# Metabolomics for clinical use and research in chronic kidney disease

*Berthold Hocher[1] and Jerzy Adamski[2]*

Abstract | Chronic kidney disease (CKD) has a high prevalence in the general population and is associated with high mortality; a need therefore exists for better biomarkers for diagnosis, monitoring of disease progression and therapy stratification. Moreover, very sensitive biomarkers are needed in drug development and clinical research to increase understanding of the efficacy and safety of potential and existing therapies. Metabolomics analyses can identify and quantify all metabolites present in a given sample, covering hundreds to thousands of metabolites. Sample preparation for metabolomics requires a very fast arrest of biochemical processes. Present key technologies for metabolomics are mass spectrometry and proton nuclear magnetic resonance spectroscopy, which require sophisticated biostatistic and bioinformatic data analyses. The use of metabolomics has been instrumental in identifying new biomarkers of CKD such as acylcarnitines, glycerolipids, dimethylarginines and metabolites of tryptophan, the citric acid cycle and the urea cycle. Biomarkers such as c-mannosyl tryptophan and pseudouridine have better performance in CKD stratification than does creatinine. Future challenges in metabolomics analyses are prospective studies and deconvolution of CKD biomarkers from those of other diseases such as metabolic syndrome, diabetes mellitus, inflammatory conditions, stress and cancer.

Uraemic toxins
Solutes that are excreted by the healthy kidney but accumulate and contribute to uraemia in patients with CKD.

[1]*Department of Basic Medicine, Medical College of Hunan University, 410006 Changsha, China.* [2]*Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany.*

*Correspondence to J.A.* adamski@helmholtz-muenchen.de

doi:10.1038/nrneph.2017.30
Published online 6 Mar 2017

High-throughput and high-content omics approaches have revolutionized biomarker research. In particular, metabolomics has been instrumental for the identification of new biomarkers of chronic kidney disease (CKD) which can be measured in tissue, plasma, serum and urine samples. Metabolites are considered to be final proxies for physiological homeostasis and gene–environment interactions. They reflect processes that are ongoing or have already taken place. In addition, metabolite measurements can be multiplexed with existing clinical chemistry assays and clinical data so have huge potential for clinical diagnostics as well as patient stratification and monitoring of therapeutic responses.

CKD is fairly frequent in the general population worldwide and is becoming a major health-care issue in populations with an increasing proportion of elderly people[1] and/or an increasing prevalence of obesity and diabetes. Progression of CKD towards end-stage renal disease (ESRD) represents a huge financial burden to health-care systems as it increases the need for costly renal replacement therapies (peritoneal dialysis, haemodialysis and haemodiafiltration) and kidney transplantation. CKD also increases the risk of cardiovascular diseases (fatal and non-fatal stroke, myocardial infarction and peripheral artery disease) and vascular calcification. Cardiovascular mortality is at least 8–10-fold higher in patients with CKD than in the general population, and has not improved substantially in the past two decades (as indicated by the lack of a substantial improvement in the annual mortality of patients on haemodialysis)[2]. Patients with CKD also have a substantially increased risk of non-fatal and fatal infectious diseases, including sepsis, as well as an increased risk of acute kidney injury (AKI). They show reduced cognitive function due to microvascular diseases of the brain and impairment of nerve cell functions owing to the effects of uraemic toxins.

Analysis of mechanisms that underlie renal failure and the identification of unique biomarkers (such as metabolite ratios) have the potential to increase our understanding of CKD and improve diagnostic algorithms. In this Review we discuss metabolites and pathways that are associated with CKD, the specificity of these metabolic biomarkers and their functional relevance.

## The need for biomarkers in CKD

CKD is defined as the presence of kidney damage for at least 3 months, independent of the underlying chronic renal disease[3]. In addition to diagnosis, classification or staging of CKD provides a tool for clinical decision making. Both diagnosis and staging of CKD are usually based on urinary albumin excretion and/or estimated

**Key points**

- The human metabolome reflects genetic variability, intrinsic biochemical processes, environmental challenges and complex interactions of all these factors
- Metabolomics is instrumental in discovering specific biomarkers in diseases with systemic effects such as chronic kidney disease (CKD)
- Metabolomics analysis can detect CKD-relevant biomarkers in tissues, plasma, serum and urine samples
- Most metabolite biomarkers of CKD are markers of glomerular filtration, markers of tubular function or metabolites that reflect a decline in mitochondrial function, alterations in the urea cycle or amino acid metabolism
- As CKD stage increases, the metabolic biomarker signatures of different renal diseases tends to become more similar and less dependent on the underlying renal disease
- Metabolic biomarkers seen in the later stages of CKD reflect a loss of glomerular filtration, tubular function and a decline in kidney metabolism and endocrine function

glomerular filtration rate (eGFR)[4]. These methods are used in daily clinical practice because they are easy and inexpensive to perform. The formulas developed for estimation of GFR do, however, have limitations. For example, estimates that are based on creatinine are influenced by biological confounders such as age, sex and in particular muscle mass. As creatinine is produced by muscle metabolism, changes in muscle mass can result in false estimates of GFR[5].

More accurate estimation of GFR is needed, in particular for children and the elderly. The introduction of cystatin-C-based estimates of GFR represents a step in the right direction. Levels of cystatin C are less dependent on age, sex and in particular muscle mass than are levels of creatinine. However, high dosages of glucocorticoids increase the production of cystatin C independent of GFR, which might complicate the use of cystatin-C-based eGFR formula in patients who require this therapy[6]. Diseases of the thyroid gland also have a GFR-independent effect on cystatin C concentrations. Patients with hypothyroidism have low cystatin C levels, which might lead to false normal cystatin C concentrations in patients with hypothyroidism and impaired GFR, whereas in patients with hyperthyroidism, cystatin C concentrations are high independent of GFR[7,8]. As cystatin C is mainly synthesized by the liver, chronic and acute liver failure also have GFR-independent effects on serum cystatin C concentrations[9]. Importantly cystatin-C-based estimates of GRF are not suitable for determining the specific type of underlying renal disease and the likelihood of disease progression.

New urinary biomarkers of kidney damage such as kidney injury molecule 1 (KIM-1), neutrophil gelatinase-associated lipocalin and liver fatty acid-binding protein (L-FABP) also have limitations. For example, serum concentrations of L-FABP reflect the severity of tubular damage but do not provide information on the aetiologic factors that are causing this damage[10,11]. Urinary KIM-1 concentrations usually need to be normalized by urinary creatinine concentrations to account for urine volume[12]. Moreover these measurements cannot be used to distinguish between patients with slow or fast CKD progression.

Reliable biomarkers are needed for differentiation of CKD entities. Although eGFR and urinary albumin excretion are useful — with the above mentioned limitations — for staging the degree of renal impairment, they do not enable diagnosis of the underlying renal disease. In most cases kidney biopsies are needed to make a definitive diagnosis. Biopsy is an invasive procedure that can lead to major complications such as bleeding and even loss of the transplanted kidney. Moreover, the biopsy must be carried out by experienced physicians and experienced pathologists are required to analyse the samples. These logistic prerequisites might explain why biopsies are often done too late to be clinically useful, and are not done in all patients in whom they would be clinically useful. Thus in daily practice, exact diagnoses of renal disease are made too rarely and often too late.

New biomarkers are also needed to help guide therapy by identifying patients with fast disease progression who might need aggressive treatment. Biomarkers might help to distinguish treatment responders from non-responders at a very early stage and to identify the optimal treatment protocol. Moreover very sensitive biomarkers are needed to improve understanding of the efficacy and safety of compounds in drug development and clinical research[13]. Metabolomic analysis represents a unique tool to address clinical questions, as these techniques enable a metabolic fingerprint to be obtained rather than a single biochemical signal. Given the opportunities of modern bioinformatics, these fingerprints will be much more informative than eGFR calculations and urinary albumin excretion.

## Metabolomics analysis

Metabolomics analysis is a process of identifying and quantifying all metabolites in a given sample[14,15]. The focus of metabolomics is metabolites with molecular masses of 80–1,200 Da. Metabolomics analyses can detect hundreds to thousands of metabolites in a single sample, including hexoses (sugars), amino acids, dipeptides, lipids and organic acids. Both endogenous and exogenous metabolites can be detected by metabolomics and, depending on the indication and research questions, either group might contain differentiating biomarkers. In studies of the impact of the genome on the metabolome, endogenous metabolites are of primary interest, whereas in studies of the impact of the environment on human health, exogenous metabolites such as plant or bacterial-derived compounds (for example flavonoids or bile acids), xenobiotics (for example pesticides or plasticizers such as phthalates) and drugs (for example sleeping aids, pain killers or nephrotoxic substances such as gentamycin and ciclosporin) might be equally indicative. Phase I metabolites, which introduce reactive and polar groups into substrates, and phase II metabolites, which conjugate substrates with charged species such as glutathione, sulfate, glycine or glucuronic acid, can also be visualized by metabolomics. The multitude of metabolites that can be identified using metabolomics is termed the metabolome[14,16–18]. Many molecules are deposited in databanks such as LipidMaps[19], which specializes in lipids, or the Human Metabolome Database[20],

Metabolic fingerprint
A snapshot of the metabolites present in sample under specific conditions.

Mass spectrometry
An analytical method by which ionised molecules are detected according to their mass-to-charge ratio.

Proton nuclear magnetic resonance spectroscopy
An analytical method that analyses the absorption and re-emission of energy of proton nuclei in a strong magnetic field.

which includes all chemical classes of metabolites. These two databanks each describe >40,000 compounds. The size of the human metabolome is still being estimated but might be around 100,000 molecules or could greatly exceed this amount[20,21] if all intermediates and secondary metabolites are taken into account.

Metabolomics analysis works very well with many types of sample of human or animal origin, including body fluids (urine, serum, plasma, dried blood spots, saliva, cerebrospinal, lung lavage, rhinitis, lenses or seminal fluid), tissue (including biopsy samples), stool, cell culture or exhaled breath. Studies that analysed the metabolite signatures present in different body fluids of the same individual have been successfully accomplished[22,23]. These studies enable the construction of metabolic networks that link disease-associated metabolites from the studied biofluids with intervention outcomes in preclinical and clinical studies or across different timescales of glycaemic control. These kinds of metabolic networks will provide a much better understanding of pathways leading to the development of disease and potentially provide insight into disease pathogenesis[22,23]. Sample components that are not the metabolites of interest are termed the matrix. These matrix components might influence the performance of the metabolomics assays. The most commonly used sample types for metabolome research in CKD are urine, serum and plasma. Importantly several critical requirements exist for pre-analytical sample processing in metabolomics research (BOX 1).

Although it was anticipated that individual responses to nutritional factors might render metabolomics inapplicable to the analysis of biomarkers of disease, detailed studies have shown that such individual variations can be discriminated from disease markers[24]. Furthermore, many pathways of the human metabolome are stable over periods of several months[25] or years[26], and any observed changes are linked to genetic or environmental factors, including disease. Major confounding factors that impact metabolomics fingerprints include genetic background, sex, age, body mass index, medication use, lifestyle (for example exercise, smoking, coffee and alcohol consumption), circadian rhythms (shift work and mealtimes), hormonal status and nutrition (vegetarian, low carbohydrate, low fat, macrobiotic or Mediterranean diets). These factors should be considered when constructing biostatistical models for data analyses[24,27–36].

## Metabolomics technologies

Several technologies have been used to cover most of the metabolome and these are constantly being developed[37]. Most research and development in analytics for metabolomics has been performed using mass spectrometry (MS) (FIG. 1) and proton nuclear magnetic resonance spectroscopy (NMR)[38,39]. Although MS apparatus can be successfully operated in flow injection analysis mode (with direct injection of the sample), the mass spectrometer is usually coupled to an additional device to increase the resolution of analytes by gas chromatography (GC) or liquid chromatography (LC)[40]. The chromatographic step separates individual metabolites and substantially reduces the levels of components of the sample matrix, so increases the relative intensities of the analyte signals. Unfortunately, not all analytes will be separated chromatographically and some might be lost during this step. Analytes must

Box 1 | **Critical requirements for sample processing and handling in metabolomics analysis**[183–189]

**Collection**

- A standardized collection protocol should be used with strict adherence to the required steps and processing times to avoid generation of multicentre bias or variance between samples from different laboratories
- Sample labelling should be readable by both human (alphanumeric code) and machine (barcode) and withstand storage at low temperatures
- Endogenous sample metabolism should be stopped as quickly as possible by snap freezing or extraction with ice-cold organic solvents

**Storage and transport**

- Samples should be stored either at −80 °C or in liquid nitrogen (−20 °C storage is detrimental to the metabolic profile)
- Samples should be stored in aliquots suitable for analytical steps to avoid freeze-thaw cycles
- Sample transport should preferably be on dry ice

**Analysis**

- Normalization measures should be planned in advance and may include reference to total signal of all metabolites, wet weight of tissue, number of cells, osmolarity measurement, creatinine assay and/or addition of stable labelled metabolites
- Reference samples (preferably with the same matrix type) should be used with every batch of samples analysed to allow comparison of instrument performance and sample stability assessment; the reference sample might be pooled plasma or urine obtained in large amounts and stored in frozen aliquots
- Samples should be randomized before analyses; randomization should consider confounders such as sex, age, body mass index, medication use, nutrition or lifestyle, but not multicentre origin
- Samples that originated from the same individual should be included in the same analytical batch
- Analytical equipment should be calibrated before and after the study and the metabolome coverage of a reference sample monitored over the whole experiment

Tandem mass spectrometers
A mass spectrometer apparatus consisting of two quadrupole mass spectrometers connected by a single quadrupole.

