Characterization of metabolome-wide biomarkers associated with kidney function in the Chinese population

Hongquan Peng (✉ hpeng93170@gmail.com)

Kiang Wu Hospital

Xun Liu

The Third Affiliated Hospital of Sun Yat-sen University

Chi wa Ao leong

Kiang Wu Hospital

Tao Tao

Kiang Wu Hospital

Tsung yang Tsai

Kiang Wu Hospital

Ngai kam Leong

Kiang Wu Hospital

hao I Cheang

Kiang Wu Hospital

Zhi Liu

University of Macau

Peijia Liu

The Third Affiliated Hospital of Sun Yat-sen University

Abstract

Background Chronic kidney disease (CKD) is a global public health problem. Identifying sensitive filtration biomarkers is a key diagnostic value contributing to an understanding of CKD at the molecular level. A metabolomics study indicated a snapshot of the biochemical activity of the human body at a particular time in the progression of CKD. This metabolome-wide biomarker study verified whether plasma or serum metabolite profiles are significantly different in CKD at various stages and characterized potential markers to assess kidney function in the Chinese population.

Methods: An analysis of plasma and serum metabolites using ultrahigh-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) was performed on 198 participants (53 serum samples and 145 plasma samples) based on their measured GFR (by iohexol plasma clearance). Participants started recruiting in late 2019, and untargeted metabolomics assays (N ¼ 214) were conducted at the Calibra-Metabolon Joint Laboratory (Hangzhou, China) using Metabolon's HD4 Discovery untargeted metabolomics platform in early 2021.

Results A large number of metabolomics related to the mGFR were selected as the top 30 metabolites by the random forest method, and we found 15 amino acids, 8 nucleotides, and 2 carbohydrates strongly related to kidney function in the combined group (serum and plasma). Thirteen amino acids, 9 nucleotides, and 3 carbohydrates were identified in the plasma group, while 13 amino acids, 7 nucleotides, and 3 carbohydrates were found in the serum group. We observed that 10 of the top 15 ranked metabolites were concordant between the plasma and serum groups. Major differences in metabolite profiles with increasing stage of CKD were observed, including altered tryptophan metabolism and pyrimidine metabolism.

Conclusions Global metabolite profiling of plasma/serum uncovered potential biomarkers of stages of CKD. In addition, these novel biomarkers provide insight into possible pathophysiologic processes that may contribute to the progression of CKD and a higher risk of comorbidities and mortality.

Background

Chronic kidney disease (CKD) is a growing burden on people worldwide and has become a major public health concern affecting approximately 10% of the population and substantially increases the risk of cardiovascular morbidity and mortality [1–3]. Biomarkers promote clinicians to offer more appropriate diagnosis and treatments. To assess kidney function, creatinine is a well-established biomarker [4]. However, creatinine is not a sensitive marker and does not warrant early detection[5, 6]. It is well known that blood metabolite concentrations are influenced by kidney function, and a few metabolites are used for its estimation—e.g., the above creatinine is used to estimate the glomerular filtration rate (GFR). However, creatinine has some limitations, rising only after almost half of kidney function loss and a dependence on race, reflecting underlying differences in muscle mass. Assessment of GFR is vitally

important to the estimation of renal function in clinical practice and public health, and measured GFR (mGFR) remains the reference standard ("gold standard") [8, 9].

Using a panel of filtration markers can improve precision, reduce errors caused by variation in the non-GFR determinants of each marker and decrease the need to use race and clinical characteristics as surrogates for the non-GFR determinants [10, 11]. Our study aims to track a few new metabolites biomarkers to optimize the development of GFR and identify novel glomerular filtration−related markers that are better than or equal to creatinine.

Most metabolites are eliminated by the kidney. Changes in serum or plasma metabolite concentration levels may be caused by impaired renal function and can be used to estimate GFR. In this study, GFR was measured with the plasma clearance rate of iohexol[12−14], and at the same time, estimated GFR (eGFR) on the basis of creatinine and cystatin C levels was also assessed [7]. Thus, our study yields a wide-range list of metabolites associated with mGFR and highlights potential novel filtration markers that may aid in improving the estimation of GFR.

Despite the high incidence and increasing prevalence of CKD, the underlying pathophysiologic mechanisms have not yet been elucidated. Therefore, the identification of novel biomarkers of kidney function is still clinically useful, as evidenced by the recent addition of a second marker to estimate GFR, cystatin C, which accurately improves estimation of GFR and predicts future risk of end-stage renal disease (ESRD) and death [15].

Blood metabolite levels are altered in the progression of CKD, promoting the investigation of the metabolome of interest in nephrology. Previously, a series of nontargeted or targeted metabolomics studies were mostly conducted in a non-Chinese population-based study, and a few new biomarkers have been ascertained [16−20]. The goal of our study was to identify and replicate novel and known metabolites that reproducibly associate with mGFR and try to reveal the characterization of biomarkers, to obtain additional insights into the pathophysiology of CKD in the Chinese population and to reveal the discrepancy among different races.

Methods

Study Participants

A total of 198 participants (96 females, 48.5%), including 10 healthy volunteers and 188 CKD patients with varying degrees of renal dysfunction, were enrolled in this study. Mean age was 58.2 ± 18.5years (range 18−96years). The mean body mass index was 24.2 ± 4.2 kg/m^2 (range 15.0-48.6 kg/m^2). The mean serum creatinine was 1.82 ± 2.01 g/l (range 0.42− 10.79 mg/l). Cystatin C 1.17 ± 1.16mg/L(range 0.62-5.82mg/L). Plasma samples were obtained from Kiang Wu Hospital, Macao, and serum samples were obtained from the Third Affiliated Hospital of Sun Yat-sen University, Guang Zhou, China. All samples were stored at -80°C for this study. A total of 198 participants (53 serum samples and 145

plasma samples) based on their measured GFR (by iohexol plasma clearance) were selected for untargeted metabolomics assays (N ¼ 214) conducted at the Calibra-Metabolon Joint Laboratory (Hangzhou, China) using Metabolon's HD4 Discovery untargeted metabolomics platform in early 2021. Informed consent was obtained from all participants. This study was approved by the local ethics committee. All volunteers were informed and signed the consent form.

