproliferative glomerulonephritis. Forty-eight volunteers with normal kidney function (eGFR $\geq$90 mL/min/1.73 m^2 and ACR $<$30 mg/g) were enrolled as controls, taking into consideration their age and sex to ensure a proper match with CKD group.

Urine sample and metadata collection

The patient is instructed to collect a random mid-stream urine sample. To do so, they should first wash their hands thoroughly with soap and water. Then, the genital area is cleaned with a mild antiseptic wipe, moving from front to back. After voiding the first portion of urine into the toilet, they should collect the middle portion of the urine stream in a sterile, wide-mouthed container, while ensuring that the container does not come into contact with the skin or any other surfaces. Once an adequate amount of mid-stream urine is collected (usually about 10 mL), the container should be tightly sealed and labelled with the patient's name, date, and time of collection. Urine samples were centrifuged for 15 min at 3000 rpm. The supernatant was isolated and stored at –80°C until analysis. All urine samples were processed with the same standardized experimental protocols and stored in the same type of plastic vials and boxes. Metadata information, including clinical demographic and kidney-related biochemical parameters, was gathered.

MALDI-TOF MS peptide profiling

Each urine sample was mixed with a matrix solution of α-cyano-4-hydroxycinnamic acid (0.3 g/L) at a ratio of 1:10. A total of 2 µL of sample/matrix solution was spotted onto the MALDI target (IntelliBio, Qingdao, China) and left to crystallize fully and equably at 40°C. The samples from both study groups were evaluated in a random order, and the disease status was blinded to minimize variability and systematic errors. MS analyses were conducted with a QuanTOF mass spectrometer (IntelliBio, Qingdao, China) in linear positive mode. Positively charged ions were detected in the m/z range of 1000–10 000 Da. The MS spectra were externally calibrated with a mixture of peptide calibration standards, including 757.40, 1045.00, 2465.20, 3494.65 and 5734.50 Da. The average mass deviation was less than 2 ppm. The following ion source parameters were used as follow: source voltage, 20 kV; detector voltage, 0.48 kV; laser pulse energy, 4.8 μJ; pulse frequency, 3000 Hz; focus mass, 5000 Da; motion scanning speed, 1 mm/sec; averaged shots per spectrum, 800. QuanTOF viewer software (IntelliBio, Qingdao, China) was applied for the acquisition and processing of the spectra. Data analysis of spectra from all urine samples was conducted using R software (v.4.0.3). Detailed methods are provided in the Supplementary data.

Statistical analysis

Principal component analysis (PCA) was conducted using the prcomp function in the 'stats' package (version 4.1.1). Bray–Curtis distance matrices were calculated based on the feature matrix using the vegdist function in the 'vegan' package (version 2.5.7). Differences in groups represented in PCA plots were tested by permutational multivariate ANOVA with 1000 permutations.