Quadrupole linear ion trap (Q-TRAP)–MS
A mass spectrometer equipped with a quadrupole that is used to trap charged molecules.

Quadrupole time-of-flight (Q-TOF)–MS
A mass spectrometer that determines the mass-to-charge ratio of molecules based on their flight time in the apparatus.

Orbitrap
A mass analyser that converts frequency signals from trapped ions to mass spectrum using the Fourier transform.

Derivatization steps
Chemical modification of molecules to increases their ability to ionize before mass analyses.

undergo ionization before MS analysis but the efficacy of this process differs between metabolites and some might not be suitable for analysis in the mass spectrometer[41].

The MS can be constructed as quadrupole tandem mass spectrometers (MS/MS), which are very robust and sensitive but have lower mass resolution than other MS technologies. Quadrupole linear ion trap (Q-TRAP)-MS (FIG. 1) combines the advantages of tandem mass spectrometers with the additional multiple fragmentation and identification options of an ion trap[42,43]. Tandem mass spectrometers are very popular for targeted metabolomics (analytics that are set up to quantify selected metabolites, generally a few hundred metabolites per sample), whereas non-targeted metabolomics (analytics that are tuned to cover as many metabolites as possible (>700 per sample) for global profiling) use different apparatus. Quadrupole time-of-flight (Q-TOF)-MS has the advantage of fast simultaneous analyses of many compounds as well as high mass accuracy and resolution[44]. An ultrahigh performance LC-MS/MS unit equipped with orbitrap provides very high mass resolution, mass accuracy, high dynamic range and a low limit of detection[45]. This setup enables excellent preselection of metabolites and matrix separation to achieve the sensitivity and resolution required to enable the application of metabolomics to research into personalized medicine and health care[46]. The highest possible mass resolution is achieved with Fourier transform ion cyclotron resonance (FTICR), albeit at the expense of lower throughput than that of Q-TRAP and Q-TOF[47].

LC-MS/MS is applicable to both small and large molecules (for example amino acids and lipids), works well with thermally labile and very polar molecules, and does not normally require labour-intensive derivatization steps. However, this technique does not work well with very nonpolar and volatile substances.

GC-MS is currently used for many types of screening, for example in toxicology or pesticide analytics. The strengths of this method are very good separation properties, easy identification of molecules by fingerprint MS spectra, and very high comparability between different mass spectrometers and laboratories. For many substances, GC-MS is the gold standard for quantification. Nevertheless the technique has its limitations and is not applicable for molecules with high polarity or thermally labile substances. Furthermore, for most classes of molecules derivatization steps are required. GC-MS works best with nonpolar, volatile molecules

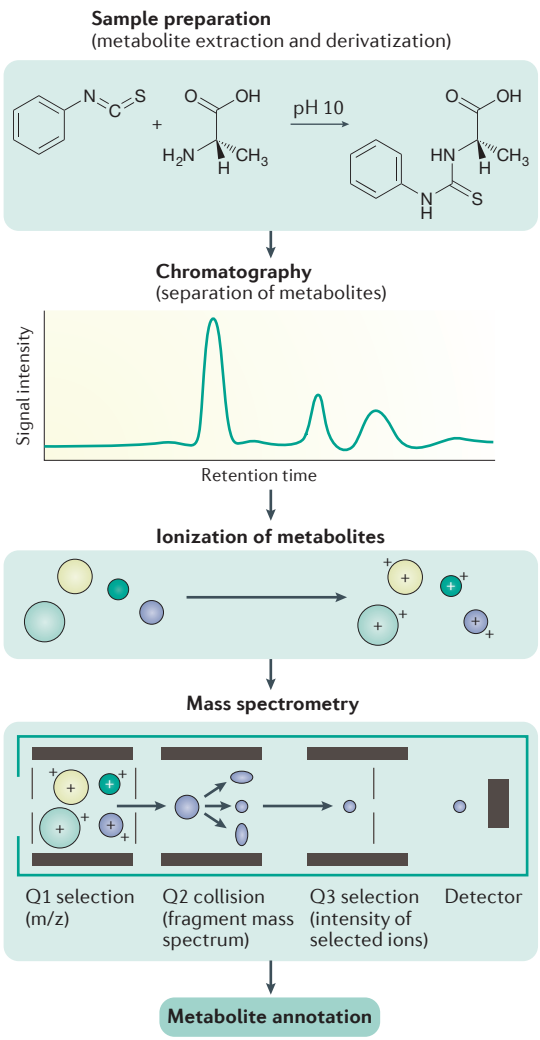



Figure 1 | **Essential elements of analytics with mass spectrometry (MS) for metabolomics.** Metabolites have to be extracted from the sample (for example using organic solvents) to facilitate further analytics. Some metabolites such as amino acids require derivatization steps to increase the mobility of their ions before MS. To increase resolution and minimize matrix effects, metabolites are separated using high performance liquid or gas chromatography. The individual peaks observed during chromatography may contain several metabolites. During chromatography, aliquots of separated metabolites are applied to ionization chambers; only ionized metabolites will be analysed later in the process. The MS provides information on mass/charge ratio (m/z), which is used to calculate the mass of individual metabolites. The mass spectrometer can be constructed in various ways; the schematic depicts a tandem mass spectrometer run in targeted mode. All parts of the mass spectrometer operate under a vacuum (depicted by the green line). Ionized molecules are pre-selected by m/z ratio in the first quadrupole (Q1), which consists of four metal rods (black bars, for clarity only two rods are shown) arranged in parallel around one axis. Radio frequency voltage is applied to the rods to accomplish the quadrupole function. Information on m/z is not sufficient to unequivocally identify a metabolite as some molecules have the same m/z but different chemical formula. In the second quadrupole (Q2) molecules released sequentially from Q1 are fragmented by collision with neutral gas. Each molecule fragments into a characteristic pattern, which can be used to unequivocally identify the molecule. To quantify the molecule of interest its fragment is selected in quadrupole 3 (Q3) and quantified in a detector. This procedure is repeated for every molecule of interest. Metabolite annotation is based on molecular mass, validation by fragmentation spectra and retention time during chromatography.

(for example free fatty acids, steroids, sugars and citric acid cycle metabolites) and runs at medium throughput (a few hundred samples per week with hundreds of metabolites identified).

Targeted and non-targeted metabolomic approaches require different sample processing and MS setups. In targeted mode LC-MS/MS can run at high throughput (>800 samples per week), whereas in non-targeted mode only lower throughput (150 samples per week) is possible. NMR spectroscopy or LC-NMR spectroscopy can run for long periods without a reduction in quality in the discovery phase at medium throughput, but due to sensitivity issues quantification is limited to a few hundred metabolites. The advantage of NMR is its high resolution in lipid analysis (for example for identification of subfractions of LDL or HDL). NMR spectroscopy analyses provide information only on 'known' metabolites (those annotated in databanks).

GC-MS, LC-MS/MS or NMR spectroscopy might overlap in identifying the same metabolites. At present LC-MS/MS has the highest metabolome coverage and resolution. In addition to 'known' metabolites LC-MS/MS might identify 'unknown' metabolites[48–50]. The latter group represents molecules that are characterized by retention time in LC, molecular mass, and fragmentation mass data in MS/MS, but without knowledge of the specific chemical formula. In most cases 'unknown' metabolites might be further processed[51] to identify them as molecules that already exist in databanks without association with given LC or MS parameters. The LC-MS/MS analyses might also discover metabolites that have not previously been identified or annotated.

MS or NMR spectroscopy approaches can be run in parallel on identical machines to increase the throughput without compromising resolution, sensitivity, accuracy and precision. Furthermore LC-MS, GC-MS and NMR spectroscopy analyses can all be performed on the same sample. Human serum and urine metabolomes have been visualized with different parallel metabolomics methods to reveal at least 4,229 and 2,206 metabolites, respectively[52,53]. Unfortunately, the diversity and complexity of the whole metabolome cannot be visualized in a single experiment at present.

Commonly used diagnostics such as ELISA or radioimmunoassays might be more sensitive than MS or NMR spectroscopy for detection of individual metabolites such as steroid hormones, but are not suitable for metabolic profiling of similar molecules due to cross-reactivity of antibodies. Most clinical laboratories already use MS (for example, for drug screening or to screen newborns for a variety of disorders), but NMR spectroscopy is not yet used for clinical diagnostics. Methods for high-throughput analyses of metabolites that are indicative of kidney dysfunction have been used with differing levels of success; the most frequently used technologies are HPLC-MS/MS or UHPLC-MS/MS[54–57].

## Statistical analysis of metabolomics data

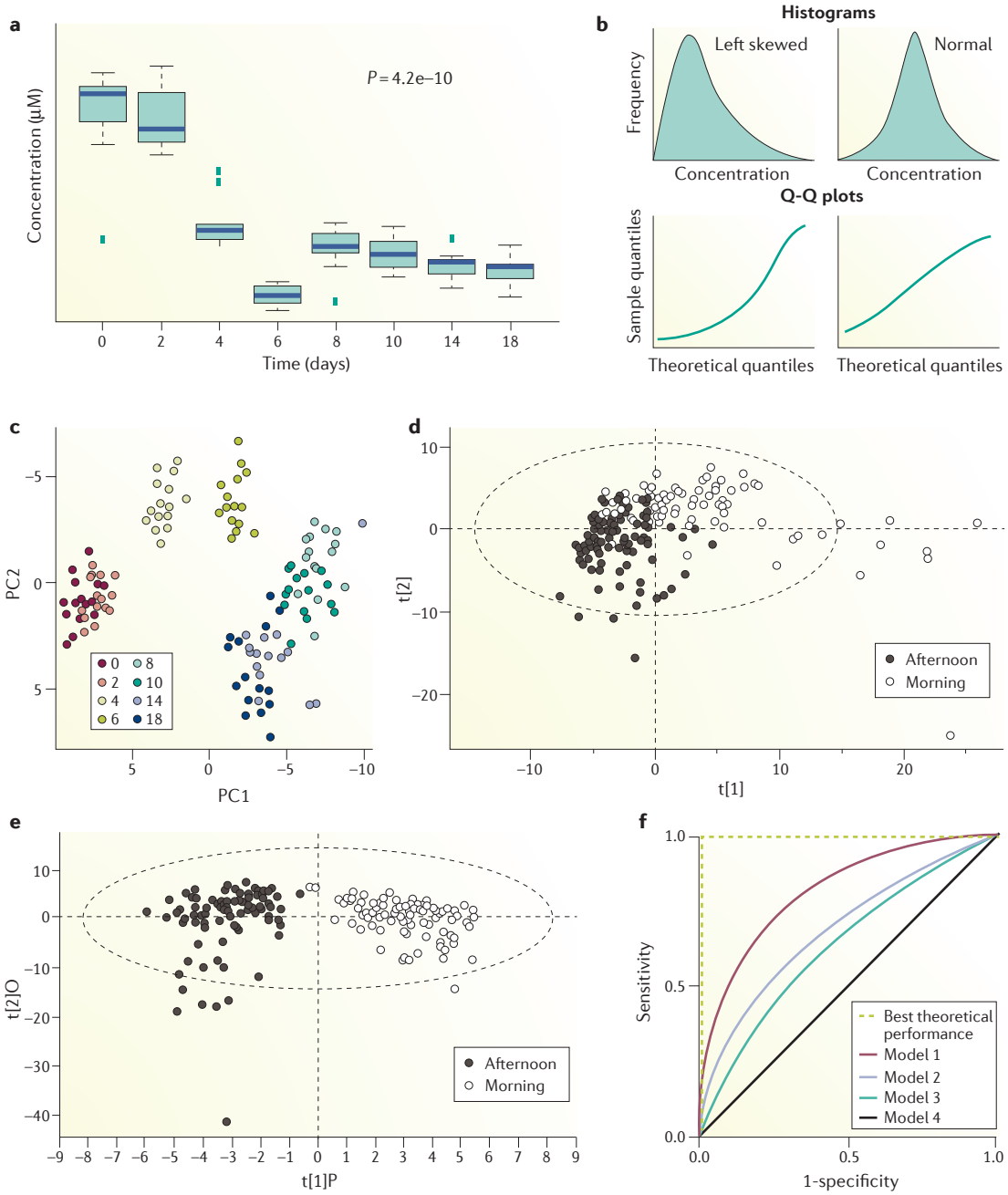
Similar to other omics approaches, metabolomics would be not possible without sophisticated bioinformatics techniques to support interpretation of large data sets[20,58,59]. Metabolomics experiments generate a wide variety of data (for example concentrations of metabolites) for each individual sample. These data have to be streamlined and simplified for further biostatistical analyses. The first main goal of the biostatistical analyses is to check if the data are valid and to present them in a meaningful way for non-specialist users.

