Proteomics and Metabolomics in Kidney Disease, including Insights into Etiology, Treatment, and Prevention

Ruth F. Dubin¹ and Eugene P. Rhee²

Abstract
In this review of the application of proteomics and metabolomics to kidney disease research, we review key concepts, highlight illustrative examples, and outline future directions. The proteome and metabolome reflect the influence of environmental exposures in addition to genetic coding. Circulating levels of proteins and metabolites are dynamic and modifiable, and thus amenable to therapeutic targeting. Design and analytic considerations in proteomics and metabolomics studies should be tailored to the investigator’s goals. For the identification of clinical biomarkers, adjustment for all potential confounding variables, particularly GFR, and strict significance thresholds are warranted. However, this approach has the potential to obscure biologic signals and can be overly conservative given the high degree of intercorrelation within the proteome and metabolome. Mass spectrometry, often coupled to up-front chromatographic separation techniques, is a major workhorse in both proteomics and metabolomics. High-throughput antibody- and aptamer-based proteomic platforms have emerged as additional, powerful approaches to assay the proteome. As the breadth of coverage for these methodologies continues to expand, machine learning tools and pathway analyses can help select the molecules of greatest interest and categorize them in distinct biologic themes. Studies to date have already made a substantial effect, for example elucidating target antigens in membranous nephropathy, identifying a signature of urinary peptides that adds prognostic information to urinary albumin in CKD, implicating circulating inflammatory proteins as potential mediators of diabetic nephropathy, demonstrating the key role of the microbiome in the uremic milieu, and highlighting kidney bioenergetics as a modifiable factor in AKI. Additional studies are required to replicate and expand on these findings in independent cohorts. Further, more work is needed to understand the longitudinal trajectory of select protein and metabolite markers, perform transomics analyses within merged datasets, and incorporate more kidney tissue–based investigation.


Introduction
As detailed in other reviews in this series, there has been tremendous progress in our ability to analyze the approximately 3 billion DNA base pairs of the human genome (or informative subsets of the genome) in large patient populations, with cascading benefits on our understanding of kidney disease pathophysiology, risk prediction, and treatment. By interrogating molecular domains downstream from the genome, proteomics and metabolomics have the potential to build on these advances, as well as add novel insights. Because the proteome and metabolome represent the summative effects of gene function, they are sometimes classified within the overarching field of “functional genomics,” which aims to understand the relationship between genotype and phenotype on a genome-wide scale (1). All of these “omics” approaches embrace an unbiased examination of human health and disease, but assay methodologies can vary widely. Further, although there are shared analytic approaches to large datasets regardless of molecule type, there are also areas of conceptual and operational divergence.

Key Concepts in Metabolomics and Proteomics in Kidney Disease
The fundamental flow of information in biologic systems is from DNA (genome) to RNA (transcriptome) to proteins (proteome) to metabolites (metabolome) (Figure 1). Although somatic mutations certainly occur (and are of critical importance in cancers), from the standpoint of kidney disease, DNA sequence can largely be viewed as static and uniform throughout the body. Thus, an assessment of the kidney genome does not require sequencing of kidney DNA, but rather can be inferred from DNA analysis of any other source (2). By contrast, the proteome and metabolome are highly dynamic—indeed, this is a major reason why an examination of these domains is critical for understanding the “functional” consequences of the genome. Further, the proteome and metabolome differ throughout the body, such that proteomic or metabolomic analysis of liver, muscle, kidney, blood, urine, etc. will yield different results (3,4). Heterogeneity also exists within tissues, for example across glomerular, endothelial, and tubular cells in the kidney (5). Thus, examination of the proteome and metabolome has clear advantages: compared with the genome, the
In considering the effect of kidney function on the proteome and metabolome, one must also understand the limitations of adjusting for GFR in statistical analyses. On one hand, GFR (even if directly measured) does not account for the full breadth of kidney function, such as tubular absorption or secretion. Thus, adjusting for GFR has the potential to obscure biologically meaningful signals. For example, a molecule might have an inverse correlation with GFR because it is nephrotoxic and causes GFR decline. This possibility is arguably more likely for circulating proteins than metabolites, as the proteome subsumes a broad range of biologic domains including inflammation, thrombosis, metabolism, cell growth, and differentiation, etc. Circulating enzymes, cytokines, and hormones are all attractive as potential participants in disease pathogenesis. Notably, recent studies have begun...
to assign unanticipated functional roles for select circulating metabolites, sometimes acting through previously "orphan" G-protein–coupled receptors (14). Many of these findings appear anchored around the crosstalk of gut microbes and host immunity and metabolism; given their relevance to inflammation, BP, and energy utilization, the metabolites in this subset are potential causal mediators of kidney disease and its complications.