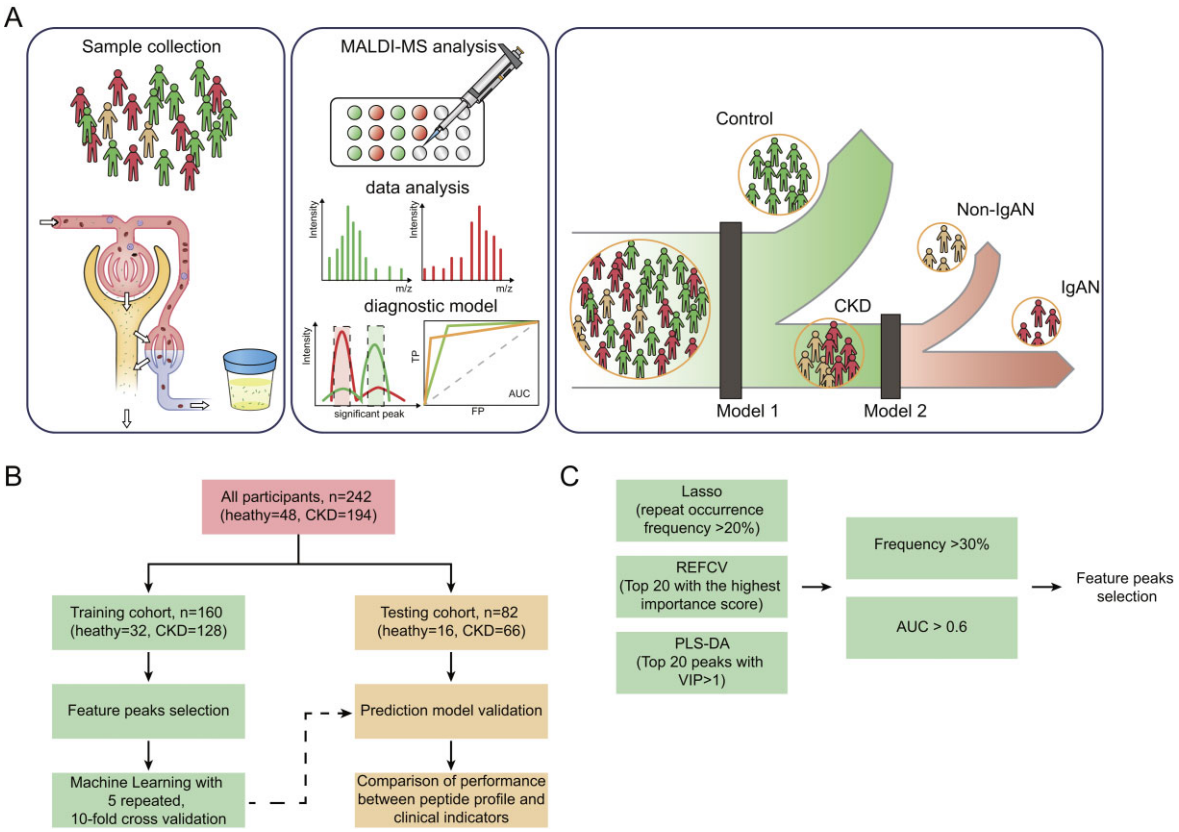

Figure 1: Study design and analysis pipeline. (**A**) Scheme of establishing a diagnostic model and prognostic index for rapid and accurate screening of CKD and its subtypes patients. Urine sample gathered from recruited participants were analysed by MALDI-TOF after simple sample pretreatment. The significant feature peaks were identified and used to construct the diagnostic model with seven machine learning algorithms. (**B**) All participants were divided into a training cohort and an independent testing cohort. In training cohort, we select disease-related feature peaks to establish classification models with different machine learning methods. The performance of models was verified by testing cohort and compared with that of models constructed by clinical indicators. (**C**) Workflow for feature selection. Three machine learning methods were employed, namely, partial least squares-discriminant analysis (PLS-DA), least absolute shrinkage and selection operator (LASSO) and recursive feature elimination with cross-validation (RFECV), to select relevant features associated with diseases in the training cohort ($n = 160$). In the PLS-DA method, variable important in projection (VIP) scores of peaks were calculated, and the top 20 were chosen as disease features. For the LASSO method, 80% of the samples in the training dataset were randomly selected with 200 repetitions, and peaks with more than 20% occurrence frequency were considered as disease features. The RFECV method selected the top 20 peaks with automatic tuning of all the features chosen by cross-validation accuracy. By combining the above feature peaks and through further empirical verification based on frequency (>30%) and AUC (>60%) between groups, the remaining features were identified as distinct peaks between the groups.

Significant differences were determined by the Wilcoxon rank-sum test between two groups and the Kruskal–Wallis test among multiple groups. The *P*-value was adjusted by the Benjamini and Hochberg method. $P < .05$ was considered statistically significant. Spearman correlation analysis was used to examine the correlation between feature peaks and laboratory tests.

RESULTS

Experimental workflow, demographic and clinical characteristics of participants

As demonstrated in Fig. 1, in this study, we established an efficient urinary peptidome workflow based on MALDI-TOF MS, and the spectrum data generated were analysed by using machine learning to identify patients with CKD and its IgAN subtype. In detail, a total of 242 participants, including 194 patients with kidney diseases and 48 controls with normal kidney function, were recruited for this study. Clean urine samples from participants were self-collected and analysed by using MALDI-TOF MS after simple preprocessing. Participants were randomly divided into a training cohort (128 CKD patients and 32 healthy controls) or a testing cohort (66 CKD patients and 16 healthy controls) with an allocation of 2:1. The characteristics of the participants are shown in Table 1. Compared with controls, CKD patients had worse renal function with a decreased eGFR and increased albuminuria. Additionally, other physiological indicators showed that CKD patients had higher systolic blood pressure and blood glucose levels, consistent with previous epidemiological reports [4, 5]. All CKD participants underwent kidney biopsy to confirm their renal pathological diagnosis with Lee's grading and the Oxford classification system.

Naturally occurring urinary peptides can be used to construct a diagnostic model to distinguish CKD patients from healthy individuals

All urine samples were prepared and analysed using an in-house standard protocol resulting in individual datasets containing information of generally 32–140 features (presumably peptides)

Table 1: The characteristics of the participants.

	Training cohort			Test cohort		
	Control ($n = 32$)	CKD ($n = 128$)	P-value	Control ($n = 16$)	CKD ($n = 66$)	P-value
Age (years)	34 (18, 61)	37.5 (15, 77)	.273	36.5 (22.0, 56.0)	39.0 (16.0, 78.0)	.413
Sex, *n* (%)						
Female	17 (53)	59 (46)	.607	6 (38)	30 (45)	.768
Male	15 (47)	69 (54)		10 (62)	36 (55)	
eGFR (mL/min/1.73 m^2)	109 (91.4, 137)	90.8 (7.45 145)	<.001	104 (91.3, 126)	74 (12.1, 138)	.00396
ACR (mg/g)	0 (0, 28.0)	1680 (10.2,47 800)	<.001	0 (0, 21.2)	2260 (10.6, 77 700)	<.001
MA_U (mg/L)	0.25 (0, 27)	1070 (4.21, 25 800)	<.001	0 (0, 24.0)	1510 (2.37, 12 400)	<.001
Ucrea (μmol/L)	11 300 (1530, 28 300)	7610 (649, 29 600)	.0194	11 900 (1700, 28 800)	6610 (437, 24 700)	.0623
Serum albumin (g/L)	44.3 (40.4, 47,4)	34.3 (11.2, 51.2)	<.001	41.7 (41.2, 47.1)	32.7 (15.7, 47.2)	<.001
TC (mmol/L)	5 (2.97, 7.40)	5.36 (0.840, 66.3)	.0536	5.10 (4.10, 7.30)	6.16 (0.850, 16.0)	.0276
Scr (μmol/L)	67 (51.0, 90.0)	83.5 (30.0, 569)	<.001	72.0 (54.0, 89.5)	96.0 (37.0, 478)	.0181
SBP (mmHg)	126 (100, 147)	131 (92.0, 192)	.105	121 (96.0, 142)	129 (92.0, 179)	.027
DBP (mmHg)	76.5 (56.0, 93.0)	78 (55.0, 122)	.257	68.0 (52.0, 89.0)	76.0 (58.0, 105)	.00892
Glu (mmol/L)	4.7 (3.70, 6.10)	4.72 (3.07, 16.3)	.98	4.75 (4.20, 5.90)	4.80 (2.64, 13.5)	.774

Data are presented as median (min, max). Chi-square test for classification variables. Wilcoxon rank-sum test for continuous variables.
Ucrea, urine creatinine; TC, total cholesterol; Scr: serum creatinine; SBP, systolic blood pressure; DBP, diastolic blood pressure; Glu, glucose.