Metabolites that are present in all samples do not require further streamlining. Quite often, however, there are samples that do not provide readings for each metabolite. The missing readings might be due to the fact that the levels of these metabolites are below the limit of detection (LOD). In targeted metabolomics missing metabolites do not generally pose an immediate threat to data integrity because metabolite measurements are normally run with internal standards that enable values to be obtained even for metabolite levels that are under the LOD. In these cases the data calculated below the LOD cannot be used directly as their linearity and absolute values are uncertain, but streamlining of data is quite straightforward. Additional data checks must be performed, for example by running outlier detection via boxplot analysis (FIG. 2a) or by data imputing for values below the LOD (for example by provision of a calculated, very low value). Data inputting is mandatory for non-targeted metabolomics, as internal standards do not normally exist for all metabolites for these kinds of measurements. In these cases data imputing is crucial to obtain the maximum amount of information possible. If none of the above approaches are feasible, these values have to be discarded as biostatistics would, in most cases, not work on data sets with missing values[60,61]. After data streamlining and imputing, the biostatistical analyses can be performed.

When comparing experimental data sets (for example control versus patient group) the sets must display enough diversity in the data for each parameter studied. As a first step the data distribution (FIG. 2b) has to be ascertained. The distribution can be assessed by statistical bootstrap[62] or a Q-Q plot[63]. Next, a decision must be made on which statistical tests to run, either univariate tests for one variable (hypothesis-driven approach) or multivariate tests for several variables (non-hypothesis driven approach).

In a hypothesis-driven approach for datasets with a normal distribution, parametric tests such as ANOVA or t-tests can be used to evaluate differences between groups. These univariate analyses might fail if the differences are very small for a single metabolite tested or if there are too many metabolites to be analysed. In the case of a non-normal distribution, non-parametric tests (such as Kruskal–Wallis or the Mann–Whitney U test) may be successfully applied[64,65].

Non-hypothesis driven approaches require multivariate analyses such as principal component analysis (PCA), partial least squares regression discriminant analysis (PLS-DA) or orthogonal partial last squares regression (OPLS), which not only differentiate metabolites but also identify possible dependencies between them (BOX 2). These different multivariate approaches are essential to reduce the data to only




include and subsequently display the important information to the user. They should be confirmed using a large validating dataset after the statistical model is built. PCA, PLS-DA and OPLS are generally done sequentially. First a PCA is produced from simplified and streamlined data that has been determined to be statistically valid. If after a PCA the data is not easy to read, PLS or PLS-DA can be performed to sharpen the data representation. OPLS and OPL-DA might then be performed to clean up the data further. Optimally, the discovery dataset should also be analysed with different approaches to cross-validate the models obtained in this manner. For example, one could start with a PLS/PLS-DA analysis and check if the same clusters are produced as for a PCA.

At present most biomarker studies are accomplished without absolute quantification of metabolites and instead provide a comparative fold change of up or downregulation. To become attractive and robust tools for diagnostics, these biomarkers have to be validated by targeted metabolomics and the discovery analyses have to display the reference ranges as well as the area under the receiver operating characteristic (ROC) curve describing the specificity and sensitivity (FIG. 2f). An in-depth description of how to calculate and interpret the ROC curve has been published previously[66].

## Metabolites as biomarkers of CKD

Several metabolites and pathways are associated with CKD[55,67,68–69] (Supplementary information S1 (figure)),

Figure 2 | **Data presentation and common biostatistical approaches in metabolomics.** **a** | A boxplot for a chosen metabolite. Each box corresponds to samples collected on a different day of the experiment. Boxplots describe the data by quartiles (Q1 and Q3) for which the point of origin is the median of the data. Extreme outliers are defined by values in the range of Q1 – 1.5*(Q3 – Q1) and Q3 + 1.5*(Q3 – Q1). If data points below the limit of detection can be defined as extreme outliers they should be discarded. Each box contains 75% of the values for the samples collected on the relevant day. The median (Q2) is depicted as a dark blue bar and outliers are shown as bright blue squares. The whiskers show one standard deviation above and below the mean. The *P* value is given for the difference of all data in the boxes (Kruskal–Wallis U test). **b** | Data distribution. The frequency of the data distribution might be normal or other (for example skewed or bimodal). The upper panels show example histograms, whereas the lower panels show the corresponding quartile–quartile (Q-Q) plots. **c** | Principal component analysis (PCA) can be applied to large datasets, which can include metabolite concentrations as well as further phenotyping data. The PCA graph shows differences between samples based on multiple metabolites ranked according to contribution strength into eigenvalue/eigenvector pairs. The eigenvectors are the solutions for which the maximum variance can be calculated by the PCA algorithm and the eigenvalues are the spread of the variance (the length of the eigenvector). The eigenvalue/eigenvector pair is calculated for each variable in the data. The highest ranked metabolite combinations are in the PC1 axis followed by PC2, PC3 etc. Here data from the same experiment as part **a** are shown for all metabolites after calculation of eigenvectors and eigenvalues. Each dot represents a different sample. **d** | Partial least squares regression discriminant analysis (PLS-DA) provides information on differences between samples, which are calculated for all metabolites. The graph shows a comparison of nuclear magnetic resonance data from urine samples collected in the morning and afternoon. t[1] and t[2] represent the linear regression model and the response vector, respectively. Each dot represents a different sample. **e** | Orthogonal partial last squares regression is used to reduce noise in PLS data. This graph shows the same data as in part **d**. The first principal component (t[1]P) is plotted against the first orthogonal component (t[2]O). **f** | A receiver operating characteristic curve is used to analyse the performance of a diagnostic assay by plotting the true positive discovery rate (sensitivity) versus the false positive rate (1-specificity). The best theoretical performance (specificity 1 and selectivity 1) is shown by the dashed line. Different biostatistics models (including several combinations of metabolites) are represented by the solid lines. The area under the curve can be calculated for each model and is given as a percentage. Permission to reproduce parts **d** and **e** obtained from American Chemical Society © Wagner, S. *et al. Anal. Chem.* **79**, 2918–2926 (2007).

and use of combinations of biomarkers enable better discrimination of patients from healthy individuals than do single metabolites[70]. For example, an elegant study that used a discovery approach followed by validation steps in human samples and parallel analyses in a rat model identified the combination of ricinoleic acid, stearic acid, cytosine, 3-methylhistidine, argininic acid and two glycerophospholipids (lipophosphatidic acid 16:0 and 18:2) as a very well performing CKD biomarker panel with a sensitivity of 83.3% and a specificity of 96.7%[70]. The levels of some CKD biomarkers might, however, also be altered in other diseases such as diabetes and hypertension. Use of state-of-the-art metabolomic technologies will enable the identification of biomarkers that provide valuable clinical information for various aspects of CKD (GFR, rate of disease progression, underlying disease and response to treatment) in a single sample.

The main pathophysiological implication for patients with CKD is known for most of the existing metabolomics biomarkers (BOX 3; see Supplementary information S2 (table)). Several biochemical biomarkers, such as creatinine, urea cycle metabolites or uric acid, are linked to GFR. Individually all of these GFR biomarkers have shortcomings owing to confounding by other non-GFR related conditions; for example creatinine levels are influenced by muscle mass, urea cycle biomarkers are affected by nutritional status and cystatin C by the function of the thyroid gland. Although these individual shortcomings cannot be overcome, modern bioinformatics tools will be able to use the individual information provided by each of these biomarkers to provide more accurate estimates of GFR. For example, formulae have been developed that include cystatin C and creatinine to improve estimation of GFR. Future formulae based on metabolomics will consider more metabolites to overcome the remaining individual disadvantages. This integrated approach will also be applied to the known biomarkers of oxidative stress that provide an indication of CKD progression. To facilitate understanding of the origin of the most frequent metabolite biomarkers, a succinct overview of their biochemistry is provided below.

### *Amino acids and their metabolites*

***Symmetric dimethylarginine (SDMA) and asymmetric dimethylarginine (ADMA).*** The arginine metabolites SDMA (function unknown) and ADMA (an inhibitor of nitric oxide synthases) differ in their position of methylation and have distinct biological properties. As SDMA is slowly metabolized and cleared from the body by the kidney, renal dysfunction results in increased plasma and serum levels of this metabolite[71,72]. The levels of ADMA are also increased in CKD[73], possibly as a result of endothelial dysfunction[74]. Increases in ADMA levels are, however, also associated with many other diseases, including vascular dysfunction[75], migraine[76], hyperglycaemia[77] and sepsis[78]. A reduction in renal methylarginine metabolism has been shown to protect against kidney damage[79] but the mechanism is unknown.

***Tryptophan.*** L-Tryptophan is a precursor in the biosynthesis of melatonin[80] and serotonin[81]. Breakdown of tryptophan through the kynurenine pathway results in niacin and quinolinate[82]. This pathway is upregulated in renal insufficiency and CKD, resulting in increased levels of these metabolites[83–85]. Another metabolite of tryptophan, the amine indoxyl sulfate[86], also accumulates in the serum of patients with CKD[87], whereas serum levels of C-glycosyltryptophan increase with age[33].

### *Nitric oxide*

Nitric oxide is a highly reactive free radical compound that has a role in nonspecific immune system signalling, and also acts as a paracrine signal transducer, resulting in vasodilation. The concentration of nitric oxide that is required for immune signalling is about 1,000 times higher than that required for vasodilation[88–91]. Metabolism of arginine by nitric oxide synthases (NOS) results in citrulline and nitric oxide. Inducible NOS has a role in nitric oxide formation during the nonspecific immune response, whereas endothelial NOS synthesizes NO in much lower quantities to stimulate vasodilation.

Box 2 | **Multivariate analyses used in biomarker research**

**Principal component analysis (PCA)**
PCA is a very popular and powerful unsupervised method that is used to explain observed variance in the metabolite dataset by a calculation of dimensionless components. In a given data set a PCA tries to find a solution to calculate the maximum variance for each variable — in the case of metabolomics the measured levels of each metabolite — and fulfils two major roles: reduction of data complexity and presentation of the data in a way that a non-specialist can easily understand[190]. A PCA will in most cases provide information on group differences and most contributing metabolites from each component right away. This test is mostly used to find previously unknown groupings in data and can be further refined or enhanced if the grouping is not sufficiently distinct using a PLS-DA analysis.

**Partial least squares regression discriminant analysis (PLS-DA)**
A PLS-DA is a supervised approach using a linear regression model on a data matrix (for example metabolite data) versus a response vector (the 'discriminant', for example a disease or control assignment)[191,192]. This analysis can be applied alone — with prior knowledge of groupings in the data (such as known biomarkers) — or in conjunction with grouping information provided or derived from a PCA. A potential problem in PLS-DA might be overfitting, which leads to noise generation and poorly readable output graphs. The same principles apply for an unsupervised PLS analysis.

**Orthogonal partial last squares regression (OPLS)**
The OPLS is a further development of PLS and provides more intuitive data interpretation by a further reduction of data dimensionality, and thus noise reduction. After the OPLS analysis, the orthogonal noise, which is unrelated to the response, is provided separately to the data, which is related to the response[193].

**Orthogonal partial last squares regression discriminant analysis (OPLS-DA)**
In biomarker research the dimensionality reduction of OPLS was combined with discriminant analysis and resulted in OPLS-DA. This method has the advantage of being, again, supervised by a categorical variable and thus being even more accurate and easier to read for the user.

ADMA blocks eNOS activity so inhibits nitric oxide formation and nitric-oxide-dependent vascular relaxation[92]. Endothelial nitric oxide synthesis is reduced in the setting of CKD[93].

### *Polyamines*

Polyamines are involved in nucleic acid binding, promoter control and regulation of oxidative stress[94–96]. The polyamine putrescine can be synthesized from arginine via the agmatine or ornithine pathways. Reactions of putrescine with $H_2O_2$ result in spermidine, which in turn can be converted into spermine by spermine oxidase. The concentrations of polyamines, including spermidine, spermine and putrescine, were first shown to be elevated in patients with CKD in the 1980s[97]. Spermidine has now been identified as a biomarker of eGFR decline[98].

### *Urea cycle metabolites*

The liver and kidney both participate in the conversion of ammonia into uric acid and urea via the urea cycle (also known as the ornithine cycle). Metabolic links between the urea cycle and polyamine metabolism, nitric oxide synthesis, the citrate cycle and arginine metabolism make data interpretation complex. A 2016 study reported that urea cycle metabolites do not change substantially with age in patients with CKD[99]. Plasma levels of citrulline, ornithine and arginine do, however, change with CKD stage[100]. Moreover, the citrulline-to-arginine and ornithine-to-arginine ratios are increased in the advanced stages of CKD[101].

### *Uric acid*

The final breakdown product of purine is uric acid, which can be further metabolized into allantoin. Urate is produced by xanthine oxidase from xanthine and hypoxanthine, which might be elevated in the setting of hypoxia or oxidative stress[102]. The levels of hypoxanthine are increased in patients on haemodialysis compared to those on peritoneal dialysis[103]. Elevated serum levels of uric acid induce renal injury with an associated inflammatory response, but the detailed mechanisms are not known. Increased plasma and urinary levels of uric acid have been reported in nephropathy models[102,104].