Metabolomic analysis

Sample Accessioning

Each sample received was accessioned into the mLIMS system and was assigned by the LIMS a unique identifier that was associated with the original source identifier only. This identifier was used to track all sample handling, tasks, results, etc. The samples (and all derived aliquots) were tracked by the LIMS system. All portions of any sample were automatically assigned their own unique identifiers by the LIMS when a new task was created; the relationship of these samples was also tracked. All samples were maintained at -80°C until processed.

Sample Preparation

Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove proteins, dissociate small molecules bound to proteins or trapped in the precipitated protein matrix, and recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation.

UPLC-MS/MS methods

The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

Data Extraction and Compound Identification

Raw data were extracted, peak-identified and QC processed using Discovery HD4 hardware and software. These systems are built on a web-service platform utilizing Microsoft's. NET technologies run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities.

Discovery HD4

Discovery HD4 maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum.

Statistical Methods and Terminology

Values are expressed as mean ± standard deviation (SD). Group comparisons were carried out with the chi-square or Mann-Whitney test. Mean values and proportions were compared using one-way analysis of variance and chi-square tests, respectively. A significance level of $p < 0.05$ was utilized in all tests, and SPSS-IBM22 for Mac was used for these analyses.

Random Forest

Random forest is a supervised classification technique based on an ensemble of decision trees[20]. For a given decision tree, a random subset of the data with identifying true class information is selected to build the tree ("bootstrap sample" or "training set"), and then the remaining data, the "out-of-bag" (OOB) variables, are passed down the tree to obtain a class prediction for each sample. This method is unbiased since the prediction for each sample is based on trees built from a subset of samples that do not include that sample. To determine which variables (biochemical) make the largest contribution to the classification, a "variable importance" measure is computed. We use the "Mean Decrease Accuracy" (MDA) as this metric. The MDA is determined by randomly permuting a variable, running the observed values through the trees, and then reassessing the prediction accuracy.

Result

We applied untargeted high-performance liquid chromatography tandem mass spectrometry (HPLC-MS) to determine metabolomics profiles in 198 participants (53 serum samples and 145 plasma samples), and we measured patients' mGFR by iohexol plasma clearance. The study population are grouped by mGFR(Table 1). Both serum and plasma were analyzed separately and combined.

A total of 198 participants (102 females, 51.5%), including 10 healthy volunteers and 188 CKD patients with varying degrees of renal dysfunction, were enrolled in this study. Table 2 shows the clinical characteristics of the population samples studied. There were no group differences in sex distribution, body mass index (BMI), or blood pressure; however, the group with impaired renal function had higher comorbidities (hypertension, diabetes, coronary artery disease prevalence), and the age was older than the group with normal kidney function.

Global Metabolite Determination and Significantly Altered Biochemicals

The present dataset comprises a total of 1094 compounds of known biochemical properties. A subset of these metabolites was identified, with significant differences in accordance with CKD stage progression ($p < 0.05$); an additional set of metabolites was identified that approached significance ($0.05 < p < 0.10$) (Tables 3–5). Analysis by two-way ANOVA identified biochemicals exhibiting significant interaction and main effects for experimental parameters of disease status and sample type. An estimate of the false discovery rate (q-value) is calculated to take into account the multiple comparisons that normally occur in metabolomic-based studies.

High Level of Metabolite Overview

Principal component analysis (PCA) is a mathematical dimension reduction procedure that allows differences across a large set of variables to be represented as a smaller set of variables. PCA permits visualization of how individuals, within a group, cluster with respect to their data-compressed principal components. As such, this tool aids in determining whether samples segregate based on differences in their overall metabolite signature. As shown in (Supplement Figs. 5–8). There is complete separation between the samples.

Identification of TOP ranking metabolite Changes

Analysis is an unbiased and supervised classification technique based on an ensemble of many decision trees. In addition to producing a metric of predictive accuracy, random forest analysis also produces an associated list of biochemical rankings in order of their importance to the classification scheme. In this study, random forest analysis was used to identify metabolites that differentiated samples from the four groups. A predictive accuracy of 80.8% was obtained in the combined plasma + serum data set (Fig. 3), 82.1% for the plasma data set (Fig. 1), and 84.9% for the serum data set (Fig. 2), compared to 25% by random chance alone. The random forest analysis performed much better than that expected by random chance, suggesting that there are significant metabolic differences that can be used to discriminate samples between the four groups, with metabolites in the amino acid and nucleotide super pathways being of most importance for the three models. A list of the top 30 biochemicals that contributed to the separation of the groups.

A large number of metabolomics related to mGFR were identified, and we selected the top 30 metabolites that were bound up with mGFR by the random forest method. As shown in Fig. 3, we found that 15 amino acids, 8 nucleotides, and 2 carbohydrates were strongly related to kidney function in the combined group (serum and plasma). Thirteen amino acids, 9 nucleotides, and 3 carbohydrates were identified in the plasma group (Fig. 1), while 13 amino acids, 7 nucleotides, and 3 carbohydrates were found in the serum group (Fig. 2). We observed that 10 of the top 15 ranked metabolites were concordant between the plasma and serum groups.

Panel of markers related to renal function

In addition to the classical clinical markers of kidney dysfunction, other small molecule metabolic markers of kidney function have been studied. In this study, in addition to creatinine and urea, a few potential biomarkers were identified, such as pseudouridine. Specifically, a novel negative biomarker of kidney disease may be 1,5-anhydroglucitol (1,5-AG). As one function of the kidney is to remove toxins and excess metabolites from the body, kidney dysfunction can result in the buildup of many endogenous and exogenous metabolites. One method the body uses to facilitate elimination is phase II metabolism, which generally increases hydrophilicity and urinary elimination. Large increases in sulfated and glucuronidated metabolites, both endogenous and exogenous, were observed in this study. This is most notable with metabolites in the aromatic amino acid, benzoate, food components, and chemical subpathways (Fig. 4).