per sample. PCA using all detected features (Fig. 2A) clearly showed that the urinary peptidomic pattern was significantly different ($P = .001$, $R^2 = 0.075$) between the healthy individuals and CKD patients based on the Bray–Curtis dissimilarity matrix. To better understand the major peptide features contributing to the shift in the urinary peptidome, we screened for biomarkers according to the filter shown in Fig. 1C by using the training cohort dataset. Thirty-one peptide features were filtered out, of which 25 peaks were enriched in healthy individuals and 6 peaks were enriched in CKD patients (Fig. 2B, Supplementary data, Table S2). Then, a classification model based on the random forest algorithm was constructed using these 31 features to distinguish the two groups. A receiver operating characteristic (ROC) curve was used to evaluate the performance of the model. In the training cohort, the area under the ROC curve (AUC) was 98.3% (Fig. 2C). We applied the model to the independent testing cohort to validate the efficiency of the classification model, and the AUC of the testing cohort was 99% (Fig. 2D), indicating that CKD patients can be well distinguished from healthy controls using the naturally occurring urinary peptidome. Furthermore, we compared the normalized intensity of these 31 CKD-related features between controls and patients with different eGFR levels (Supplementary data, Fig. S4) and identified features for which intensity was closely related to kidney function. Four of 31 features are shown in Fig. 2E–H, 3 of which (M/Z1250.48, M/Z1909.6 and M/Z1975.09) were decreased as the eGFR declined, and one of them (M/Z2925.93) was increased in both the training cohort and testing cohort. In addition, we demonstrated the correlation between these four features and clinical parameters (Fig. 2I). The three features mentioned above were negatively related to increased urinary protein loss and serum creatinine, while M/Z2925.93 showed a harmful role with clinical parameters that reflect renal function.

Distinguishing IgAN patients from other nephropathy patients (non-IgAN) using machine learning based on IgAN-specific feature peaks

IgAN is the most common glomerular disease in the clinic, and it is also the first major cause of ESRD [4, 5]. In our cohort, IgAN subtype accounted for the highest proportion among CKD patients (42.3%, 82/194), which aligns with the epidemiological survey of CKD in China (Supplementary data, Table S1) [24]. IgAN patients were younger than non-IgAN patients and exhibited slightly better renal function indicators based on clinical assessments. Additionally, IgAN patients had lower blood pressure and glucose levels (Table 2). We sought to distinguish IgAN patients from all CKD patients. PCA of all features revealed the difference between control, IgAN and non-IgAN patients (Fig. 3A, $P = .001$, $R^2 = 0.099$); among all features, 45 features either contributed to the discrimination between healthy controls and IgAN patients (Fig. 3B left) or distinguished IgAN patients from other glomerulopathy patients (Fig. 3B right) and were defined as IgAN-related features (Supplementary data, Table S3 and Fig. S5). PCA based on the IgAN-related features showed that the urinary peptide patterns of healthy controls, IgAN patients and non-IgAN patients could be significantly distinguished (Fig. 3C, $P = .001$, $R^2 = 0.205$). Then, a classification model based on the random forest algorithm was constructed using these 45 features to distinguish IgAN patients from other glomerulopathy patients. The AUCs of the ROC curve were 79% and 76.2% in the training cohort and testing cohort, respectively (Fig. 3D). IgAN is often diagnosed by kidney biopsy to specify the immunopathological lesion of the kidney tissue.

At present, the commonly used pathological grading system is Lee's classification, which is a descriptive classification method, and the Oxford classification system which was developed in 2009 as a pathological classification system for IgAN to predict the risk of disease progression [25–27]. Thus, we sought to determine the relationship between the urinary peptidome and the results of kidney biopsy. PCA (Fig. 3E) demonstrated that the urinary peptidome was shifted as IgAN progressed according to Lee's grading. The Oxford classification is regarded as a more objective indicator since it consists of five classification criteria [28, 29], namely, the mesangial score, endocapillary hypercellularity absence, segmental glomerulosclerosis absence, tubular atrophy and crescents (MEST-C), within which MST is regarded as more relevant to the progression of ESRD [26]. Thus, we explored the difference in the urinary peptidome under MST classification. PCA (Fig. 3F–H) showed that the urinary peptidome shifted from right to left on the PC1 axis as the M, S and T