### *Lipids*

Lipids are amphiphilic and their most common function is in the formation of bilayers or vesicles. They might act as transport or energy intermediates and as precursors of signal transduction (for example phospatidylinositols, eicosanoids or ceramides)[105–109]. Several classes of lipids have been instrumental in risk stratification and biomarker research in cardiovascular disease[110], diabetes[111], endometriosis[112] and kidney disorders[98,113,114]. Lipid classes that are relevant as biomarkers in CKD include fatty acids, glycerolipids, sphingolipids and sterols (including steroids). The glycerolipids are distinguished by a polar head, which might be choline, serine or ethanolamine, and by the chain lengths of two fatty acids. Sphingolipids are composed of serine and a fatty acid. In humans, the most common lipids of this class are sphingomyelins.

Quantification of various lipids can be accomplished by fast FIA-MS/MS, but analysis of fatty acid length and of the positions of double bonds is challenging in population studies. As only the total composition of the lipid classes can be determined, and side chain and substitution regioselectivity and stereochemistry is unknown, the nomenclature follows the LipidMaps system[115].

### *Acylcarnitines*

L-Carnitine is an amino acid that can be obtained from the diet (for example in red meat) and synthesized in the liver and muscle[116]. The kidney has a role in the biosynthesis of carnitine from lysine and methionine and the excretion of carnitine into urine and plasma[117]. Carnitine has protective effects in models of neurodegeneration, age-related decline in mitochondrial function and oxidative-stress-related gene expression, but the mechanisms have not yet been resolved[118–120]. Moreover, carnitine serves as a carrier for fatty acids into mitochondria[121], and concentrations of acylcarnitines are indicative of the rate of β-oxidation of fatty acids[122].

Some acylcarnitines might have distinct functions, for example, isovalerylcarnitine and hexadecanoylcarnitine are pro-apoptotic[123,124]. Moreover changes in concentrations of acylcarnitines reflect different metabolic processes; for example the concentration of

isovalerylcarnitine reflects the metabolism of valine, isoleucine and leucine[125] rather than that of fatty acyls. In the normal kidney the release of carnitine refuels the mitochondrial fatty acid shuttle in other organs. In non-pathological conditions acylcarnitine levels are very low and barely detectable. The presence of acylcarnitines in plasma, serum or urine indicates mitochondrial damage and concentrations of acylcarnitines are increased in the plasma of individuals with reduced eGFR[114].

### *Steroids*

Steroids are lipophilic signalling molecules that are synthesized by the gonads, adrenal gland, liver and lung[126–128] and control a multitude of processes beyond sexual dimorphism. Cholesterol and dehydroepiandrosterone sulfate (DHEA-S) are the most abundant steroids. These steroids are of interest in CKD because of their roles in immunomodulation[129], regulation of stress responses[130] and modulation of lipid metabolism[131]. The enzyme 11β-hydroxysteroid dehydrogenase controls the bioavailability of cortisol in the kidney, which if disturbed can result in hypertension[132,133].

### *Citric acid cycle intermediates*

The generation of ATP and cofactors such as NADH takes place in the citric acid cycle (also known as the tricarboxylic acid cycle). This cycle is a hub for a multitude of metabolites, which are also important in other pathways including gluconeogenesis (oxaloacetate), amino acid metabolism (fumarate, succinate, oxaloglutarate and oxaloacetate) or fatty acid metabolism (acetyl-coenzyme A). Impaired kidney function leads to an increase in the urinary levels of citric acid cycle intermediates. This effect is especially prominent in diabetic kidney disease[134,102,135].

### *Oxidative stress*

Oxidative stress resulting from depletion of oxidants and increased generation of free radicals[136] is a key factor that contributes to the excess morbidity and mortality of patients with CKD[137,138]. In CKD, redox signalling is associated with dysfunction of lipid metabolism resulting in increased levels of glycerophosphoethanolamines and bile acids[138]. Although the important role of oxidative stress in CKD is well recognized, clinical tools for monitoring oxidative stress are limited. Quantitative analysis of oxidative stress is very challenging as special pre-analytical conditions have to be met, which is not always realistic in human population studies. Levels of the oxidative stress markers methionine sulfoxide, the methionine sulfoxide-to-methionine ratio[139,140] and malonedialdehyde[141,142] have, however, been successfully analysed using metabolomics techniques. An imbalance of plasma amino acids, metabolites and lipids is associated with oxidative stress in patients with lysinuric protein intolerance[138].

### *Xenobiotics*

***1,5-Anhydroglucitol (1,5AG).*** 1,5AG is considered to be of exogenous origin and is present in plasma. In 1981, the levels of 1,5AG were shown to reflect glycaemic status in patients with diabetes[143]. Functional kidneys reabsorb 1,5AG, whereas glucose inhibits this process[144]. In incident CKD, serum levels of 1,5AG are reduced[145].

***Meta-cresol (m-cresol).*** M-Cresol and its isomers p-cresol and o-cresol are widely used in many cosmetics, drugs, pesticides and disinfectants, and high levels of exposure might have adverse effects such as the development of allergies and carcinogenesis[146,147]. Methylphenols[148] and p-cresol are renal toxins[149]. P-Cresol has been reported to mediate autophagy in the renal proximal tubules[150,151].

## Association studies with metabolomics

Genome-wide association studies with metabolomics have provided new insights into how genes can influence human metabolic homeostasis, environmental impact and drug exposition[48,152]. These studies have revealed

Box 3 | **Metabolomics biomarkers with pathophysiological implications for patients with CKD**

- Creatinine (GFR, muscle mass, nutritional status)
- Polyamines such as spermine or spermidine (nucleic acid binding, promoter control, oxidative stress regulation)
- Dimethylarginines such as SDMA and ADMA (endothelial function)
- Urea cycle (GFR, endothelial dysfunction via L-arginine, CKD stage)
- Uric acid (GFR, insulin resistance, metabolic syndrome, cardiovascular risk)
- Tryptophan metabolism (endothelial dysfunction, atherosclerosis, age)
- Amino acids (nutritional status, protein metabolism, lipid metabolism, signal transduction)
- Lipids (nutritional status, atherosclerosis, metabolic syndrome)
- Acylcarnitines (mitochondrial damage, cardiovascular complications of uraemia, lipid and amino acid metabolism)
- Steroids (endocrine function of the adrenal glands and gonads, stress, immunomodulation, hypertension, obesity)
- Oxidative stress (CKD progression)
- Citric acid cycle (energy status of the body, muscle energy metabolism)
- 1,5-Anhydroglucitol (microvascular outcomes, particularly in patients with diabetes)
- Meta-cresol (marker of exposure to environmental renal toxins)

ADMA, asymmetric dimethylarginine; CKD, chronic kidney disease; GFR, glomerular filtration rate; SDMA, symmetric dimethylarginine.

Isotope dilution
An approach to determine the concentration of substances by comparison with a stable isotope-labelled added internal standard.

that each individual possesses a metabolic phenotype — termed metabotype — that is considerably determined by genetic factors. Metabotypes are conserved over time and perturbations of an individual metabotype can indicate the beginning of a disease. In studies with epidemiologically balanced cohorts (matched for sex, age, BMI, eGFR and comorbidities such as diabetes or hypertension), many metabolites have been identified as indicative of kidney failure.

A cross-sectional study reported that the human serum metabolome changed with decreased GFR in the general population[114]. A targeted metabolomics (LC-MS/MS) study identified serine and the lipid glutarylcarnitine as strongly inversely associated with reduced eGFR[98]. This approach was further refined with analyses of kidney function decline and CKD, which indicated that spermidine (a marker of oxidative stress and amino acid metabolism), as well as the metabolite ratios of phosphatidylcholine diacyl C42:5-to-phosphatidylcholine acyl-alkyl C36:0 and kynurenine-to-tryptophan were associated with decline in eGFR[98]. The kynurenine-to-tryptophan ratio was also associated with a significantly higher incidence of CKD. Using the same discovery (KORA) and replication (TwinsUK) cohorts, the researchers expanded their studies in serum to non-targeted metabolomics using combined GC-MS and LC-MS/MS[153]. They showed pair-wise associations with eGFR for *N*-acetylcarnosine, C-mannosyltryptophan, *N*-acetylalanine, pseudouridine, erythronate and myo-inositol. Moreover, kidney function decline and incident CKD were associated with increased concentrations of O-sulfo-L-tyrosine, C-glycosyltryptophan and pseudouridine, indicating changes in oxidative stress, pyrimidine and amino acid metabolism, and signal transduction pathways.

C-Mannosylosyltryptophan and pseudouridine are interesting biomarkers because, in contrast to creatinine, their levels are not dependent on muscle mass[153]. A study that used the same analytical approach (GC-MS and LC-MS/MS) in plasma of individuals with CKD stages 2–4 revealed that several processes change with CKD progression, including amino acid metabolism (increased conversion of arginine to ADMA and lower levels of ornithine in CKD stage 4), coagulation and/or inflammation (increased levels of fibrinopeptide-A), carboxylate anion transport (elevated monocarboxylate and di-carboxylate anions such as γ-glutamylglutamine) and production of adrenal steroid hormones (increased cortisol)[100].

An interesting long-term (19.6 years of follow-up) longitudinal study that used non-targeted metabolomics analysis (GC-MS and LC-MS/MS) of the serum of African Americans without CKD, reported that three metabolites, creatine, carnitine and betaine, were positively associated with eGFR[145] calculated using the CKD-EPI formula[154]. By contrast, increased concentrations of 3-indoxyl sulfate (a metabolite of tryptophan) were associated with a substantial decrease in eGFR. Increases in the levels of 1,5-anhydroglucitol and 5-oxoproline were significantly associated with lower risk of incident CKD.

Analysis of plasma samples from participants in the Framingham Heart Study who did not have CKD at baseline identified multiple metabolites that were significantly associated with incident CKD during 8 years of follow-up[155]. These metabolites included amino acids and their intermediates (citrulline, kynurenine, 5-hydroxyindoleacetic acid and quinolinic acid), nucleotide metabolites (xantosine, adenosine, β-aminoisobutyric acid), citric acid cycle metabolites (isocitrate, aconitate), lipid metabolites (choline, lysophosphatidylcholines 18:2 and 18:1) and sugar metabolites (inositol, sucrose) as well as trimethylamine-N-oxide (a marker of microbial metabolism). Using isotope dilution this study differentiated biomarkers involved in renal metabolism (for example, citrulline and choline) and renal secretion (for example, kynurenic acid), and identified them as eGFR-independent predictors of CKD.

Simultaneous analysis of plasma and urine by targeted metabolomics (LC-MS/MS) was undertaken to analyse the amino acid metabolic profiles of patients with different CKD stages[101]. In the plasma of patients on haemodialysis, the concentrations of most amino acids and amines were decreased, with the exception of ADMA, hydroxyl-kynurenine, citrulline, asparagine and proline, which were increased. Moreover, plasma and urinary levels of citrulline, and plasma but not urinary concentrations of ADMA, were higher in patients with CKD stages 4 and 5 than in those with less severe renal disease (CKD stages 2 and 3). By contrast the plasma ratios of valine-to-glycine and tyrosine-to-phenylalanine were lower in patients with advanced stages of CKD. These alterations in plasma and urine metabolite profiles were not normalized by dialysis.

A comparative study of plasma and urine samples from patients with membranous nephropathy revealed that the metabolite profiles of both sample types reflect kidney dysfunction[156]. Serum samples of patients with higher urine protein levels revealed increased concentrations of several amino acids (asparagine, serine and threonine), and citric acid, and reduced concentrations of m-cresol compared with samples from patients with lower urinary protein levels. On the other hand, levels of cholesterol, dicarboxylic acids (adipic acid, 2-hydroxy-sebacic acid) and phenolic acids (vanillactic acid) were significantly elevated in the urine of patients with lower protein levels[156].

NMR analysis of urine samples from patients with stage 3 and 4 CKD revealed elevation of many uraemic solutes, including the filtration marker p-cresyl sulfate in all patients and the xenobiotic 2-hydroxyisobutyric acid and sulfate metabolite dimethyl sulfone in those who were not on dialysis[157]. Another small NMR study reported that patients with focal segmental glomerulosclerosis (FSGS) had a different metabolite pattern to that of patients with other types of glomerulopathies (membranous nephropathy, minimal change nephropathy and IgA nephropathy) and healthy controls[158]. Major differentiating metabolites with elevated concentrations in the FSGS group were glucose, dimethylamine and trimethylamine, and with lowered concentrations

were β-hydroxyisovalerate, pyruvate, citrate, valine, hippurate, isoleucine, phenylacetylglycine, tyrosine and 3-methylhistidine[151].

In a comprehensive metabolomics study that aimed to identify a urine metabolomic signature of CKD, urine samples were first analysed by NMR in the discovery phase and then validated by LC-triple quadrupole MS[159]. The analysis identified seven urinary metabolites that differed between CKD and non-CKD urine samples, including α-phenylacetylglutamine, glutamate, guanidoacetate (all of which have roles in amino acid metabolism) and trimethylamine-N-oxide (a marker of microbial metabolism). The fold changes ranged from 1.73–7.54.