**Assay Methodology**

Whereas nucleic acids comprise four to five similar chemical motifs, proteins and metabolites span a broad range of size, polarity, and structure, raising considerable analytic challenges. For a detailed review of proteomic and metabolomic methodologies, the reader is referred to other in-depth reviews (15–18). Mass spectrometry (MS), which differentiates ions—for example peptide fragments or metabolites—on the basis of mass-to-charge ratio, has been a major workhorse for both proteomics and metabolomics. Typically, MS is coupled to up-front separation techniques, such as liquid chromatography (LC) or gas chromatography, that can have a substantial effect on which analytes are ultimately measured. Two-dimensional gel electrophoresis has also been used extensively in proteomics, whereas nuclear magnetic resonance spectrometry has been an important alternative for metabolomics. For MS, proteins need to be enzymatically digested into multiple smaller peptides and peptide fragments before analysis. This requirement for complex preanalytic sample preparation has historically limited the throughput of MS-based proteomics. More recently, highly multiplexed methods that capture and detect hundreds or thousands of proteins have been developed (15,19). Specific protein capture is achieved using nucleotide-labeled antibodies or aptamers (nucleic acid–based), and then detection and quantitation are on the basis of nucleic acid oligomerization. The breadth, sensitivity, and throughput of these approaches make them powerful research tools that can be applied to large study populations. For both proteomics and metabolomics, assay validation is important. Traditionally, this has involved using an orthogonal approach to confirm top hits, for example ELISA for proteins or tandem MS for metabolites. Because proteomic and metabolomic approaches usually provide information on relative protein or metabolite levels, rather than absolute concentrations, orthogonal methods may also be required for absolute quantitation.

**Analytic Considerations**

As with genomics, correction for multiple hypothesis testing is required in proteomics and metabolomics to reduce the risk of false discovery. However, there is no uniformly applied threshold analogous to the $P<5\times10^{-8}$ genome-wide threshold, in part because the actual sizes of the proteome and metabolome are uncertain (and contingent on whether one includes post-translational modifications, xenobiotics, etc.) and different platforms provide incomplete and only partially overlapping coverage of these molecular domains. This latter limitation has made independent replication across cohorts and meta-analysis more challenging in proteomics and metabolomics than genomics. For discovery analyses, most investigators have favored using either Bonferroni adjustment or false discovery rate, for example using the Benjamini–Hochberg procedure (20). Given the intercorrelation of many analytes in the context of biologic pathways, however, these approaches may be overly conservative and, as with adjustment for GFR, have the potential to obscure biologic signals. If the goal is to identify individual biomarkers with robust disease associations above and beyond other established predictors, then fully adjusted models and strict significance thresholds are ideal. A focus on biologic discovery may warrant more flexibility (Figure 3).

Machine learning and pathway analyses are potent methods to help organize and decode omics datasets. Machine learning methods such as least absolute shrinkage and selection operator, elastic net, and random forest regression each have different strengths. In omics datasets, many variables may be collinear, and random forest regression may handle this better than other methods. If the goal is to create a parsimonious risk model, least absolute shrinkage and selection operator or elastic net may be more suitable. Support-vector machines or clustering are methods that use algorithms to categorize variables in supervised or unsupervised machine learning. These methods can be tuned to select the number of “top hits” appropriate to the research

**Figure 2.** Kidney arteriovenous gradients for the circulating metabolome and proteome. Panels show median venous-to-arterial (V/A) ratio for approximately 300 nolipid metabolites measured by liquid chromatography–mass spectrometry and approximately 1000 proteins measured by aptamer in samples obtained from ten individuals. PTH, parathyroid hormone.
A handful of top hits may be ideal when designing a biomarker risk panel, whereas several hundred proteins or metabolites will be more amenable to pathway analyses. Pathway analysis is a broad term encompassing several methods to organize large datasets into functionally distinct biologic themes (21,22). These methods are more established and have been used more widely in proteomics (and genomics) than in metabolomics. There are many publicly available software programs for pathway analysis, including Gene Set Enrichment Analysis; Database for Annotation, Visualization, and Integrated Discovery; and Search Tool for the Retrieval of Interacting Genes/Proteins. Ingenuity Pathway Analysis is also widely used and requires a private license.