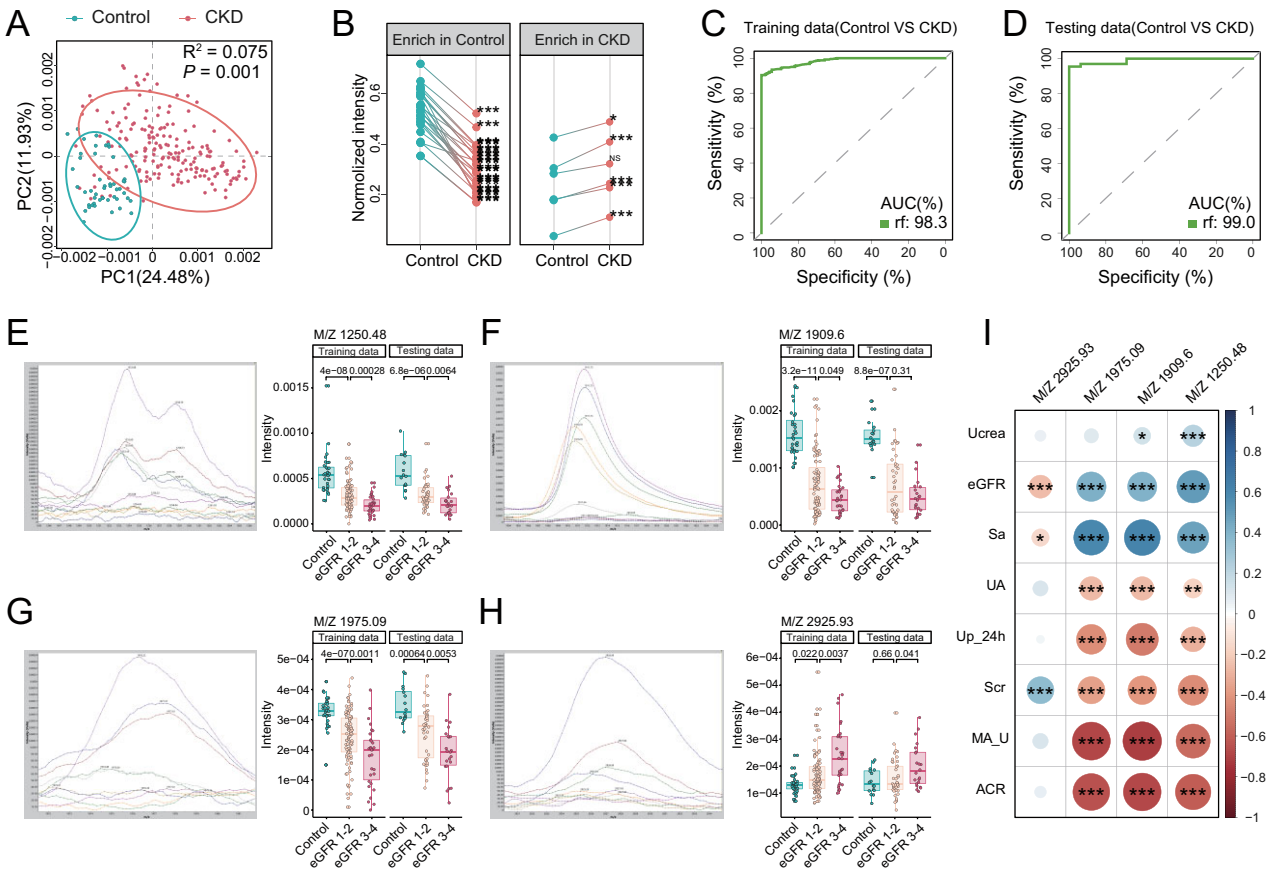

Figure 2: The shifted urinary peptide profile between controls and CKD patients. (**A**) PCA analysis using urinary peptide profile showed significant difference between healthy participants and CKD patients. (**B**) The feature selection workflow used to identify 31 distinct peaks between control participants and CKD patients. The line chart represented the normalized intensity of these feature peaks. Among the 31 peaks, 25 were significantly enriched in healthy participants, while 5 out of the 6 peaks were significantly enriched in CKD patients. (**C**) ROC curves of machine learning models distinguishing between healthy participants and CKD patients with five repeats and 10-fold cross-validation in training cohort. (**D**) ROC curves of machine learning models distinguishing between healthy participants and CKD patients in testing cohost. (**E–H**) Four feature peaks with significantly increased/decreased intensity among control, eGFR1–2 and eGFR3–4, respectively. In each subfigure, the left panel displays partial raw mass spectra from MALDI-TOF MS, while the right panel illustrates the intensity of urinary peptides among groups. (**I**) The association of six feature peaks with kidney-related clinical parameters. Sa, serum albumin; UA, uric acid; Up_24 h, 24-h urinary protein quantity; Scr, serum creatinine.

classification grading was aggravated. Furthermore, we found many features associated with Lee's classification grading, in which M/Z1752.24 (Fig. 3I) and M/Z1932.05 (Fig. 3J) decreased as Lee's score increased. These findings indicated that IgAN-specific features detected by MALDI-TOF MS may be associated with the progression of ESRD and can be recognized as potential biomarkers.

The robustness of the diagnostic models and batch results indicated satisfactory reproducibility

As mentioned above, we used the random forest algorithm to build diagnostic models, which resulted in an AUC of approximately 99% for the identification of CKD patients and nearly 80% for IgAN patients. The random forest algorithm can be used to process data of high dimensions and includes a classifier with multiple decision trees. Although no algorithm solves all problems perfectly in the field of machine learning, especially for supervised learning (e.g. predictive modelling), in the current study we tried a variety of machine learning methods and used algorithms that are not sensitive to imbalanced data (e.g. ctree, SVM) to verify the robustness of the results. In addition to random forest, we used six different machine learning algorithms to generate diagnostic models. In the model that distinguished CKD patients and healthy controls, the six models achieved AUCs of 90.2%–98.5% in the training cohort and 83.8%–99.7% in the testing cohort (Fig. 4A). In the model that distinguished IgAN from non-IgAN patients (Fig. 4C), the other six algorithms achieved AUCs from 61.3% to 80.0% in the training cohort and 61.2%–85.1% in the testing cohort.

The coefficients of variation of the seven models' AUC-ROCs were calculated to assess the reproducibility of the machine learning modelling of the MALDI-TOF MS spectrum acquired on five different days. In the CKD and healthy control models, nearly all coefficients of variation (CVs) of the AUC-ROCs of the seven models generated from the spectrum acquired from five different batches were less than 5% in the training and testing data (Fig. 4B), and in the model that identified IgAN from CKD patients, CVs were approximately 10% due to the rather small amount of data (Fig. 4D). These results show that the reproducibility of machine learning modelling using five different batches of MALDI-TOF MS spectra is satisfactory in discriminating CKD patients from healthy controls.

Finally, we compared the error counts by using urinary peptidomic modelling of urinary microalbumin or ACR to screen for CKD. By using a random forest model for diagnosis, two control