All of the association studies described above have an intrinsic shortcoming that hampers further progress in the field: classification of patients in CKD studies is usually based on eGFR rather than measured GFR. This limitation creates noise in the data sets. Moreover data on factors that have a proven impact on the metabolome such as medication use, comorbidities, coexisting infectious diseases, current nutrition (for example, western or vegetarian diet), exposure to environmental toxins, and perhaps most importantly, precise clinical examination are often missing, incomplete or wrong. As analytical technical and biostatistical tools are well advanced and biochemical analysis is of a high standard, the remaining challenge for further progress in CKD metabolomics research is to improve the quality of the clinical phenotyping.

## Factors that affect the kidney metabolome

### *Diabetes mellitus*

Kidney failure is a very frequent complication of diabetes mellitus and metabolomics has been applied to study this issue[160]. The PREVEND case-controlled study showed that patients with type 2 diabetes mellitus (T2DM) and normoabuminuria had no changes in their metabolite profile, whereas those with microalbuminuria had lower levels of histidine and higher levels of the lipid butenoylcarnitine in their plasma than normoalbuminuric patients. As revealed by FIA-MS/MS and LC-MS/MS, the urine of patients with microalbuminuria also had reduced levels of glutamine, tyrosine and hexoses. These findings were used to create a new model to improve risk prediction for macroalbuminuria[160].

A GC-MS study showed that diabetic individuals (T1DM and T2DM) with and without CKD differ in their urine metabolome[161]. Distinguishing metabolites included 3-hydroxyisovalerate, aconitate, citrate, 2-ethyl,3-hydroxypropionate, glycolate, 3-hydroxyisobutyrate, 2-methylacetoacetate, 3-methyladipic acid, 3-hydroxypropionate and uracil, indicating that mitochondrial function is dysregulated in diabetic kidney disease. Moreover, a GC-MS and LC-MS analysis of baseline 24 h urine samples from patients with T1DM who did not have any sign of diabetic kidney disease at baseline identified urinary biomarkers that differentiated between progressive and non-progressive forms of albuminuria, including discriminating metabolites in tryptophan metabolism, acylcarnitines and acylglycines[162]. This metabolite profile reached an accuracy of 75% and a precision of 73%, enabling discrimination between those patients who developed diabetic kidney disease and those who did not progress to diabetic kidney disease during the study duration of 5.5 years.

In 2016, a NMR study demonstrated that deteriorating glycaemic control affects patterns of serum metabolites[163]. Among patients with CKD, those with T2DM had higher levels of branched chain amino acids, tyrosine, and formate but lower levels of urea, creatinine, arginine and pyruvate[163].

### *Toxins*

Environmental toxins clearly impact kidney tissue and the urinary metabolome, and are associated with CKD[164]. For example, among individuals with long-term environmental exposure to cadmium, those with higher urinary cadmium levels (>5 ug/l) had significantly higher urinary levels of amino acids and their metabolites (glutamine, cysteine, tyrosine, *N*-methyl-L-histidine, histidinol, taurine, phenylacetylglutamine, hippurate, and pyroglutamic acid), galactose, myo-inositol, xanthine, urea, deoxyadenosine monophosphate and steroids (17α-hydroxyprogesterone, corticosterone, tetrahydrocortisone and estrone) than those with lower cadmium levels, indicating cadmium-induced damage to renal tubular cells[164].

A study in rats exposed to agents with known nephrotoxic effects (including 2-bromoethanamine, *N*-phenylanthranilic acid, gentamicin, ciclosporin or cisplatin, showed significant changes in plasma levels of three metabolites: 3-methylhistidine, 3-indoxyl sulfate and guanidoacetate[165]. These biomarkers reflect different processes elicited by the challenge; guanidoacetate is a marker of changed metabolism of arginine and glycine, whereas 3-methylhistidine and 3-indoxyl sulfate directly reflect altered GFR.

The pesticide phorate has also been shown to be nephrotoxic by metabolomic analysis of rat urine[166]. Administration of this agent resulted in increased levels of diethylthiophosphate, indoxyl sulfic acid, cholic acid and 7-methylguanine, and decreased levels of uric acid, suberic acid and citric acid, indicating pleiotropic effects on energy-related metabolism, liver and kidney function, the antioxidant system and DNA damage.

### *Stress*

Stress is a strong confounder in metabolomics profiling[24]. An LC-MS study in rats without CKD showed that key metabolic pathways are altered in response to stress (induced by hydrocortisone treatment), resulting in decreases in energy metabolism (lactic acid and acetylcarnitine), lipid metabolism (free fatty acids, 1-monolinoleoylglycerol and cholesterol), gut microbiota metabolism (indole-3-propionic acid), biosynthesis of catecholamine (noradrenaline), and elevated alanine metabolism[24]. This important observation should be considered when designing studies and considering pre-analytical parameters in human studies. Thus the evaluation of patient stress levels, for example by using validated stress questionnaires, needs to be considered

when planning new CKD studies. Such an approach will enable correction for stress as a confounding factor when analysing the data.

### *Drug treatment*

Metabolomics has also identified many pathways that are affected by off-target drug effects. For example, statin treatment to lower cholesterol levels has been shown to affect the plasma metabolites of patients with CKD[167]. After 1 year of statin treatment, the levels of malone dialdehyde and kynurenine, as well as the kynurenine-to-tryptophan ratio were improved in comparison to untreated patients, indicating potential beneficial cholesterol-lowering independent off-target effects of statins.

Aristolochic acid has a long history in the treatment of renal disease[168] but has been shown to be nephrotoxic by metabolomics analysis (LC-Q-TOF)[169]. In rats, numerous urine metabolites of distinct pathways were increased by aristolochic acid treatment, including creatinine, uric acid, hippuric acid, allantoin, phenylacetylglycine, kynurenic acid, cholic acid, taurine, indoxyl sulfate, indole-3-carboxylic acid, glucose, citrate, fumarate, aconitate, p-cresyl sulfate and spermine. Further studies by the same group expanded the metabolite panel to lipids and identified changes in the kidney tissue levels of acylcarnitines, phospatidylcholines, lysophosphatidylcholines, phosphatidylethanolamines, lyso-phosphatidyl-ethanol-amines, ceramides and triglycerides in rats in response to aristolochic acid[170]. These changes in lipid metabolites preceded changes in renal histology and function. A subsequent study in rats subjected to aristolochic-acid-induced nephropathy identified 12-ketodeoxycholic acid, taurochenodeoxycholic acid, lipophosphatidic acid (15:0) and docosahexaenoic acid as biomarkers for early detection of tubulo-interstitial nephropathy[138]. This interesting study also showed that treatment with irbesartan reversed these biomarker changes in the rat model and in patients with CKD[138]. Taken together, these data show that exposure to aristolochic acid results in dynamic changes in the metabolism of fatty acid, phospholipid, and glycerolipid in kidney tissue before and after the onset of detectable changes in renal function or histology. These findings indicate that an altered tissue lipid metabolism is involved in the pathogenesis of toxin-induced nephropathy.

The antiviral acyclovir also has nephrotoxic effects. In a male rat model, metabolomics analysis showed that acyclovir treatment interrupts the metabolism of arachidonic acid, tryptophan, arginine, proline, glycerophospholipid and other pathways[171]. These findings indicate that in addition to its effects on viral metabolism, acyclovir affects the metabolism of eukaryotic cells.

Antibiotics such as aminoglycosides might also cause kidney injury. Metabolomics analysis of rat urine has shown that even at low doses, gentamicin strongly affects tryptophan metabolism, resulting in an increase in tryptophan and a decrease in kynurenic acid[172], possibly reflecting immunomodulating effects of aminoglycosides, as kynurenic acid has immunosuppressive properties.

### *Ischaemic injury*

A murine study demonstrated that ischaemic kidney injury impacts the plasma and kidney metabolomes[173]. Plasma, kidney medulla and cortex samples taken during experimental bilateral ischaemia and during the recovery phase were profiled using non-targeted GC-MS and LC-MS/MS. Ischaemia-induced changes were first observed in the kidney cortex followed by medulla and plasma. In the early ischaemic phase, levels of 3-indoxyl sulfate were increased, whereas in the later phase levels of oxidated lipids such as prostaglandins were increased. The analysis also indicated an adaptation in energy source (switch from glucose to lipids) in response to ischaemia. These findings suggest that metabolomics might be a suitable tool to differentiate between different stages of AKI after ischaemia–reperfusion injury. These different stages might require different therapeutic approaches, but this hypothesis remains to be proven.

### *Graft rejection*

Renal allograft rejection remains a major problem and diagnosis is mainly based on analysis of biopsy samples. Thus a need exists for new and reliable biomarkers to differentiate decreases in GFR after kidney transplantation owing to rejection from decreases due to other causes such as nephrotoxic effects of immuosuppressive agents, kidney infections or delayed graft function. Serum LC-MS analysis of patients with acute kidney graft rejection identified a rejection-indicative panel of metabolites, consisting of creatinine, kynurenine, uric acid, phosphatidylcholines, sphingomyelins, lyso-phosphatidylcholines and polyunsaturated fatty acids[174]. This panel might potentially be useful for clinical decision making, such as deciding whether a biopsy is necessary.

Interestingly, even before transplantation, patients who experienced acute graft rejection had lower serum levels of the steroid DHEA-S than those who did not reject their grafts[174]. If this finding is confirmed in further independent studies, this biomarker might be a useful tool for personalized immunosuppressive approaches based on the rejection risk estimated using simple DHEA-S measurements together with analysis of other risk factors.

### *Renal cancer*

Analysis of frozen tumour tissue from patients with clear cell renal cell carcinoma (ccRCC) using non-targeted metabolomics (GC-MS and LC-MS/MS) has revealed metabolic adaptation in a multitude of pathways[175]. For example, greater upregulation of aerobic glycolysis (indicated by a net increase in lactate) was identified in higher grade tumour cell lines compared to lower grade tumour cell lines. Whether or not these biomarkers are suitable tools for tumour grading now needs to be investigated in clinical studies[175]. Metabolomic findings in tumour cell lines provide important information on altered cellular pathways in these tumour cells. Overall the most pronounced alterations were seen in cancer cell lines

with the most aggressive phenotypes. Glutaminolysis was connected to inhibition of ROS as manifested by upregulation of the glutathione pathway. Increased levels of acylcarnitines (medium chain laurylcarnitine, hexanoylcarnitine, octanoylcarnitine and long chain myristoylcarnitine, palmitoylcarnitine and stearoylcarnitine) indicated inhibition of mitochondrial β-oxidation of fatty acids. In the matrix analysed, the Warburg effect was very prominent; glycolysis, the fatty acid pathway and the glutamine pathway stopped supplying metabolites to the citric acid cycle, and consequently this cycle was downregulated[175]. These findings might provide insights into the disease mechanism and hence have implications for treatment.

### *Disease progression*

In their early stages different diseases are associated with distinct metabolite profiles. For example, lysophosphatidylcholine 18:2, acetylcarnitine and glycine have been found to be predictive of T2DM[111,176], whereas branched amino acids (valine, leucine, isoleucine) in serum or plasma are a hallmark of overt diabetes mellitus[177]. Moreover progression of diabetes mellitus is significantly associated with changes in the concentrations of branched chain and aromatic amino acids (phenylalanine and tyrosine)[178]. No deterioration of kidney function was observed in these studies.

A metabolomics study of plasma samples obtained from the aortas and renal veins of individuals with established CKD identified arginine, methionine and threonine as biomarkers of renal prognosis[179]. Biomarkers that are associated with progression towards ESRD have also been identified (isethionate, saccharate, trimethylamine *N*-oxide, 4-oxopentanoate, cytidine, gluconate, glucuronate, guanidinosuccinate, 2-hydroxyisobutyrate, uridine, 5-oxoproline, pimelate, *N*-acetylneuraminate, 3-methylhistidine, phthalate and citramalate)[180].

Whether these new biomarkers have better specify and sensitivity for CKD progression than do classical risk factors such as failure to respond to treatment, degree of urinary albumin excretion, failure to control blood pressure, and type of underlying CKD remains to be determined. Interestingly, the metabolic biomarkers that are observed in the late stages of CKD tend to be similar to those seen in advanced diabetic nephropathy, graft rejection, ischaemic injury or renal cancer, and indicate a loss of glomerular filtration and a decline in tubular function and kidney metabolism.

## Prediction of CKD

To date, integrative epidemiological analyses and genetic approaches for the prediction of kidney failure have not been successful[57]. Several models for CKD prediction have now been constructed using metabolic biomarkers. The predictive accuracy of these models is more than 80%[69,181] and different metabolite panels can be used to increase their selectivity and specificity[70]. Although proteomic and/or metabolomic biomarkers for CKD prediction do not currently perform better than eGFR[182], developments in this field are very promising, as illustrated by the results of studies that have been designed to account for known confounders, such as age, sex, stress, medication use or comorbidities. Given the limitations of the existing equations for estimation of GFR, further improvement of the predictive power of metabolomics for CKD research and diagnosis will require metabolomics data to be correlated with measured GFR values obtained using gold-standard methods such as inulin clearance. From the perspective of study design this task will be challenging, but such an approach is clearly required to further improve our understanding of CKD.