Discussion

Kidneys are the organs filtering waste products from blood, and creating urine and kidney function is frequently measured as the glomerular filtration rate (GFR). Recently, advances in mass methodology have allowed comprehensive studies of metabolomics and its relationship with kidney function [21–25]. Metabolomics studies can identify and quantify all metabolites present in a given sample, covering hundreds to thousands of metabolites. Thus, data such as the heatmap and plots, as well as principal component analysis and random forest analysis, are performed on the entirety of the data set, the plasma-only data set, and the serum-only data set. Major differences in metabolite profiles in the various severities of CKD were observed. A large number of biochemicals increased with the progression of CKD; on the other hand, a small number of biochemicals were reduced. These differences may reveal stage-specific biomarkers of CKD.

In this study, we found a number of metabolites associated with mGFR, and we selected the top 30 metabolites that were strongly related to mGFR by the random forest method,As shown in(Fig. 1–3), Here, we find 10 of the top ranking 15 metabolites substances were concordant between plasm group and serum group .This indicates that the results of serum samples and plasma samples for the determination of metabolomics may be generally consistent;however, still need more studies to verify.

Sekula et al. [26] reported 56 metabolites that replicated as associated with eGFRcr, including 6 metabolites that were consistently strongly correlated with eGFRcr (pseudouridine, c-mannosyltryptophan, N-acetylalanine, erythronate, myo-inositol and N-acetylcarnosine). However, Coresh J et al [16] reported a few candidate novel filtration markers of metabolites in a panel including pseudouridine, acetylthreonine, myoinositol, phenylacetylglutamine and tryptophan and high correlation with mGFR (including all of the above metabolites except N-acetylcarnosine).

In our research, we found that C-glycosyltryptophan (also known as C-mannosyltryptophan) pseudouridine, N-acetylalanine, erythronate, myo-inositol and even N-acetylcarnosine were highly correlated with mGFR, except acetylthreonine. N-acetylcarnosine and pseudouridine showed markedly increasing levels with increased nephropathy in our study, consistent with the results reported in Sekula et al. [26]. Both metabolites can be indicators of protein turnover as N-acetylation of amino acids.

Pseudouridine is a derivative of uridine and is a modified nucleoside found in RNA. Interestingly, pseudouridine may be an ideal biomarker ranking in the top 5 among the above studies, meaning it is a stable indicator and nondependent on race.

Additionally, both N6-carbamoylthreonyladenosine and hydroxyasparagine were unique in this study. Hydroxyasparagine, known as β-hydroxyasparagine (beta-hydroxyasparagine), is associated with mGFR and CKD and is a modified asparagine amino acid. However, we know little about this metabolite. It appears in posttranslational modifications of EGF-like domains that can occur in humans and others Eukaryotes. The modified amino acid residue is found in fibrillin-1[27]. C-glycosyltryptophan was identified to be associated with mGFR and CKD, as well as with prospective endpoints of eGFR decline, incident CKD and ESRD[26].

Specifically, a potential negative biomarker of kidney disease is 1,5-anhydroglucitol (1,5-AG). 1,5-AG. In the kidney, 1,5-AG is filtered by the glomerulus, and the majority is reabsorbed in the proximal tubule back to the blood. Glucose, which trended higher in the moderate and severe nephropathy groups than in the normal kidney function group, is a competitive inhibitor of this reabsorption. Thus, if glucose rises, more 1,5-AG is excreted in the urine, lowering blood levels. The level of 1,5-AG was decreased in severe nephropathy compared to the three other groups and in moderate nephropathy compared to mild nephropathy. In a recent study, 1,5-AG may also have prognostic value in relation to cardiovascular events in patients with coronary heart disease[28].

Creatine kinase catalyzes both the transfer of a high-energy phosphate from ATP to creatine and the regeneration of ATP from creatine phosphate and ADP. In solution, creatine slowly and spontaneously cyclizes to creatinine, which is eliminated in the urine and can be used as a marker of kidney function. Creatinine levels increase with increasing nephropathy. The urea cycle is necessary for organisms to detoxify and safely eliminate waste ammonia generated from the catabolism of amino acids. Urea also increases with increasing nephropathy.

In addition to removing waste from the organism, another function of the kidney is to regulate the electrolyte and fluid volume for the body. Thus, with nephropathy, derangements in molecules necessary for osmotic regulation are expected. As seen in Fig. 4, increases in small molecules involved in osmotic regulation, such as inositol metabolism, myo-inositol and chiro-inositol, erythronate, and trimethylamine N-oxide (TMAO), were observed with increasing nephropathy. Other metabolites, such as 3-indoxyl sulfate, phenylacetylglutamine, 1-methylguanidine, and guanidinosuccinate, have been shown to be uremic toxins or metabolites that accumulate during uremia. All three of these metabolites show increasing concentrations with increased nephropathy. Guanidine (G), 1-methylguanidine (MG), and 1,1-dimethylguanidine (DMG) have long been implicated as uremic "toxins." N-acetylalanine-aminopeptidase is a new enzyme found in human erythrocytes. Enzymatic activity has not been found in the cytosolic compartment of highly purified human leucocytes [29]. Its physiological function in erythrocytes is still unknown.