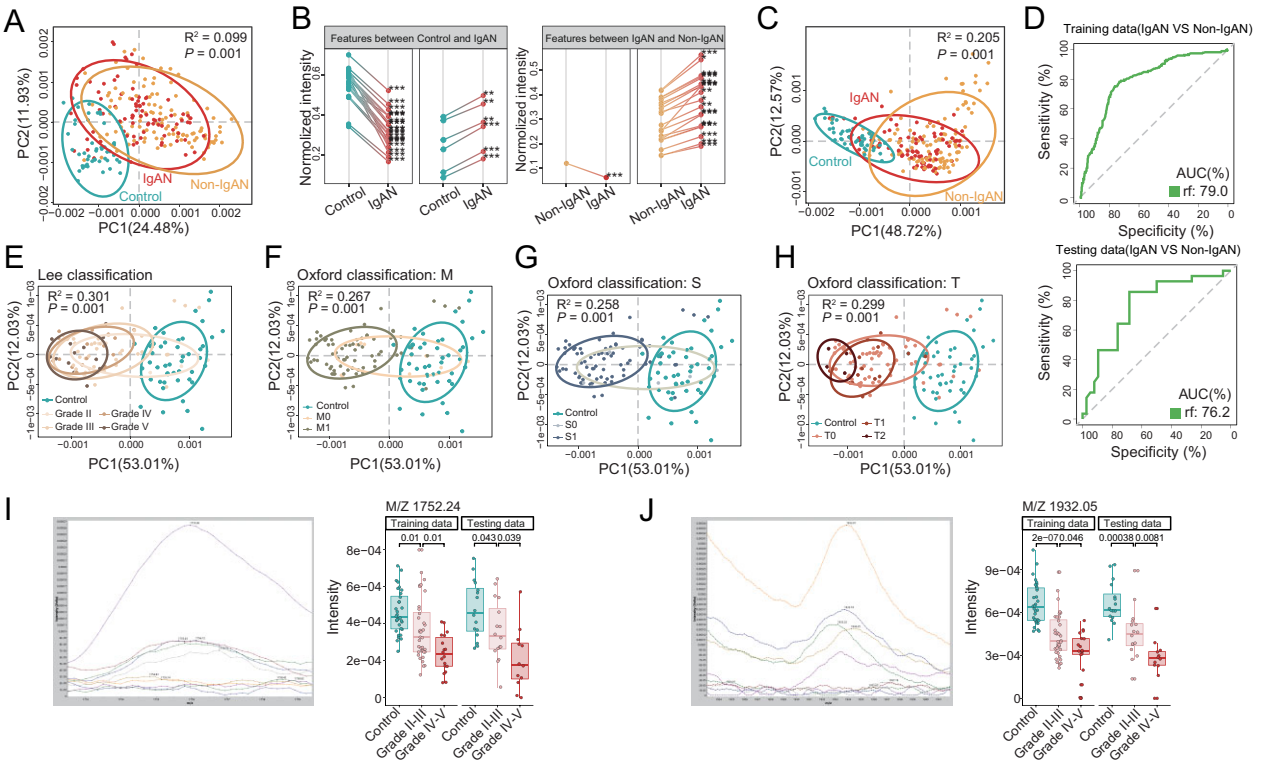

Figure 3: The shifted urinary peptide profile between controls, and IgAN and non-IgAN patients. (**A**) PCA analysis using total urinary peptide profile showed significant difference between groups. (**B**) The line chart showed the 45 feature peaks selected according to feature selection workflow. Among 45 peaks, 28 peaks (22 peaks enriched significantly in healthy participants compared with IgAN patients, while 6 peaks enriched significantly in IgAN patients) were selected between healthy participants and IgAN patients and 21 peaks (one peak enriched significantly in non-IgAN subpopulations compared with IgAN patients, while 20 peaks enriched significantly in IgAN patients) were selected between IgAN patients and non-IgAN subpopulations. (**C**) PCA analysis using the 45 features peaks. (**D**) ROC curves by machine learning method. (**E**) PCA analysis using 45 feature peaks showed urine profile shift progressively during development of IgAN in Lee classification. (**F–H**) PCA analysis using 45 feature peaks showed urine profile shift progressively during development of IgAN in Oxford classification M (F), S (G) and T (H), respectively. (**I, J**) The intensity of two feature peaks significantly increased/deceased between controls, Grade II-III and Grade IV-V, respectively. For the Lee's grading system for IgAN, patients classified as Grade II-III are reflecting moderate disease severity. Those in Grade IV-V are indicating severe disease with significant kidney damage.

Table 2: The characteristics of the participants among control, IgAN and non-IgAN groups.

	Control (N = 48)	IgAN (N = 82)	Non-IgAN (N = 112)	P-value
Age (years)	35.5 (18.0, 61.0)	33.0 (15.0, 73.0)	46.5 (16.0, 78.0)	<.001
Sex, *n* (%)				
Female	23 (47.9)	46 (56.1)	43 (38.4)	.049
Male	25 (52.1)	36 (43.9)	69 (61.6)	
eGFR (mL/min/1.73 m^2)	108 (91.3, 137)	89.1 (7.92, 138)	85.8 (7.45, 145)	<.001
ACR (mg/g)	0 (0, 28.0)	986 (10.2, 30 700)	2490 (10.6, 77 700)	<.001
MA_U (mg/L)	0 (0, 27.0)	808 (2.37, 16 600)	2040 (4.21, 25 800)	<.001
Ucrea (μmol/L)	11 700 (1530, 28 800)	6940 (437, 29 600)	7640 (895, 24 900)	.00929
Serum albumin (g/L)	41.8 (40.4, 47.4)	37.9 (15.6, 50.8)	25.8 (11.2, 51.2)	<.001
TC (mmol/L)	5.00 (2.97, 7.40)	4.83 (0.850, 18.9)	6.33 (0.840, 66.3)	<.001
Scr (μmol/L)	68.5 (51.0, 90.0)	88.0 (45.0, 569)	86.0 (30.0, 521)	<.001
SBP (mmHg)	124 (96.0, 147)	121 (92.0, 186)	134 (92.0, 192)	<.001
DBP (mmHg)	73.0 (52.0, 93.0)	76.0 (55.0, 115)	79.0 (56.0, 122)	.0451
Glu (mmol/L)	4.70 (3.70, 6.10)	4.67 (3.07, 9.62)	4.87 (2.64, 16.3)	.0488

Data are presented as median (min, max). Chi-square test for classification variables. Kruskal–Wallis test for continuous variables.
Ucrea, urine creatinine; TC, total cholesterol; Scr: serum creatinine; SBP, systolic blood pressure; DBP, diastolic blood pressure; Glu, glucose.

and one CKD patients were incorrectly diagnosed, while when using urinary microalbumin or ACR to screen for CKD in current practice, five or seven CKD patients were incorrectly diagnosed separately, indicating that urinary peptide-based RF modelling had a better accuracy for CKD diagnosis (Fig. 4E).