## Conclusions and future perspectives

Metabolomics profits from progress in bioinformatics and study design for other high content analyses, including genetics and proteomics. Future challenges in metabolomics analyses lie both in study design and in the deconvolution of CKD biomarkers from those of other diseases such as the metabolic syndrome, diabetes, inflammation, stress or cancer. Prospective studies with very precise phenotypic characterization of patients, using for example measured GFR instead of eGFR, are clearly needed because poor phenotypic characterization often limits the opportunities provided by metabolomics.

Another target for metabolomics research is to support clinical decision-making, for example by distinguishing between patients with fast or slow CKD progression and by monitoring therapy efficacy. Moreover, new biomarkers are needed as non-invasive tools to aid the accurate diagnosis of renal diseases with similar clinical features (such as urinary protein excretion in various glomerular diseases) but differing treatment and prognosis. These goals are very realistic as analysis of the mechanisms that underlie renal failure and the differentiation of unique biomarkers (such as metabolite ratios) are rapidly developing.

Bioinformatics will enable complex information obtained from metabolomics to be condensed into key basic medical information relevant for therapeutic decision making for individual patients. Combinations of GFR biomarkers will enable the development of more accurate equations to estimate GFR. Other classes of metabolites will provide information about disease progression (for example oxidative stress biomarkers), microvascular disease and the risk of cardiovascular disease. Eventually metabolomics diagnostic approaches might generate useful patient-specific clinical data on kidney function, rate of CKD progression, cardiovascular risk and degree of insulin resistance, and enable personalized treatment based on individual risk profiles. So far metabolomics data have been used in system biology approaches to identify molecular pathways, but this information is often of limited clinical value. The next step should be the application of these tools to identify not just signatures of biochemical pathways but also of medical and pathophysiological conditions. This integrated approach will enable personalized treatment of patients, but large scale clinical studies using metabolomics are needed to reach this goal.

1. Zhang, Q. L. & Rothenbacher, D. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health* **8**, 117 (2008).
2. Jha, V., Wang, A. Y. & Wang, H. The impact of CKD identification in large countries: the burden of illness. *Nephrol. Dial. Transplant.* **27** (Suppl. 3), iii32–iii38 (2012).
3. Stevens, P. E., Levin, A. & Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann. Intern. Med.* **158**, 825–830 (2013).
4. KDIGO. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int. Suppl.* **3**, 4 (2013).
**Guidelines for the diagnosis and treatment of CKD that are used wordwide in clinical practice and research.**
5. Ferguson, T. W., Komenda, P. & Tangri, N. Cystatin C as a biomarker for estimating glomerular filtration rate. *Curr. Opin. Nephrol. Hypertens.* **24**, 295–300 (2015).
6. Bokenkamp, A., van Wijk, J. A., Lentze, M. J. & Stoffel-Wagner, B. Effect of corticosteroid therapy on serum cystatin C and beta2-microglobulin concentrations. *Clin. Chem.* **48**, 1123–1126 (2002).
7. Knight, E. L. *et al.* Factors influencing serum cystatin C levels other than renal function and the impact on renal function measurement. *Kidney Int.* **65**, 1416–1421 (2004).
**Analysis of factors other than renal function that influence levels of serum cystatin C in patients with CKD.**
8. Fricker, M., Wiesli, P., Brandle, M., Schwegler, B. & Schmid, C. Impact of thyroid dysfunction on serum cystatin C. *Kidney Int.* **63**, 1944–1947 (2003).
9. Wan, Z. *et al.* Combining serum cystatin C with total bilirubin improves short-term mortality prediction in patients with HBV-related acute-on-chronic liver failure. *PLoS ONE* **10**, e0116968 (2015).
10. Ichikawa, D. *et al.* Utility of urinary tubular markers for monitoring chronic tubulointerstitial injury after ischemia-reperfusion. *Nephrology (Carlton)* http://dx.doi.org/10.1111/nep.12998 (2017).
11. Foster, M. C. *et al.* Urinary biomarkers and risk of ESRD in the Atherosclerosis Risk in Communities Study. *Clin. J. Am. Soc. Nephrol.* **10**, 1956–1963 (2015).
12. Peralta, C. A. *et al.* Associations of urinary levels of kidney injury molecule 1 (KIM-1) and neutrophil gelatinase-associated lipocalin (NGAL) with kidney function decline in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am. J. Kidney Dis.* **60**, 904–911 (2012).
13. Levin, A., Lancashire, W. & Fassett, R. G. Targets, trends, excesses, and deficiencies: refocusing clinical investigation to improve patient outcomes. *Kidney Int.* **83**, 1001–1009 (2013).
14. Tweeddale, H., Notley-McRobb, L. & Ferenci, T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. *J. Bacteriol.* **180**, 5109–5116 (1998).
15. Nicholson, J. K., Lindon, J. C. & Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189 (1999).
**The concept of metabolomic research for biomedicine.**
16. Su, L. J. *et al.* The use of metabolomics in population-based research. *Adv. Nutr.* **5**, 785–788 (2014).
17. Griffin, J. L. The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Phil. Trans. R. Soc. B* **361**, 147–161 (2006).
18. Lindon, J. C. & Nicholson, J. K. The emergent role of metabolic phenotyping in dynamic patient stratification. *Expert Opin. Drug Metab. Toxicol.* **10**, 915–919 (2014).
**An explanation of the concept and use of metabolomic phenotyping in human cohorts.**
19. Fahy, E., Sud, M., Cotter, D. & Subramaniam, S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* **35**, W606–W612 (2007). 20.
Wishart, D. S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37**, D603–D610 (2009).
**Overview of the Human Metabolome Database, which provides a versatile resource for metabolite function and characteristics.**
21. Zamboni, N., Saghatelian, A. & Patti, G. J. Defining the metabolome: size, flux, and regulation. *Mol. Cell* **58**, 699–706 (2015).
22. Huang, D. *et al.* An integrated plasma and urinary metabonomic study using UHPLC-MS: intervention effects of Epimedium koreanum on 'Kidney-Yang Deficiency syndrome' rats. *J. Pharm. Biomed. Anal.* **76**, 200–206 (2013).
23. Yousri, N. A. *et al.* A systems view of type 2 diabetes-associated metabolic perturbations in saliva, blood and urine at different timescales of glycaemic control. *Diabetologia* **58**, 1855–1867 (2015).
24. Krug, S. *et al.* The dynamic range of the human metabolome revealed by challenges. *FASEB J.* **26**, 2607–2619 (2012).
25. Floegel, A. *et al.* Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS ONE* **6**, e21103 (2011).
26. Yousri, N. A. *et al.* Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics* **10**, 1005–1017 (2014).
27. Mook-Kanamori, D. O. *et al.* Type 2 diabetes is associated with postprandial amino acid measures. *Arch. Biochem. Biophys.* **589**, 138–144 (2016).
28. Jacobs, S. *et al.* Evaluation of various biomarkers as potential mediators of the association between coffee consumption and incident type 2 diabetes in the EPIC-Potsdam Study. *Am. J. Clin. Nutr.* **100**, 891–900 (2014).
29. Xu, T. *et al.* Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med.* **11**, 60 (2013).
30. Lankinen, M. *et al.* Metabolomic analysis of plasma metabolites that may mediate effects of rye bread on satiety and weight maintenance in postmenopausal women. *J. Nutr.* **141**, 31–36 (2011).
31. Jaremek, M. *et al.* Alcohol-induced metabolomic differences in humans. *Transl Psychiatry* **3**, e276 (2013).
32. Yu, Z. *et al.* Human serum metabolic profiles are age dependent. *Aging Cell* **11**, 960–967 (2012).
33. Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int. J. Epidemiol.* **42**, 1111–1119 (2013).
34. Asher, G. & Sassone-Corsi, P. Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock. *Cell* **161**, 84–92 (2015).
35. Guo, L. *et al.* Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. *Proc. Natl Acad. Sci. USA* **112**, E4901–E4910 (2015).
36. Heinzmann, S. S. *et al.* Stability and robustness of human metabolic phenotypes in response to sequential food challenges. *J. Proteome Res.* **11**, 643–655 (2012).
37. Hyotylainen, T. Novel methodologies in metabolic profiling with a focus on molecular diagnostic applications. *Expert Rev. Mol. Diagn.* **12**, 527–538 (2012).
38. Whitfield, P. D., German, A. J. & Noble, P. J. Metabolomics: an emerging post-genomic tool for nutrition. *Br. J. Nutr.* **92**, 549–555 (2004).
39. Gebregiworgis, T. & Powers, R. Application of NMR metabolomics to search for human disease biomarkers. *Comb. Chem. High Throughput Screen.* **15**, 595–610 (2012).
40. Artati, A., Prehn, C., Möller, G. & Adamski, J. in *Genetics Meets Metabolomics: From Experiment to Systems Biology* (ed. Suhre, K.) 13–38 (Springer, 2012).
**Discussion of study design and tools used for metabolomics analyses in human populations.**
41. Baldwin, M. A. Mass spectrometers for the analysis of biomolecules. *Methods Enzymol.* **402**, 3–48 (2005).
**Summary of the principles of mass spectrometry.**
42. Hopfgartner, G. *et al.* Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J. Mass Spectrom.* **39**, 845–855 (2004).
43. Khamis, M. M., Adamko, D. J. & El-Aneed, A. Mass spectrometric based approaches in urine metabolomics and biomarker discovery. *Mass Spectrom. Rev.* **36**, 115–134 (2017).
44. Campbell, J. L. & Le Blanc, J. C. Using high-resolution quadrupole TOF technology in DMPK analyses. *Bioanalysis* **4**, 487–500 (2012).
45. Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M. & Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **81**, 6656–6667 (2009).
46. Cacciatore, S. & Loda, M. Innovation in metabolomics to improve personalized healthcare. *Ann. NY Acad. Sci.* **1346**, 57–62 (2015).
47. Nicolardi, S., Bogdanov, B., Deelder, A. M., Palmblad, M. & van der Burgt, Y. E. Developments in FTICR-MS and its potential for body fluid signatures. *Int. J. Mol. Sci.* **16**, 27133–27144 (2015).
48. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54–60 (2011).
**Study of the impact of the human genome on the metabolome and functional annotation of the genome.**
49. Hirayama, A. *et al.* Metabolic profiling reveals new serum biomarkers for differentiating diabetic nephropathy. *Anal. Bioanal. Chem.* **404**, 3101–3109 (2012).
50. Tsutsui, H. *et al.* Biomarker discovery in biological specimens (plasma, hair, liver and kidney) of diabetic mice based upon metabolite profiling using ultra-performance liquid chromatography with electrospray ionization time-of-flight mass spectrometry. *Clin. Chim. Acta* **412**, 861–872 (2011).
51. Krumsiek, J. *et al.* Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet.* **8**, e1003005 (2012).
52. Psychogios, N. *et al.* The human serum metabolome. *PLoS ONE* **6**, e16957 (2011).
53. Bouatra, S. *et al.* The human urine metabolome. *PLoS ONE* **8**, e73076 (2013).
54. Klepacki, J., Klawitter, J., Klawitter, J., Thurman, J. M. & Christians, U. A high-performance liquid chromatography — tandem mass spectrometry — based targeted metabolomics kidney dysfunction marker panel in human urine. *Clin. Chim. Acta* **446**, 43–53 (2015).
55. Rhee, E. P. Metabolomics and renal disease. *Curr. Opin. Nephrol. Hypertens.* **24**, 371–379 (2015).
**General review on metabolomics applied to renal disease.**
56. Zhang, A., Sun, H., Qiu, S. & Wang, X. Metabolomics insights into pathophysiological mechanisms of nephrology. *Int. Urol. Nephrol.* **46**, 1025–1030 (2014).
57. Schena, F. P. Biomarkers and personalized therapy in chronic kidney diseases. *Expert Opin. Investig. Drugs* **23**, 1051–1054 (2014).
58. Bartel, J., Krumsiek, J. & Theis, F. J. Statistical methods for the analysis of high-throughput metabolomics data. *Comput. Struct. Biotechnol. J.* **4**, e201301009 (2013).
59. Issaq, H. J., Van, Q. N., Waybright, T. J., Muschik, G. M. & Veenstra, T. D. Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.* **32**, 2183–2199 (2009).
60. Armitage, E. G., Godzien, J., Alonso-Herranz, V., Lopez-Gonzalvez, A. & Barbas, C. Missing value imputation strategies for metabolomics data. *Electrophoresis* **36**, 3050–3060 (2015).
61. Succop, P. A., Clark, S., Chen, M. & Galke, W. Imputation of data values that are less than a detection limit. *J. Occup. Environ. Hyg.* **1**, 436–441 (2004).
62. Berk, M., Ebbels, T. & Montana, G. A statistical framework for biomarker discovery in metabolomic time course data. *Bioinformatics* **27**, 1979–1985 (2011).
63. Wartenberg, D. & Northridge, M. Defining exposure in case-control studies: a new approach. *Am. J. Epidemiol.* **133**, 1058–1071 (1991).
64. Kruskal, W. H. & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).
65. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).
66. Xia, J., Broadhurst, D. I., Wilson, M. & Wishart, D. S. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* **9**, 280–299 (2013).
**An introduction to metabolites as biomarkers.**
67. Lopez-Giacoman, S. & Madero, M. Biomarkers in chronic kidney disease, from kidney function to kidney damage. *World J. Nephrol.* **4**, 57–73 (2015).
68. Fassett, R. G. *et al.* Biomarkers in chronic kidney disease: a review. *Kidney Int.* **80**, 806–821 (2011).
69. Breit, M. & Weinberger, K. M. Metabolic biomarkers for chronic kidney disease. *Arch. Biochem. Biophys.* **589**, 62–80 (2016).
70. Zhang, Z. H. *et al.* Metabolomic signatures of chronic kidney disease of diverse etiologies in the rats and humans. *J. Proteome Res.* **15**, 3802–3812 (2016).