Tryptophan, an essential aromatic amino acid, can feed into a number of processes, including production of the neurotransmitter serotonin, the anti-inflammatory metabolite kynurenine, and downstream of kynurenine, NAD + production. An enzyme that catalyzes the first step in the conversion of tryptophan to kynurenine is tryptophan 2,3-dioxygenase (TDO), which is primarily expressed in the liver. Additionally, changes in tryptophan metabolites can also reflect increasing inflammation: indoleamine 2,3-dioxygenase (IDO), which also catalyzes the conversion of tryptophan to kynurenine, is activated by proinflammatory cytokines (e.g., IFN-γ and TNF-α). While kynurenine is produced in this way by an inflammatory process, it has an anti-inflammatory function, serving as a brake on the immune response. Multiple metabolites in the tryptophan metabolic pathway, including kynurenine, kynurenate, anthranilate, and xanthurenate, were increased in the severe and moderate nephropathy groups, with some metabolites increased in the mild group compared to the normal group, indicative of increased inflammation (data not shown). C-glycosyltryptophan results from a posttranslational modification of tryptophan by linking a sugar by a carbon−carbon bond[30−31]. Certain posttranslational modifications, such as carbamylation, have been associated with different chronic conditions, including CKD[32]. Studies in humans showed consistently increased levels of C-glycosyltryptophan in people with decreased eGFR[33−35]. Moreover, studies of humans and animal models suggested C-glycosyltryptophan to be a good biomarker of kidney function with more favorable properties than serum creatinine[35, 36].

In our study, numerous xenobiotics and metabolized xenobiotics and other metaboletes were observed. Some of note are differences in therapeutic drugs. These drugs, which can have significant systemic effects, are a confounding factor in the data.

Conclusion

The differences in the blood plasma and serum metabolome of human subjects with increasing levels of nephropathy, as well as healthy control subjects, were examined. Overall, the changes in the metabolome between the four groups were significant. In the PCA, there was group separation of the plasma and serum samples with orthogonal separation due to the experimental group. Our study identified 6 novel and potential metabolites that reproducibly strongly associate with mGFR, including pseudouridine, C-glycosyltryptophan, erythronate, N-acetylalanine, myo-inositol, and N-acetylcarnosine, but not acetylthreonine. However, pseudouridine may be an ideal biomarker that is nondependent on race. In addition, both N6-carbamoylthreonyladenosine and hydroxyasparagine are strongly associated with mGFR and CKD and are unique in this study. Specifically, a potential negative biomarker of kidney disease may be 1,5-anhydroglucitol (1,5-AG). In this study, serum samples and plasma samples for the determination of metabolomics may be generally consistent, although slightly different; however, more samples are needed to confirm this hypothesis. Future studies will utilize the potential 3−5 novel biomarkers in estimating the glomerular filtration rate without race input.

Abbreviations

CKD: Chronic kidney disease; ESRD: End-Stage Renal Disease; GFR: Glomerular filtration rate; mGFR: Measured glomerular filtration rate; UPLC-MS/MS: Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy; 1,5-AG: 1,5-anhydroglucitol; eGFR: Estimated glomerular filtration rate, PCA: Principal Components Analysis; LIMS: Laboratory Information Management System; IDO: Indoleamine 2,3-dioxygenase; TDO: Tryptophan 2,3-dioxygenase; TMAO: Trimethylamine N-oxide; ATP: Adenosine triphosphate; ADP: Adenosine diphosphate.

Declarations

Ethics approval and consent to participate

The protocol was approved by the institutional review board of the Kiang.

Wu Hospital (KWH 2018–001). All volunteers were informed and signed the

consent form. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

All authors have read the manuscript and consent for publications.

Availability of data and materials

The datasets used and analyzed during the current study are available from

the corresponding author on reasonable request.

Competing interests

The authors have declared that no competing interests exist.

Funding

Author Contributions

HQP and XL conceived the study. HQP CWAI, TT, and TYT collected the data and carried out the experiment. HQP and ZL performed the data analyses and verified the analytical methods. Both HQP and XL contributed to the final version of the manuscript. All authors have read and approved the manuscript.

Acknowledgment

References

1. Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, Chen M, He Q, Liao Y, Yu X, Chen N, Zhang JE, Hu Z, Liu F, Hong D, Ma L, Liu H, Zhou X, Chen J, Pan L, Chen W, Wang W, Li X, Wang H. Prevalence of chronic kidney disease in China: a cross-sectional survey. Lancet. 2012, 379(9818):815–22.

2. Tonelli M, Wiebe N, Culleton B, House A, Rabbat C, Fok M, McAlister F, Gary AX.Chronic kidney disease and mortality risk: a systematic review. J. Am. Soc. Nephrol.2006,17(7):2034–2047.

3. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N. Engl. J. Med. 2004,351(13):1296–1305.

4. Levin A, Stevens PE, Bilous RW et al. KDIGO 2012 clinical practice guidelinefor the evaluation and management of chronic kidney disease. KidneyInt. Suppl. 2013, 3:1–150.

5. Levey AS, Inker LA, Coresh J. GFR estimation: from physiology to public health. Am. J. Kidney Dis. 2014, 63(5):820–34.

6. Jinxia Chen, Hua Tang, Hui Huang, LinshengLv, Yanni Wang, Xun Liu, Tanqi Lou. Development and Validation of New Glomerular Filtration Rate Predicting Models for Chinese Patients with Type 2 Diabetes. J. Transl. Med. 2015, 13:317(2)

7. Levey AS, Coresh J. Chronic kidney disease. Lancet. 2012, 379(9811):165–80.

8. Brown SC, O'Reilly PH. Iohexol clearance for the determination of glomerularfiltration rate in clinical practice: evidence for a new gold standard. J. Urol.1991, 146(3):675–9.

9. Stevens LA, Levey AS. Measured GFR as a confirmatory test for estimatedGFR. J Am. Soc.Nephrol. 2009, 20(11):2305–13.

10. Levey AS, Coresh J, Tighiouart H, Greene T, Inker LA. Measured and estimated glomerular filtration rate: current status and future directions.Nat. Rev.Nephrol. 2020, 16(1):51–64.

11. Inker LA, Levey AS, Coresh J. Estimated glomerular filtration rate from a panel of filtration markers-hope for increased accuracy beyond measured glomerular filtration rate? Adv. Chronic Kidney Dis. 2018, 25, 67–75.