DISCUSSION

In this study, we developed a rapid and accurate pipeline for detecting CKD and its IgAN subgroup based on multivariate urinary peptidomic pattern detection with MALDI-TOF MS. These

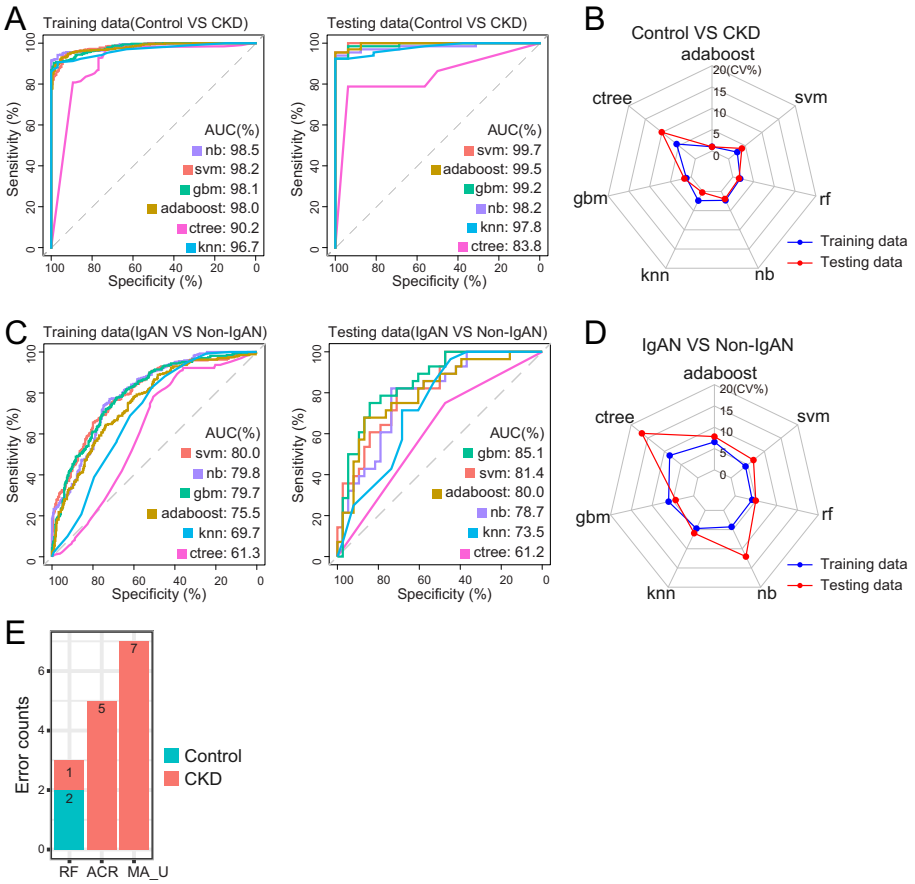

Figure 4: Results of repeated testing of six machine learning models. (**A**) The ROC curves of various machine learning models (including nb, svm, gbm, adaboost, ctree and knn algorithms) used to distinguish between healthy participants and CKD patients. These models were trained using five repeats and 10-fold cross-validation in the training cohort. The right side of the figure shows the ROC curves of the same models used to distinguish between healthy participants and CKD patients in the testing cohort. (**B**) Radar plot showing the CV of seven machine learning models in distinguishing between controls and CKD. This plot represents five-repeats testing. (**C**) The ROC curves of machine learning models (nb, svm, gbm, adaboost, ctree and knn algorithms) used to distinguish between IgAN and Non-IgAN patients. The models were trained using five repeats and 10-fold cross-validation in the training cohort. The right side of the figure shows the ROC curves of the same models used to distinguish between IgAN and non-IgAN patients in the testing cohort. (**D**) Radar plot showing the CV of seven machine learning models in distinguishing between IgAN and non-IgAN. This plot represents five-repeats testing. (**E**) The error counts for the random forest model, ACR and MA_U are displayed. CKD errors are denoted in red, while control errors are in blue.

patterns have superior performance to current clinical parameters (Supplementary data, Fig. S1). Our data were robustly verified by an independent testing cohort as well as multiple machine learning algorithms with five repeats and 10-fold cross-validation. Of note, our results suggest that these patterns maintain consistent diagnostic value across time and batches.

We established a CKD cohort with confirmed pathological diagnosis by kidney biopsy, and clinical data suggested that our cohort was at an early stage of kidney disease (Supplementary data, Fig. S1A). Although urinary microalbumin and ACR are currently good indicators of nephropathy in the clinic, from our data, the AUCs of ACR and urine microalbumin (MA_U) in distinguishing control and CKD were 99.66% and 99.11%, respectively (Supplementary data, Fig. S1B), while MA_U varied greatly between individuals (Table 1), and the use of MA_U's reference range alone to diagnose CKD resulted in a greater false-negative rate in the current cohort (Fig. 4E).

We identified a panel of IgAN-specific urinary biomarkers, and the combination of feature peaks compared with clinical indicators showed potential diagnostic value for distinguishing IgAN patients from other CKD patients. In addition, we screened out certain urinary peptides that were strongly related to kidney-related clinical parameters, including the eGFR and urinary microalbumin level. The confirmation of IgAN diagnosis relies on kidney biopsy [4, 5]; currently, Lee's grading and the Oxford classification are widely used in the clinic [30]. Lee's grading system was proposed in 1982 and is easy to use and apply [31]. The Oxford classification system was first published in 2009 and has been increasingly used in clinical practice around the world. Both Lee's grading and the Oxford classification were used to predict the clinical prognosis of IgAN patients based on renal biopsy [26, 27]. In this study, all CKD patients underwent kidney biopsy and were pathologically evaluated by Lee's grading and the Oxford classification. We found a significant association between Lee's grading and the M/S/T features of the Oxford classification (Supplementary data, Figs S2 and S3) [28]. It has also been reported [26] that the M/S/T features of the Oxford classification can be used in the prediction of IgAN prognosis; thus, we proposed that the urinary peptide pattern may be a biomarker associated with the prognosis of IgAN.