71. Fleck, C. *et al.* Serum concentrations of asymmetric (ADMA) and symmetric (SDMA) dimethylarginine in renal failure patients. *Kidney Int. Suppl.* **78**, S14–S18 (2001).
**Study showing that the levels of biomarkers of endothelial dysfunction are increased in CKD.**
72. Wahbi, N. *et al.* Dimethylarginines in chronic renal failure. *J. Clin. Pathol.* **54**, 470–473 (2001).
73. Vallance, P., Leone, A., Calver, A., Collier, J. & Moncada, S. Accumulation of an endogenous inhibitor of nitric oxide synthesis in chronic renal failure. *Lancet* **339**, 572–575 (1992).
74. Raptis, V., Kapoulas, S. & Grekas, D. Role of asymmetrical dimethylarginine in the progression of renal disease. *Nephrology (Carlton)* **18**, 11–21 (2013).
75. Notsu, Y., Yano, S., Shibata, H., Nagai, A. & Nabika, T. Plasma arginine/ADMA ratio as a sensitive risk marker for atherosclerosis: Shimane CoHRE study. *Atherosclerosis* **239**, 61–66 (2015).
76. Neri, M. *et al.* A meta-analysis of biomarkers related to oxidative stress and nitric oxide pathway in migraine. *Cephalalgia* **35**, 931–937 (2015).
77. Trgo, G. *et al.* Association of asymmetric dimethylarginine with acute pancreatitis-induced hyperglycemia. *Pancreas* **45**, 694–699 (2016).
78. Aydemir, O., Ozcan, B., Yucel, H., Yagmur Bas, A. & Demirel, N. Asymmetric dimethylarginine and L-arginine levels in neonatal sepsis and septic shock. *J. Matern. Fetal Neonatal Med.* **28**, 977–982 (2015).
79. Tomlinson, J. A. *et al.* Reduced renal methylarginine metabolism protects against progressive kidney damage. *J. Am. Soc. Nephrol.* **26**, 3045–3059 (2015).
80. Wurtman, R. J., Larin, F., Axelrod, J., Shein, H. M. & Rosasco, K. Formation of melatonin and 5-hydroxyindole acetic acid from 14C-tryptophan by rat pineal glands in organ culture. *Nature* **217**, 953–954 (1968).
81. Udenfriend, S., Weissbach, H. & Bogdanski, D. F. Increase in tissue serotonin following administration of its precursor 5-hydroxytryptophan. *J. Biol. Chem.* **224**, 803–810 (1957).
82. Hankes, L. V., Brown, R. R. & Schmaeler, M. Metabolism of isomers of 3-hydroxykynureninec14 to quinolinic acid, niacin metabolites and carbon dioxide. *Proc. Soc. Exp. Biol. Med.* **121**, 253–259 (1966).
83. Saito, K. *et al.* Mechanism of increases in L-kynurenine and quinolinic acid in renal insufficiency. *Am. J. Physiol. Renal Physiol.* **279**, F565–F572 (2000).
84. Schefold, J. C. *et al.* Increased indoleamine 2,3-dioxygenase (IDO) activity and elevated serum levels of tryptophan catabolites in patients with chronic kidney disease: a possible link between chronic inflammation and uraemic symptoms. *Nephrol. Dial. Transplant.* **24**, 1901–1908 (2009).
85. Streja, E. *et al.* Niacin and progression of CKD. *Am. J. Kidney Dis.* **65**, 785–798 (2015).
86. Bryan, G. T. Quantitative studies on the urinary excretion of indoxyl sulfate (indican) in man following administration of L-tryptophan and acetyl-L-tryptophan. *Am. J. Clin. Nutr.* **19**, 105–112 (1966).
87. Niwa, T. Uremic toxicity of indoxyl sulfate. *Nagoya J. Med. Sci.* **72**, 1–11 (2010).
88. Sudar-Milovanovic, E. *et al.* Hormonal regulation of nitric oxide (NO) in cardio-metabolic diseases. *Curr. Pharm. Des.* http://dx.doi.org/10.2174/1381612823666170124124855 (2017).
89. Persichini, T. *et al.* Cross-talk between NO synthase isoforms in neuro-inflammation: possible implications in HIV-associated neurocognitive disorders. *Curr. Med. Chem.* **23**, 2706–2714 (2016).
90. Ghimire, K., Altmann, H. M., Straub, A. & Isenberg, J. S. Nitric oxide: what's new to NO? *Am. J. Physiol. Cell Physiol.* http://dx.doi.org/10.1152/ajpcell.00315.2016 (2016).
91. Erkens, R. *et al.* Modulation of local and systemic heterocellular communication by mechanical forces: a role of eNOS. *Antioxid. Redox Signal.* http://dx.doi.org/10.1089/ars.2016.6904 (2016).
92. Leiper, J. & Vallance, P. Biological significance of endogenous methylarginines that inhibit nitric oxide synthases. *Cardiovasc. Res.* **43**, 542–548 (1999).
93. Baylis, C. Nitric oxide deficiency in chronic kidney disease. *Am. J. Physiol. Renal Physiol.* **294**, F1–F9 (2008).
94. Popolo, A., Adesso, S., Pinto, A., Autore, G. & Marzocco, S. L-Arginine and its metabolites in kidney and cardiovascular disease. *Amino Acids* **46**, 2271–2286 (2014).
95. Pegg, A. E. The function of spermine. *IUBMB Life* **66**, 8–18 (2014).
96. Ramani, D., De Bandt, J. P. & Cynober, L. Aliphatic polyamines in physiology and diseases. *Clin. Nutr.* **33**, 14–22 (2014).
97. Saito, A., Takagi, T., Chung, T. G. & Ohta, K. Serum levels of polyamines in patients with chronic renal failure. *Kidney Int. Suppl.* **16**, S234–S237 (1983).
98. Goek, O. N. *et al.* Metabolites associate with kidney function decline and incident chronic kidney disease in the general population. *Nephrol. Dial. Transplant.* **28**, 2131–2138 (2013).
99. Rroji, M. *et al.* Association of advanced age with concentrations of uraemic toxins in CKD. *J. Nephrol.* **29**, 81–91 (2016).
100. Shah, V. O. *et al.* Plasma metabolomic profiles in different stages of CKD. *Clin. J. Am. Soc. Nephrol.* **8**, 363–370 (2013).
101. Duranton, F. *et al.* Plasma and urinary amino acid metabolomic profiling in patients with different levels of kidney function. *Clin. J. Am. Soc. Nephrol.* **9**, 37–45 (2014).
**A comparison of plasma and urinary biomarkers in patients with kidney dysfunction.**
102. Liu, J. *et al.* 1H NMR-based metabonomic analysis of serum and urine in a nonhuman primate model of diabetic nephropathy. *Mol. Biosyst.* **9**, 2645–2652 (2013).
103. Choi, J. Y. *et al.* Dialysis modality-dependent changes in serum metabolites: accumulation of inosine and hypoxanthine in patients on haemodialysis. *Nephrol. Dial. Transplant.* **26**, 1304–1313 (2011).
104. Abbiss, H. *et al.* Development of a non-targeted metabolomics method to investigate urine in a rat model of polycystic kidney disease. *Nephrology (Carlton)* **17**, 104–110 (2012).
105. Riehle, R. D., Cornea, S. & Degterev, A. Role of phosphatidylinositol 3,4,5-trisphosphate in cell signaling. *Adv. Exp. Med. Biol.* **991**, 105–139 (2013).
106. Wakelam, M. J., Pettitt, T. R. & Postle, A. D. Lipidomic analysis of signaling pathways. *Methods Enzymol.* **432**, 233–246 (2007).
107. Serhan, C. N., Haeggstrom, J. Z. & Leslie, C. C. Lipid mediator networks in cell signaling: update and impact of cytokines. *FASEB J.* **10**, 1147–1158 (1996).
108. Chaurasia, B. & Summers, S. A. Ceramides — lipotoxic inducers of metabolic disorders. *Trends Endocrinol. Metab.* **26**, 538–550 (2015).
109. Dennis, E. A. & Norris, P. C. Eicosanoid storm in infection and inflammation. *Nat. Rev. Immunol.* **15**, 511–523 (2015).
110. Wurtz, P. *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131**, 774–785 (2015).
111. Wang-Sattler, R. *et al.* Novel biomarkers for pre-diabetes identified by metabolomics. *Mol. Syst. Biol.* **8**, 615 (2012).
112. Vouk, K. *et al.* Discovery of phosphatidylcholines and sphingomyelins as biomarkers for ovarian endometriosis. *Hum. Reprod.* **27**, 2955–2965 (2012).
113. Albrecht, E. *et al.* Metabolite profiling reveals new insights into the regulation of serum urate in humans. *Metabolomics* **10**, 141–151 (2014).
114. Goek, O. N. *et al.* Serum metabolite concentrations and decreased GFR in the general population. *Am. J. Kidney Dis.* **60**, 197–206 (2012).
115. Fahy, E. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **50** (Suppl.), S9–S14 (2009).
**An explanation of lipid nomenclature.**
116. Pearson, D. J. & Tubbs, P. K. Acetyl-carnitine in heart and liver. *Nature* **202**, 91 (1964).
117. Guder, W. G. & Wagner, S. The role of the kidney in carnitine metabolism. *J. Clin. Chem. Clin. Biochem.* **28**, 347–350 (1990).
118. Bernard, A., Rigault, C., Mazue, F., Le Borgne, F. & Demarquoy, J. L-Carnitine supplementation and physical exercise restore age-associated decline in some mitochondrial functions in the rat. *J. Gerontol. A Biol. Sci. Med. Sci.* **63**, 1027–1033 (2008).
119. Pescosolido, N., Imperatrice, B. & Karavitis, P. The aging eye and the role of L-carnitine and its derivatives. *Drugs R. D.* **9** (Suppl. 1), 3–14 (2008).
120. Zambrano, S. *et al.* The renoprotective effect of L-carnitine in hypertensive rats is mediated by modulation of oxidative stress-related gene expression. *Eur. J. Nutr.* **52**, 1649–1659 (2013).
121. Bremer, J. Carnitine as a fatty acid carrier in intermediary metabolism. *Nature* **196**, 993–994 (1962).
122. Pande, S. V. A mitochondrial carnitine acylcarnitine translocase system. *Proc. Natl Acad. Sci. USA* **72**, 883–887 (1975).
123. Ferrara, F., Bertelli, A. & Falchi, M. Evaluation of carnitine, acetylcarnitine and isovalerylcarnitine on immune function and apoptosis. *Drugs Exp. Clin. Res.* **31**, 109–114 (2005).
124. Mutomba, M. C. *et al.* Regulation of the activity of caspases by L-carnitine and palmitoylcarnitine. *FEBS Lett.* **478**, 19–25 (2000).
125. Roe, C. R. *et al.* Isolated isobutyryl-CoA dehydrogenase deficiency: an unrecognized defect in human valine metabolism. *Mol. Genet. Metab.* **65**, 264–271 (1998).
126. Miller, W. L. Steroid hormone synthesis in mitochondria. *Mol. Cell. Endocrinol.* **379**, 62–73 (2013).
127. Prehn, C., Möller, G. & Adamski, J. Recent advances in 17beta-hydroxysteroid dehydrogenases. *J. Steroid Biochem. Mol. Biol.* **114**, 72–77 (2009).
128. Labrie, F. Intracrinology in action: importance of extragonadal sex steroid biosynthesis and inactivation in peripheral tissues in both women and men. *J. Steroid Biochem. Mol. Biol.* **145**, 131–132 (2015).
129. Chapman, K. E. *et al.* Changing glucocorticoid action: 11beta-hydroxysteroid dehydrogenase type 1 in acute and chronic inflammation. *J. Steroid Biochem. Mol. Biol.* **137**, 82–92 (2013).
130. Hardy, M. P. *et al.* Stress hormone and male reproductive function. *Cell Tissue Res.* **322**, 147–153 (2005).
131. Morton, N. M. *et al.* Improved lipid and lipoprotein profile, hepatic insulin sensitivity, and glucose tolerance in 11beta-hydroxysteroid dehydrogenase type 1 null mice. *J. Biol. Chem.* **276**, 41293–41300 (2001).
132. Draper, N. & Stewart, P. M. 11beta-hydroxysteroid dehydrogenase and the pre-receptor regulation of corticosteroid hormone action. *J. Endocrinol.* **186**, 251–271 (2005).
133. Kotelevtsev, Y. *et al.* 11β-hydroxysteroid dehydrogenase type 1 knockout mice show attenuated glucocorticoid-inducible responses and resist hyperglycemia on obesity or stress. *Proc. Natl Acad. Sci. USA* **94**, 14924–14929 (1997).
**This key review describes the impact of oxidative stress in CKD.**
134. You, Y. H., Quach, T., Saito, R., Pham, J. & Sharma, K. Metabolomics reveals a key role for fumarate in mediating the effects of NADPH oxidase 4 in diabetic kidney disease. *J. Am. Soc. Nephrol.* **27**, 466–481 (2016).
135. Li, M. *et al.* GC/TOFMS analysis of metabolites in serum and urine reveals metabolic perturbation of TCA cycle in db/db mice involved in diabetic nephropathy. *Am. J. Physiol. Renal Physiol.* **304**, F1317–F1324 (2013).
136. Hasselwander, O. & Young, I. S. Oxidative stress in chronic renal failure. *Free Radic. Res.* **29**, 1–11 (1998).
137. Himmelfarb, J., Stenvinkel, P., Ikizler, T. A. & Hakim, R. M. The elephant in uremia: oxidant stress as a unifying concept of cardiovascular disease in uremia. *Kidney Int.* **62**, 1524–1538 (2002).
138. Chen, H. *et al.* Metabolomics insights into activated redox signaling and lipid metabolism dysfunction in chronic kidney disease progression. *Redox Biol.* **10**, 168–178 (2016).
139. Mashima, R., Nakanishi-Ueda, T. & Yamamoto, Y. Simultaneous determination of methionine sulfoxide and methionine in blood plasma using gas chromatography-mass spectrometry. *Anal. Biochem.* **313**, 28–33 (2003).
140. Breier, M. *et al.* Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. *PLoS ONE* **9**, e89728 (2014).
141. Simic-Ogrizovic, S. *et al.* Markers of oxidative stress after renal transplantation. *Transpl. Int.* **11** (Suppl. 1), S125–S129 (1998).
142. Fonseca, I. *et al.* Oxidative stress in kidney transplantation: malondialdehyde is an early predictive marker of graft dysfunction. *Transplantation* **97**, 1058–1065 (2014).
143. Akanuma, H., Ogawa, K., Lee, Y. & Akanuma, Y. Reduced levels of plasma 1,5-anhydroglucitol in diabetic patients. *J. Biochem.* **90**, 157–162 (1981).
144. Kametani, S., Hashimoto, Y., Yamanouchi, T., Akanuma, Y. & Akanuma, H. Reduced renal reabsorption of 1,5-anhydro-D-glucitol in diabetic rats and mice. *J. Biochem.* **102**, 1599–1607 (1987).