12. Soveri I, Berg UB, Björk J, et al. Measuring GFR: a systematic review. Am. J.Kidney Dis. 2014, 64(3):411–24.

13. Schwartz GJ, Abraham AG, Furth SL, Warady BA, Muñoz A. Optimizingiohexol plasma disappearance curves to measure the glomerular filtration rate in children with chronic kidney disease. Kidney Int. 2010, 77(1):65–71

14. Zhang Y, Sui Z, Yu Z, Li TF, Feng WY, Zuo L. Accuracy of iohexol plasma clearance for GFR-determination: a comparison between single and dual sampling.BMC Nephrol. 2018, 19(1):174.

15. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, Kusek JW, Manzi J, Van Lente F, Zhang YL, Coresh J, Levey AS; CKD-EPI Investigators. Estimating glomerular filtration rate from serum creatinine and cystatin C. N. Engl. J. Med. 2012, 367(1):20–9.

16. Coresh J, Inker LA, Sang Y, et al. Metabolomic profiling to improve glomerular filtration rate estimation: a proof-of-concept study.Nephrol Dial Transplant. 2019, 34(5):825–833.

17. Shah VO, Townsend RR, Feldman HI, Pappan KL, Kensicki E, Vander Jagt DL. Plasma metabolomic profiles in different stages of CKD.Clin. J. Am. Soc.Nephrol. 2013, 8(3):363–370.

18. Rhee EP, Clish CB, Wenger J, et al. Metabolomics of Chronic Kidney Disease Progression: A Case-Control Analysis in the Chronic Renal Insufficiency Cohort Study.Am. J.Nephrol. 2016, 43(5):366–374.

19. Benito S, Sánchez-Ortega A, Unceta N, et al. Plasma biomarker discovery for early chronic kidney disease diagnosis based on chemometric approaches using LC-QTOF targeted metabolomics data.J. Pharm. Biomed Anal. 2018, 149:46–56.

20. Breiman, L. Random Forests.Machine Learning. 2001, 45:5–32.

21. Beckonert O, Keun HC, Ebbels TMD, Bundy J, Holmes E, Lindon JC Nicholson JK. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts.Nat. Protoc. 2007, 2:2692–2703.

22. Evans, AM, Bridgewater B, Liu Q, Mitchell M. High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics.Metabolomics. 2014, 4:2:10000132.

23. JohnsonCH, Ivanisevic J,Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms.Nat. Rev. Mol. Cell. Biol. 2016, 17:451–459.

24. Hocher B, Adamski, J. Metabolomics for clinical use and research in chronic kidney disease.Nat. Rev. Nephrol. 2017, 13(5):269–284.

25. Kottgen A, Raffler J,Sekula P,Kastenmuller G. Genome-wide association studies of metabolite concentrations (mGWAS): relevance for nephrology.Semin. Nephrol. 2018, 38(2):151–174.

26. Sekula P, Goek ON, Quaye L, et al. A Metabolome-Wide Association Study of Kidney Function and Disease in the General Population.J. Am. Soc.Nephrol. 2016, 27(4):1175–88.

27. Glanville RW, Qian RQ, McClure DW, Maslen CL. Calcium binding, hydroxylation, and glycosylation of the precursor epidermal growth factor-like domains of fibrillin-1, the Marfan gene protein. J. Biol. Chem. 1994, 269(43):26630–4.

28. Ignatova YS, Karetnikova VN, Horlampenko AA, Gruzdeva OV, Dyleva YA, Barbarash OL. The marker of adverse prognosis 1.5-anhydroglucitol in patients with coronary heart disease in the long-term period after planned myocardial revascularization.TerArkh. 2019, 91(4):48–52.

29. Schönberger OL, Tschesche H. N-Acetylalanine aminopeptidase, a new enzyme from human erythrocytes.Hoppe Seylers Z Physiol Chem. 1981, 362(7):865–873.

30. Furmanek A,Hofsteenge J. Protein C-mannosylation: facts and questions.ActaBiochim. Pol. 2000, 47:781–789.

31. Manabe S, Marui Y, Ito Y. Total synthesis of mannosyl tryptophan and its derivatives.Chemistry-A European Jouranl. 2003, 9(6):1435–1447.

32. Gillery P, Jaisson S. Post-translational modification derived products (PTMDPs): toxins in chronic diseases?.Clin. Chem. Lab. Med. 2014, 52(1):33–38.

33. Shah VO, Townsend RR, Feldman HI, Pappan KL, Kensicki E, Vander Jagt DL. Plasma metabolomic profiles in different stages of CKD. Clin. J. Am. Soc.Nephrol. 2013, 8(3):363–370.

34. Niewczas MA, Sirich TL, Mathew AV, et al. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. Kidney Int. 2014, 85(5):1214–24.

35. Sekula P, Dettmer K, Vogl FC, et al. From Discovery to Translation: Characterization of C-Mannosyltryptophan and Pseudouridine as Markers of Kidney Function. Sci Rep. 2017, 7(1):17400.

36. Takahira R, Yonemura K, Yonekawa O, et al. Tryptophan glycoconjugate as a novel marker of renal function.Am. J. Med. 2001, 110(3):192–197.

Tables

Due to technical limitations, the tables are only available as a download in the supplemental files section.

Figures

The 30 top ranking biochemicals in the importance plot suggest key differences in

- Amino Acids (13)
- Nucleotides (9)
- Carbohydrates (3)
- Xenobiotics (3)

Random Forest Confusion Matrix

Predicted Group

	Normal	Mild	Moder.	Severe	Class Error
Normal	36	1	0	0	0.027
Mild	9	26	4	0	0.333
Moder.	2	6	37	2	0.213
Severe	0	0	2	20	0.091
Predictive accuracy = 82.1%					

(Actual Group — row axis)