In all, 25 of 31 CKD-related and 31 of 45 IgAN-specific urinary peptides were identified by liquid chromatography–MS/MS

without proteolysis (Supplementary data, Tables S4 and S5); among those, uromodulin had the most pronounced change detected in this study. Previous study indicated that uromodulin protein has emerged as a promising biomarker for assessing tubular function and nephron mass, offering a potential alternative to existing markers that primarily focus on glomerular function. Our study further identified a specific fragment of uromodulin, the 1909.6 m/z peptide, to have the strongest association with renal function, which suggests that measuring this specific fragment may be clinically relevant for assessing renal impairment or CKD. However, the detailed mechanisms underlying this association and the potential clinical applications require further investigation. Notably, we have discovered a potential new biomarker called Methyl-CpG-binding domain protein 5 (MBD5 fragment, 1283.46 m/z), a family member of MBD2, which has been reported to be associated with CKD [32, 33]. In the current study, MBD5 fragment 1283.46 m/z was downregulated in CKD; further research will reveal the specific mechanisms and explore the potential of MBD5 as a biomarker in CKD.

Prior to this study, Good *et al.* [34] reported that a support-vector classification model based on 273 urinary peptides detected by capillary electrophoresis–MS can distinguish between healthy subjects and CKD patients. In addition, in small sample size cohorts [35], urinary peptides have been found to be significantly different between healthy and CKD individuals, even those with distinct types of kidney disease, based on a platform combining a magnetic bead separation system with MS. In this study, our urinary peptides were detected based on the MALDI-TOF MS platform, which is equipped in most diagnostic labs and requires simple preprocessing. We validated that urinary peptidomics can overcome the batch effect and has stable performance with acceptable CV% in repeated experiments. These results indicate that the urinary peptidome has potential value as a novel, noninvasive and robust biomarker for kidney diseases.

In conclusion, our current results indicated that the identified CKD-related and IgAN-specific urinary peptides enable accurate detection of CKD and its IgAN subtype, providing a robust pipeline for screening diseases associated with kidney dysfunction. The results of our study have provided valuable insights into the early screening of CKD patients. In addition, our study has shown that the peptide we identified holds promising potential as a diagnostic biomarker for use in clinical settings, utilizing both MALDI-TOF MS platform and immunology methods. These findings have opened up new avenues for improving the diagnosis and management of CKD and IgAN patients. Despite the potential diagnostic value, in the future, larger independent cohorts are needed to verify the reliability of model extrapolation before being applied in the clinic as a routine screening method to understand the pathogenesis of nephropathy and provide potential protein targets for the clinical treatment of nephropathy.

SUPPLEMENTARY DATA

Supplementary data are available at *ckj* online.

ACKNOWLEDGEMENTS

We thank the Ministry of Science and Technology of China and Natural Science Foundation of China for financially supporting this study.

FUNDING

This work was funded by the National key research and development program of China (2022YFA0806400), National Natural Science Foundation of China (Key Program, No. 82 130 068) and National Science Fund for Distinguished Young Scholars (No. 81 925 026) to H.Zhou.

DATA AVAILABILITY STATEMENT

Data collected for the study, including deidentified individual participant data and a data dictionary defining each field in the set, and the raw data for MALDI-TOF MS and liquid chromatography–MS/MS of this study will be made available from the corresponding author on reasonable request.

AUTHORS' CONTRIBUTIONS

D. Zheng and H. Zhou designed the study; Z.Li, N.Zeng performed the experiments and wrote the manuscript., Z. Li and G. Liang analyzed the data, X. Zhao, X. Chen, H. Liu, J. Lin, P. Zheng, X. Lin enrolled the participants and collected samples, all authors reviewed and proofed the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

REFERENCES

1. Stevens PE, Levin A. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med* 2013;**158**:825–30. https://doi.org/10.7326/0003-4819-158-11-201306040-00007
2. Jha V, Garcia-Garcia G, Iseki K *et al*. Chronic kidney disease: global dimension and perspectives. *Lancet* 2013;**382**:260–72. https://doi.org/10.1016/S0140-6736(13)60687-X
3. Ene-Iordache B, Perico N, Bikbov B *et al*. Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study. *Lancet Glob Health* 2016;**4**:e307–19. https://doi.org/10.1016/S2214-109X(16)00071-1
4. Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med* 2013;**368**:2402–14. https://doi.org/10.1056/NEJMra1206793
5. Rodrigues JC, Haas M, Reich HN. IgA nephropathy. *Clin J Am Soc Nephrol* 2017;**12**:677–86. https://doi.org/10.2215/CJN.07420716
6. Webster AC, Nagler EV, Morton RL *et al*. Chronic kidney disease. *Lancet* 2017;**389**:1238–52. https://doi.org/10.1016/S0140-6736(16)32064-5
7. Chen L, Su W, Chen H *et al*. Proteomics for biomarker identification and clinical application in kidney disease. *Adv Clin Chem* 2018;**85**:91–113. https://doi.org/10.1016/bs.acc.2018.02.005
8. Zürbig P, Renfrow MB, Schiffer E *et al*. Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation. *Electrophoresis* 2006;**27**:2111–25. https://doi.org/10.1002/elps.200500827
9. Molin L, Seraglia R, Lapolla A *et al*. A comparison between MALDI-MS and CE-MS data for biomarker assessment in chronic kidney diseases. *J Proteomics* 2012;**75**:5888–97. https://doi.org/10.1016/j.jprot.2012.07.024