145. Yu, B. *et al.* Serum metabolomic profiling and incident CKD among African Americans. *Clin. J. Am. Soc. Nephrol.* **9**, 1410–1417 (2014).
146. Dooms-Goossens, A., Degreef, H., Vanhee, J., Kerkhofs, L. & Chrispeels, M. T. Chlorocresol and chloracetamide: allergens in medications, glues, and cosmetics. *Contact Dermatitis* **7**, 51–52 (1981).
147. Sanders, J. M. *et al.* Carcinogenesis studies of cresols in rats and mice. *Toxicology* **257**, 33–39 (2009).
148. Ford, S. M., Hook, J. B. & Bond, J. T. The effects of butylated hydroxyanisole and butylated hydroxytoluene on renal function in the rat. I. Effects on fluid and electrolyte excretion. *Food Cosmet Toxicol.* **18**, 15–20 (1980).
149. De Smet, R. *et al.* Toxicity of free p-cresol: a prospective and cross-sectional analysis. *Clin. Chem.* **49**, 470–478 (2003).
150. Lin, H. H., Huang, C. C., Lin, T. Y. & Lin, C. Y. p-Cresol mediates autophagic cell death in renal proximal tubular cells. *Toxicol. Lett.* **234**, 20–29 (2015).
151. Brocca, A., Virzi, G. M., de Cal, M., Cantaluppi, V. & Ronco, C. Cytotoxic effects of p-cresol in renal epithelial tubular cells. *Blood Purif.* **36**, 219–225 (2013).
152. Adamski, J. Genome-wide association studies with metabolomics. *Genome Med.* **4**, 34 (2012).
**An introduction to metabolomics for genome-wide association studies.**
153. Sekula, P. *et al.* A metabolome-wide association study of kidney function and disease in the general population. *J. Am. Soc. Nephrol.* **27**, 1175–1188 (2016).
154. Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).
155. Rhee, E. P. *et al.* A combined epidemiologic and metabolomic approach improves CKD prediction. *J. Am. Soc. Nephrol.* **24**, 1330–1338 (2013).
156. Gao, X. *et al.* Systematic variations associated with renal disease uncovered by parallel metabolomics of urine and serum. *BMC Syst. Biol.* **6** (Suppl. 1), S14 (2012).
157. Mutsaers, H. A. *et al.* Optimized metabolomic approach to identify uremic solutes in plasma of stage 3–4 chronic kidney disease patients. *PLoS ONE* **8**, e71199 (2013).
158. Hao, X. *et al.* Distinct metabolic profile of primary focal segmental glomerulosclerosis revealed by NMR-based metabolomics. *PLoS ONE* **8**, e78531 (2013).
159. Posada-Ayala, M. *et al.* Identification of a urine metabolomic signature in patients with advanced-stage chronic kidney disease. *Kidney Int.* **85**, 103–111 (2014).
160. Pena, M. J. *et al.* Urine and plasma metabolites predict the development of diabetic nephropathy in individuals with Type 2 diabetes mellitus. *Diabet. Med.* **31**, 1138–1147 (2014).
161. Sharma, K. *et al.* Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. *J. Am. Soc. Nephrol.* **24**, 1901–1912 (2013).
162. van der Kloet, F. M. *et al.* Discovery of early-stage biomarkers for diabetic kidney disease using ms-based metabolomics (FinnDiane study). *Metabolomics* **8**, 109–119 (2012).
163. Lee, J. *et al.* Changes in serum metabolites with the stage of chronic kidney disease: comparison of diabetes and non-diabetes. *Clin. Chim. Acta* **459**, 123–131 (2016).
164. Gao, Y. *et al.* Identifying early urinary metabolic changes with long-term environmental exposure to cadmium by mass-spectrometry-based metabolomics. *Environ. Sci. Technol.* **48**, 6409–6418 (2014).
165. Uehara, T. *et al.* Identification of metabolomic biomarkers for drug-induced acute kidney injury in rats. *J. Appl. Toxicol.* **34**, 1087–1095 (2014).
166. Sun, X. *et al.* Metabonomics evaluation of urine from rats administered with phorate under long-term and low-level exposure by ultra-performance liquid chromatography-mass spectrometry. *J. Appl. Toxicol.* **34**, 176–183 (2014).
167. Zinellu, A. *et al.* Impact of cholesterol lowering treatment on plasma kynurenine and tryptophan concentrations in chronic kidney disease: relationship with oxidative stress improvement. *Nutr. Metab. Cardiovasc. Dis.* **25**, 153–159 (2015).
**Study showing that amelioration of both oxidative and inflammation status after cholesterol-lowering treatment in CKD might be mediated by restoration of antioxidant taurine concentrations during therapy.**
168. Vanhaelen, M., Vanhaelen-Fastre, R., But, P. & Vanherweghem, J. L. Identification of aristolochic acid in Chinese herbs. *Lancet* **343**, 174 (1994).
169. Zhao, Y. Y. *et al.* Urinary metabolomics and biomarkers of aristolochic acid nephrotoxicity by UPLC-QTOF/HDMS. *Bioanalysis* **7**, 685–700 (2015).
170. Zhao, Y. Y. *et al.* Metabolomics analysis reveals the association between lipid abnormalities and oxidative stress, inflammation, fibrosis, and Nrf2 dysfunction in aristolochic acid-induced nephropathy. *Sci. Rep.* **5**, 12936 (2015).
171. Zhang, X. *et al.* Plasma metabolic profiling analysis of nephrotoxicity induced by acyclovir using metabonomics coupled with multivariate data analysis. *J. Pharm. Biomed. Anal.* **97**, 151–156 (2014).
172. Hanna, M. H. *et al.* Urinary metabolomic markers of aminoglycoside nephrotoxicity in newborn rats. *Pediatr. Res.* **73**, 585–591 (2013).
173. Wei, Q., Xiao, X., Fogle, P. & Dong, Z. Changes in metabolic profiles during acute kidney injury and recovery following ischemia/reperfusion. *PLoS ONE* **9**, e106647 (2014).
174. Zhao, X., Chen, J., Ye, L. & Xu, G. Serum metabolomics study of the acute graft rejection in human renal transplantation based on liquid chromatography-mass spectrometry. *J. Proteome Res.* **13**, 2659–2667 (2014).
175. Wettersten, H. I. *et al.* Grade-dependent metabolic reprogramming in kidney cancer revealed by combined proteomics and metabolomics analysis. *Cancer Res.* **75**, 2541–2552 (2015).
176. Floegel, A. *et al.* Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **62**, 639–648 (2013).
177. Felig, P., Wahren, J., Sherwin, R. & Palaiologos, G. Amino acid and protein metabolism in diabetes mellitus. *Arch. Intern. Med.* **137**, 507–513 (1977).
178. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17**, 448–453 (2011).
179. Rhee, E. P. *et al.* Metabolomics of chronic kidney disease progression: a case-control analysis in the chronic renal insufficiency cohort study. *Am. J. Nephrol.* **43**, 366–374 (2016).
180. Kimura, T. *et al.* Identification of biomarkers for development of end-stage kidney disease in chronic kidney disease by metabolomic profiling. *Sci. Rep.* **6**, 26138 (2016).
181. Kobayashi, T. *et al.* A metabolomics-based approach for predicting stages of chronic kidney disease. *Biochem. Biophys. Res. Commun.* **445**, 412–416 (2014).
182. Nkuipou-Kenfack, E. *et al.* Assessment of metabolomic and proteomic biomarkers in detection and prognosis of progression of renal function in chronic kidney disease. *PLoS ONE* **9**, e96955 (2014).
183. Mussap, M., Antonucci, R., Noto, A. & Fanos, V. The role of metabolomics in neonatal and pediatric laboratory medicine. *Clin. Chim. Acta* **426**, 127–138 (2013).
184. Yin, P., Lehmann, R. & Xu, G. Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal. Bioanal. Chem.* **407**, 4879–4892 (2015).
185. Bernini, P. *et al.* Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *J. Biomol. NMR* **49**, 231–243 (2011).
186. Anton, G. *et al.* Pre-analytical sample quality: metabolite ratios as an intrinsic marker for prolonged room temperature exposure of serum samples. *PLoS ONE* **10**, e0121495 (2015).
**Analysis of the impact of pre-analytical sample handling on metabolomic fingerprinting.**
187. Herberth, G. *et al.* Endogenous metabolites and inflammasome activity in early childhood and links to respiratory diseases. *J. Allergy Clin. Immunol.* **136**, 495–497 (2015).
188. Viswan, A., Sharma, R. K., Azim, A. & Sinha, N. NMR-based metabolic snapshot from minibronchoalveolar lavage fluid: an approach to unfold human respiratory metabolomics. *J. Proteome Res.* **15**, 302–310 (2016).
189. Tsentalovich, Y. P. *et al.* Metabolomic composition of normal aged and cataractous human lenses. *Exp. Eye Res.* **134**, 15–23 (2015).
190. Courant, F., Antignac, J. P., Monteau, F. & Le Bizec, B. Metabolomics as a potential new approach for investigating human reproductive disorders. *J. Proteome Res.* **12**, 2914–2920 (2013).
191. Lutz, U., Lutz, R. W. & Lutz, W. K. Metabolic profiling of glucuronides in human urine by LC-MS/MS and partial least-squares discriminant analysis for classification and prediction of gender. *Anal. Chem.* **78**, 4564–4571 (2006).
192. Fonville, J. M., Richards, S. E., Barton, R. H., Boulange, C. L. & Ebbels, T. M. D. The evolution of partial least square models and related chemometric approaches in metabonomics and metabolite phenotyping. *J. Chemom.* **24**, 636–649 (2010).
193. Trygg, J. & Wold, S. Orthogonal projections to latent structures (OPLS). *J. Chemom.* **16**, 119–128 (2002).



### Acknowledgements
The authors' work is supported by Innovative Medicines Initiative Joint Undertaking under Grant agreement number 115439 (StemBANCC), and number 115317 (DIRECT), the German Federal Ministry of Education and Research (BMBF) to the German Center Diabetes Research (DZD e.V.) grant to J.A. and the Deutsche Forschungsgemeinschaft (DFG) number Ho 1665/5-1, Ho 1665/5-2 and Ho 1665/5-3 to B.H. The authors would like to express their gratitude for critical reading of the manuscript to Dr. Cornelia Prehn and Dr. Alexander Cecil at Helmholtz Zentrum München, Genome Analysis Center, Neuherberg, Germany.


### Author contributions
Both authors researched the data for the article, discussed the content, wrote the text and revised or edited the manuscript before submission.

### Competing interests statement
The authors declare no competing interests.

**SUPPLEMENTARY INFORMATION**
**See online article:** S1 (figure) | S2 (table)
**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**