Biochemical Importance Plot

Increasing Importance to Group Separation →

mean-decrease-accuracy

Biochemicals (top to bottom):
hydroxyasparagine**, N-acetylneuraminate, C-glycosyltryptophan, pseudouridine, N,N-dimethyl-pro-pro, erythronate*, 3-(3-amino-3-carboxypropyl)uridine*, N2,N2-dimethylguanosine, gulonate*, erythritol, O-sulfo-L-tyrosine, N-acetyltaurine, dimethylarginine (SDMA + ADMA), N-acetylserine, N6-carbamoylthreonyladenosine, N6-succinyladenosine, N1-methylinosine, 5-(galactosylhydroxy)-L-lysine, 1-methylhistidine, 4-acetamidobutanoate, 5,6-dihydrouridine, S-adenosylhomocysteine (SAH), N,N,N-trimethyl-alanylproline betaine (TMAP), N4-acetylcytidine, 5-methylthioribose**, vanillactate, orotidine, N-acetylalanine, arabonate/xylonate, gluconate

Super Pathway
- Amino Acid
- Carbohydrate
- Cofactors and Vitamins
- Nucleotide
- Peptide
- Xenobiotics

Figure 1

Random forest analysis of plasma from subjects with normal kidney function, mild nephropathy, moderate nephropathy, and severe nephropathy. Random forest analysis of plasma from subjects with normal kidney function, mild nephropathy, moderate nephropathy, and severe nephropathy.

The 30 top ranking biochemicals in the importance plot suggest key differences in

- Amino Acids (13)
- Nucleotides (7)
- Carbohydrates (3)
- Lipids (3)
- Cofactors and Vitamins (2)

Random Forest Confusion Matrix

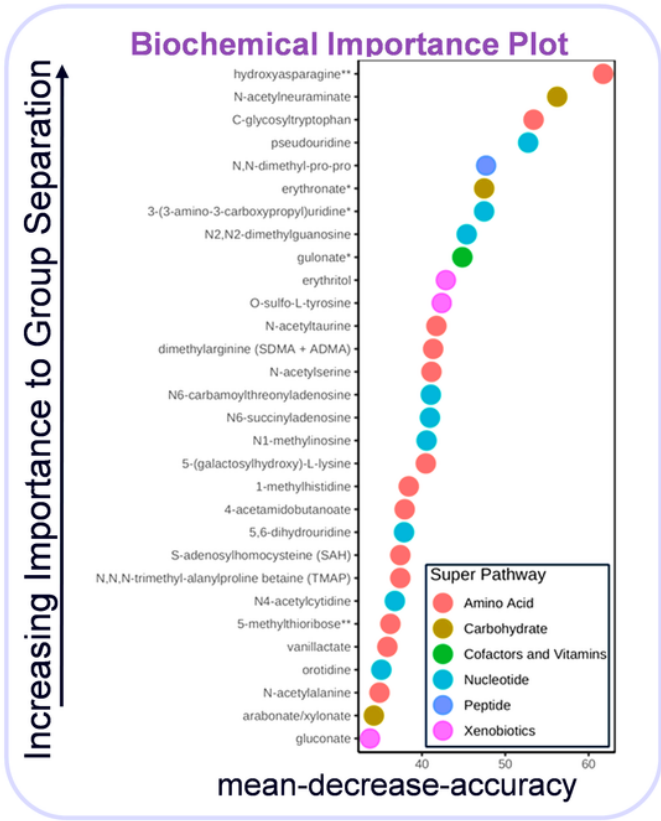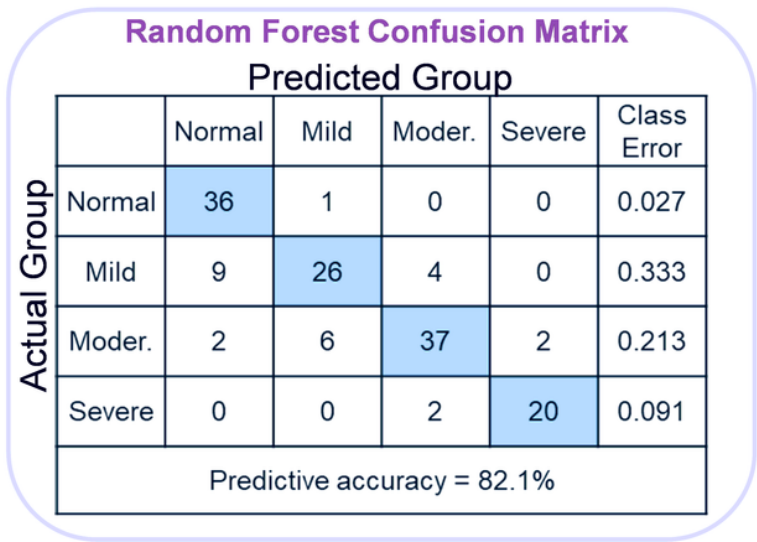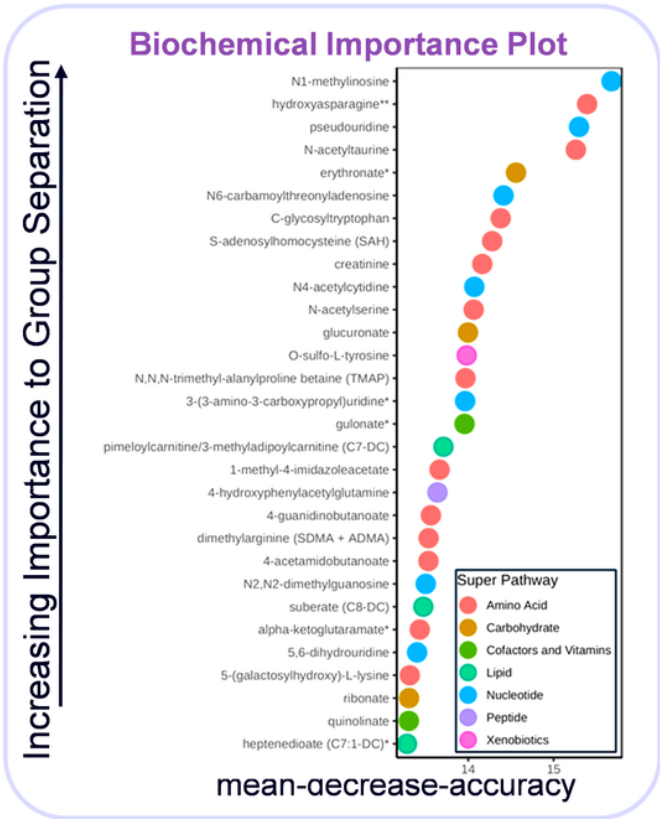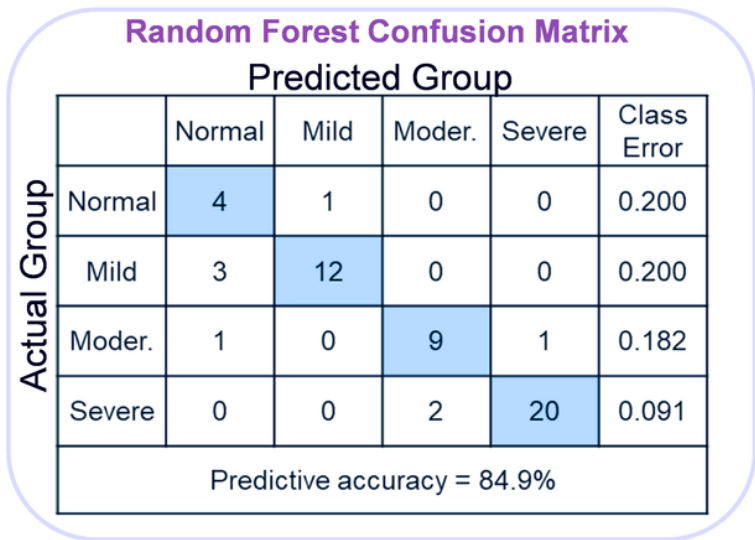|		Predicted Group			
		Normal	Mild	Moder.	Severe	Class Error
Actual Group	Normal	4	1	0	0	0.200
	Mild	3	12	0	0	0.200
	Moder.	1	0	9	1	0.182
	Severe	0	0	2	20	0.091
	Predictive accuracy = 84.9%					