10. Zhao YY, Wu SP, Liu S *et al*. Ultra-performance liquid chromatography-mass spectrometry as a sensitive and powerful technology in lipidomic applications. *Chem Biol Interact* 2014;**220**:181–92. https://doi.org/10.1016/j.cbi.2014.06.029
11. Zhao YY, Cheng XL, Vaziri ND *et al*. UPLC-based metabonomic applications for discovering biomarkers of diseases in clinical chemistry. *Clin Biochem* 2014;**47**:16–26. https://doi.org/10.1016/j.clinbiochem.2014.07.019
12. Zhao YY, Cheng XL, Cui JH *et al*. Effect of ergosta-4,6,8(14),22-tetraen-3-one (ergone) on adenine-induced chronic renal failure rat: a serum metabonomic study based on ultra performance liquid chromatography/high-sensitivity mass spectrometry coupled with MassLynx i-FIT algorithm. *Clin Chim Acta* 2012;**413**:1438–45. https://doi.org/10.1016/j.cca.2012.06.005
13. Klein J, Bascands JL, Mischak H *et al*. The role of urinary peptidomics in kidney disease research. *Kidney Int* 2016;**89**:539–45. https://doi.org/10.1016/j.kint.2015.10.010
14. Mischak H, Delles C, Vlahou A *et al*. Proteomic biomarkers in kidney disease: issues in development and implementation. *Nat Rev Nephrol* 2015;**11**:221–32. https://doi.org/10.1038/nrneph.2014.247
15. Schanstra JP, Mischak H. Proteomic urinary biomarker approach in renal disease: from discovery to implementation. *Pediatr Nephrol* 2015;**30**:713–25. https://doi.org/10.1007/s00467-014-2790-y
16. Klein J, Lacroix C, Caubet C *et al*. Fetal urinary peptides to predict postnatal outcome of renal disease in fetuses with posterior urethral valves (PUV). *Sci Transl Med* 2013;**5**:198ra106. https://doi.org/10.1126/scitranslmed.3005807
17. Zimmerli LU, Schiffer E, Zürbig P *et al*. Urinary proteomic biomarkers in coronary artery disease. *Mol Cell Proteomics* 2008;**7**:290–8. https://doi.org/10.1074/mcp.M700394-MCP200
18. Weissinger EM, Metzger J, Dobbelstein C *et al*. Proteomic peptide profiling for preemptive diagnosis of acute graft-versus-host disease after allogeneic stem cell transplantation. *Leukemia* 2014;**28**:842–52. https://doi.org/10.1038/leu.2013.210
19. Metzger J, Negm AA, Plentz RR *et al*. Urine proteomic analysis differentiates cholangiocarcinoma from primary sclerosing cholangitis and other benign biliary disorders. *Gut* 2013;**62**:122–30. https://doi.org/10.1136/gutjnl-2012-302047
20. Yan L, Yi J, Huang C *et al*. Rapid detection of COVID-19 using MALDI-TOF-based serum peptidome profiling. *Anal Chem* 2021;**93**:4782–7. https://doi.org/10.1021/acs.analchem.0c04590
21. Nachtigall FM, Pereira A, Trofymchuk OS *et al*. Detection of SARS-CoV-2 in nasal swabs using MALDI-MS. *Nat Biotechnol* 2020;**38**:1168–73. https://doi.org/10.1038/s41587-020-0644-7
22. Muller M, Hummelink K, Hurkmans DP *et al*. A serum protein classifier identifying patients with advanced non-small cell lung cancer who derive clinical benefit from treatment with immune checkpoint inhibitors. *Clin Cancer Res* 2020;**26**:5188–97. https://doi.org/10.1158/1078-0432.CCR-20-0538
23. Weber JS, Sznol M, Sullivan RJ *et al*. A serum protein signature associated with outcome after anti-PD-1 therapy in metastatic melanoma. *Cancer Immunol Res* 2018;**6**:79–86. https://doi.org/10.1158/2326-6066.CIR-17-0412
24. Li LS, Liu ZH. Epidemiologic data of renal diseases from a single unit in China: analysis based on 13,519 renal biopsies. *Kidney Int* 2004;**66**:920–3. https://doi.org/10.1111/j.1523-1755.2004.00837.x
25. Haas M. Histologic subclassification of IgA nephropathy: a clinicopathologic study of 244 cases. *Am J Kidney Dis* 1997;**29**:829–42. https://doi.org/10.1016/S0272-6386(97)90456-X
26. Trimarchi H, Barratt J, Cattran DC *et al*. Oxford Classification of IgA nephropathy 2016: an update from the IgA Nephropathy Classification Working Group. *Kidney Int* 2017;**91**:1014–21. https://doi.org/10.1016/j.kint.2017.02.003
27. Lee HS, Lee MS, Lee SM *et al*. Histological grading of IgA nephropathy predicting renal outcome: revisiting H. S. Lee's glomerular grading system. *Nephrol Dial Transplant* 2005;**20**:342–8. https://doi.org/10.1093/ndt/gfh633
28. Cattran DC, Coppo R, Cook HT *et al*. The Oxford classification of IgA nephropathy: rationale, clinicopathological correlations, and classification. *Kidney Int* 2009;**76**:534–45. https://doi.org/10.1038/ki.2009.243
29. Roberts IS, Cook HT, Troyanov S *et al*. The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility. *Kidney Int* 2009;**76**:546–56. https://doi.org/10.1038/ki.2009.168
30. Hao Y, Zhao Y, Huang R *et al*. Analysis of the relationship between Oxford classification, IgM deposition and multiple indexes and the adverse prognosis of patients with primary IgA nephropathy and related risk factors. *Exp Ther Med* 2019;**17**:1234–9.
31. Lee SM, Rao VM, Franklin WA *et al*. IgA nephropathy: morphologic predictors of progressive renal disease. *Hum Pathol* 1982;**13**:314–22. https://doi.org/10.1016/S0046-8177(82)80221-9
32. Ai K, Pan J, Zhang P *et al*. Methyl-CpG-binding domain protein 2 contributes to renal fibrosis through promoting polarized M1 macrophages. *Cell Death Dis* 2022;**13**:125. https://doi.org/10.1038/s41419-022-04577-3
33. Ai K, Li X, Zhang P *et al*. Genetic or siRNA inhibition of MBD2 attenuates the UUO- and I/R-induced renal fibrosis via downregulation of EGR1. *Mol Ther Nucleic Acids* 2022;**28**:77–86. https://doi.org/10.1016/j.omtn.2022.02.015
34. Good DM, Zürbig P, Argilés A *et al*. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Mol Cell Proteomics* 2010;**9**:2424–37. https://doi.org/10.1074/mcp.M110.001917
35. Wu J, Wang N, Wang J *et al*. Identification of a uromodulin fragment for diagnosis of IgA nephropathy. *Rapid Commun Mass Spectrom* 2010;**24**:1971–8. https://doi.org/10.1002/rcm.4601

Received: 9.5.2023; Editorial decision: 22.8.2023