Biochemical Importance Plot

Figure 2

Random forest analysis of serum from subjects with normal kidney function, mild nephropathy, moderate nephropathy, and severe nephropathy. Random forest analysis of serum from subjects with normal kidney function, mild nephropathy, moderate nephropathy, and severe nephropathy

The 30 top ranking biochemicals in the importance plot suggest key differences in

- Amino Acids (15)
- Nucleotides (8)
- Carbohydrates (2)
- Cofactors and Vitamins (2)
- Xenobiotics (2)

Random Forest Confusion Matrix

Predicted Group

Actual Group	Normal	Mild	Moder.	Severe	Class Error
Normal	37	5	0	0	0.119
Mild	11	38	5	0	0.296
Moder.	2	7	44	5	0.241
Severe	0	0	3	41	0.068
Predictive accuracy = 80.8%					

Biochemical Importance Plot

Increasing Importance to Group Separation

hydroxyasparagine**
C-glycosyltryptophan
pseudouridine
N-acetylneuraminate
3-(3-amino-3-carboxypropyl)uridine*
N,N-dimethyl-pro-pro
N2,N2-dimethylguanosine
4-acetamidobutanoate
erythronate*
N6-carbamoylthreonyladenosine
gulonate*
N1-methylinosine
O-sulfo-L-tyrosine
dimethylarginine (SDMA + ADMA)
5,6-dihydrouridine
S-adenosylhomocysteine (SAH)
N-acetylserine
erythritol
N-acetyltaurine
N,N,N-trimethyl-alanylproline betaine (TMAP)
hydroxy-N6,N6,N6-trimethyllysine*
N6-succinyladenosine
5-(galactosylhydroxy)-L-lysine
vanillactate
1-ribosyl-imidazoleacetate*
N-acetylalanine
2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA)*
orotidine
N-acetylhomocitrulline
quinolinate

Super Pathway
- Amino Acid
- Carbohydrate
- Cofactors and Vitamins
- Nucleotide
- Peptide
- Xenobiotics
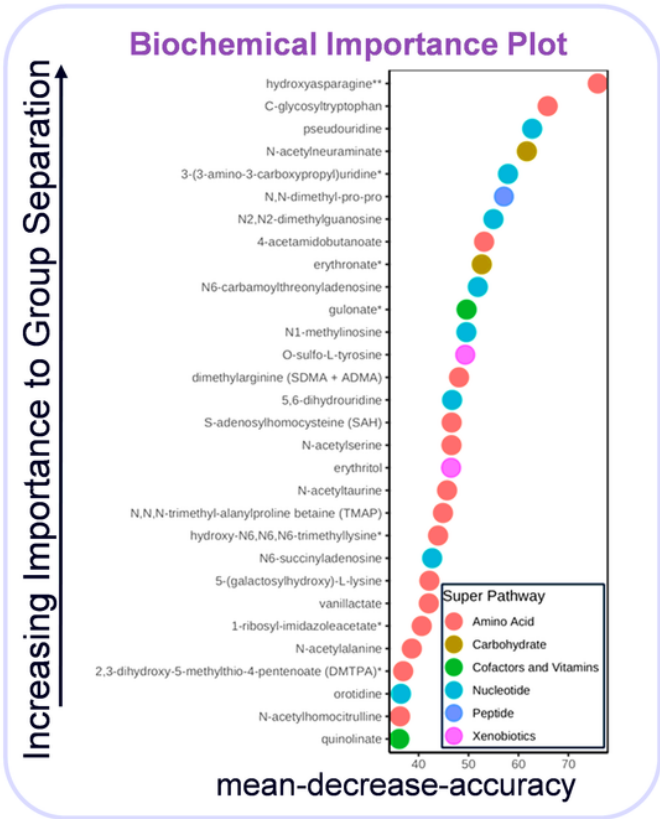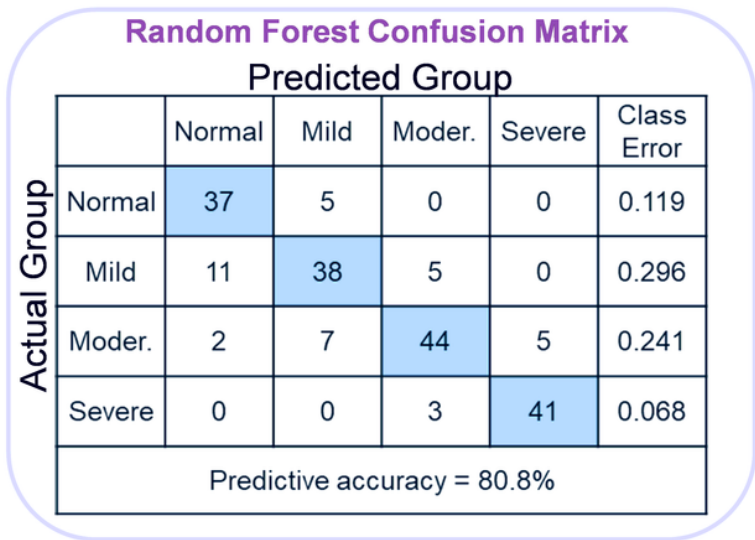
mean-decrease-accuracy

Figure 3

Random forest analysis of plasma and serum from subjects with normal kidney function, mild nephropathy, moderate nephropathy, and severe nephropathy. Random forest analysis of plasma and serum from subjects with normal kidney function, mild nephropathy, moderate nephropathy, and severe nephropathy.

| Sub Pathway	Biochemical Name	Mild Normal	Moder. Normal	Severe Normal	Moder. Mild	Severe Mild	Severe Moder.
Glycolysis & Gluconeogenesis	glucose	1.04	1.17	1.19	1.12	1.14	1.02
	1,5-anhydroglucitol (1,5-AG)	1.14	0.90	0.52	0.79	0.46	0.58
Inositol Metabolism	myo-inositol	1.18	1.74	5.42	1.47	4.59	3.12
	chiro-inositol	1.40	2.83	10.77	2.02	7.68	3.80
Aminosugar Metabolism	erythronate*	1.33	2.21	6.18	1.66	4.66	2.80
Food Component/Plant	erythritol	4.99	2.62	16.49	0.52	3.31	6.30
Phospholipid Metabolism	trimethylamine N-oxide	1.70	2.13	6.11	1.25	3.59	2.87
Alanine & Aspartate Metabolism	N-acetylalanine	1.26	1.81	3.08	1.44	2.44	1.70
Tryptophan Metabolism	3-indoxyl sulfate	1.23	2.13	7.56	1.72	6.13	3.56
Histidine Metabolism	N-acetylcarnosine	1.14	1.64	2.98	1.43	2.61	1.82
Acetylated Peptides	phenylacetylglutamine	1.97	3.79	18.02	1.92	9.16	4.76
Urea cycle; Arginine & Proline Metabolism	urea	1.17	1.90	6.49	1.62	5.53	3.42
Creatine Metabolism	creatinine	1.23	1.76	2.95	1.43	2.39	1.67
Guanidino & Acetamido Metabolism	1-methylguanidine	1.11	2.82	81.30	2.54	73.22	28.86
	guanidinosuccinate	1.35	3.61	47.06	2.67	34.79	13.04
Pyrimidine Metabolism, Uracil containing	pseudouridine	1.29	2.16	5.82	1.68	4.51	2.69
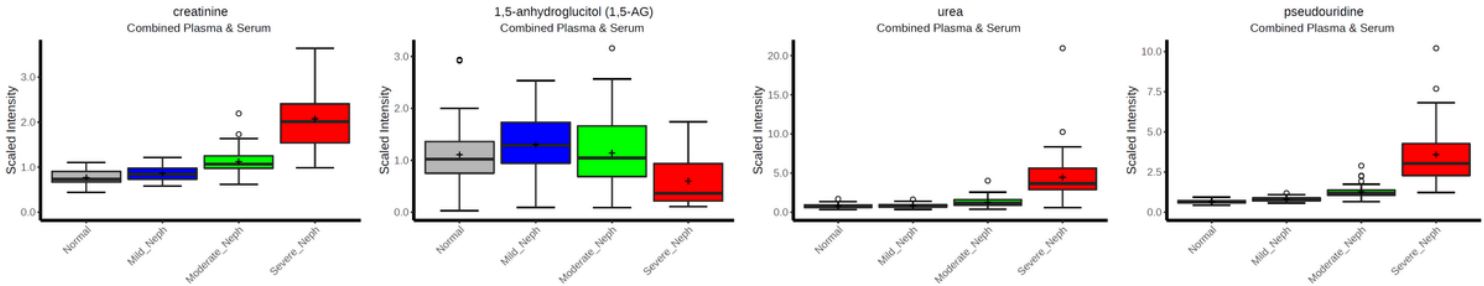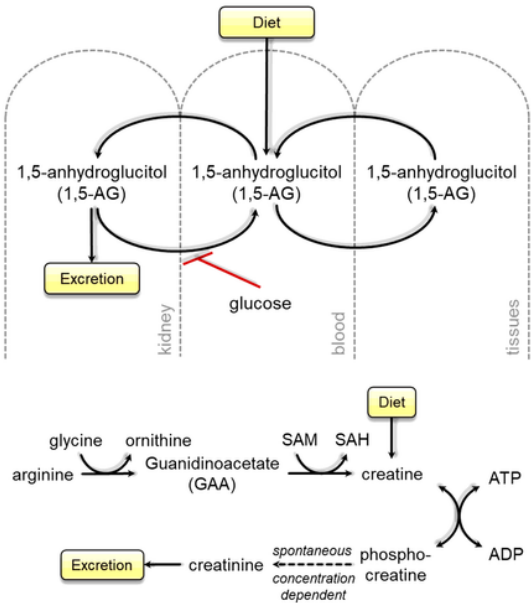
Figure 4

Metabolites with Kidney function

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.pdf