Illustrative Examples in the Literature

Application of Proteomics to Understanding Biology of Membranous Nephropathy

The story of the M-type phospholipase A(2) receptor (PLA2R) illustrates the potential for proteomic discovery to inform pathobiology and guide diagnosis and treatment. On the basis of its histopathology, primary membranous nephropathy was long recognized as an antibody-mediated autoimmune glomerular disease, but the target antigen was unknown. Beck et al. performed western blots of protein extracts of normal human glomeruli with serum from patients with primary membranous nephropathy and secondary membranous nephropathy and from normal controls. The investigators identified a 185-kDa reactive protein band specific to primary membranous nephropathy and then applied liquid chromatography–mass spectrometry (LC-MS) to the reactive protein band to detect PLA2R. Further, they showed that PLA2R was expressed in podocytes from healthy kidneys and that antibodies to PLA2R were found in the glomerular immune deposits of patients with primary membranous nephropathy (23). A similar approach, western blotting followed by LC-MS, was used to identify thrombospondin type-1 domain-containing 7A (THSD7A) as an additional target antigen for autoantibodies in some patients with anti-PLA2R-negative membranous nephropathy (24). Since these landmark findings, autoantibodies to PLA2R have been found to account for approximately 70%–80% of cases of primary membranous nephropathy, with autoantibodies to THSD7A accounting for a much smaller fraction (25). These discoveries have transformed clinical care, because serum autoantibody levels have been found to correlate with membranous nephropathy disease activity and response to immunosuppression (26,27). A recent exploratory study applied LC-MS–based proteomics to serum from patients with idiopathic membranous nephropathy who did or did not achieve complete remission on calcineurin inhibitors, and identified serum amyloid A1 protein as a candidate marker that may provide further prognostic information on top of PLA2R antibody levels (28).

Utility of Proteomics for Predicting CKD Incidence and Prognosis

Whereas the example of membranous nephropathy demonstrates how proteomic discovery can focus interest
on one or two specific proteins, other studies have considered a wider array of molecules. Mischak and colleagues (29) have endorsed the value of combining multiple peptide or protein biomarkers, on the premise that the high variability of single biomarkers limits accuracy. Using capillary electrophoresis–MS, these investigators compared urine from 230 patients with a range of CKD causes and 379 controls with normal kidney function. Cluster analysis by support vector machines was used to combine 273 urinary peptides that differed between CKD and healthy controls into an aggregate urinary peptide marker termed CKD273 (30). Subsequent studies that included several thousand patient samples have shown that CKD273 may be of higher utility than standard metrics for detection and prediction of CKD (31,32). The US Food and Drug Administration has issued a letter of support encouraging further studies with CKD273 as a biomarker panel in early phase clinical trials “to determine its clinically utility for prognostic enrichment, drug development decisions, and study design considerations,” but also noting that CKD273 does not appear to perform well in prediction of progression to ESKD or doubling of serum creatinine in patients with more advanced disease (29).

Major components of CKD273 include collagen fragments, perhaps indicative of accumulation of intrakidney extracellular matrix, along with fragments of blood-derived proteins involved in inflammation and tissue repair. Notably, chronic inflammation has drawn considerable interest as a risk marker and potential etiologic factor in diabetic kidney disease progression. For example, Niewczas and colleagues have highlighted a strong association between circulating TNF receptors 1 and 2 (TNF-R1 and TNF-R2) and the risk of ESKD in diabetic kidney disease (33). To more comprehensively assess inflammation in diabetic kidney disease, these investigators recently used an aptamer-based approach to measure a panel of 194 proteins that includes most of the circulating inflammatory proteins known in the literature and most proteins previously studied in diabetic kidney disease (34). Discovery analysis was performed in 219 subjects with type 1 diabetes from the Joslin Kidney Study, followed by validation in 144 subjects with type 2 diabetes from the Joslin Kidney Study and replication in 162 subjects with type 2 diabetes in the Pima Indian Study. This study identified a Kidney Risk Inflammatory Signature (KRIS), comprised of 17 inflammatory proteins enriched for TNF-R superfamily members, significantly associated with the 10-year risk of ESKD. The investigators also assessed KRIS proteins in urine from a small case-control analysis of rapid GFR decline and found that most were elevated in cases and none were reduced. Further, they found no significant correlation between the majority of circulating KRIS protein levels and their corresponding gene expression in kidney biopsy specimens obtained from 56 Pima Indians. On the basis of these observations, the authors posit that nonkidney sources (such as leukocytes) play an important role in generating circulating KRIS proteins, and perhaps as etiologic drivers of diabetic kidney disease.

**Metabolomics and the Contribution of the Gut Microbiome to Uremia**

Many metabolomics studies have examined blood or urine with the goal to identify metabolite alterations associated with CKD or to identify metabolite predictors of CKD progression. To date, no single metabolite has emerged as an unequivocal causal factor in the pathogenesis of CKD or its complications; this relates in part to the nonoverlap across platforms noted above and the heterogeneity of study populations, but also to the sheer number of metabolomic perturbations that manifest as kidney clearance is lost. However, one major theme that has emerged is the fundamental contribution made by the gut microbiome to the blood metabolome in CKD. For example, Meyer and colleagues (35) used LC-MS to profile blood obtained from nine patients on hemodialysis with intact colons and six patients on hemodialysis who had previously undergone colectomy. Compared with patients with ESKD with intact colons, patients with ESKD who had undergone colectomy had markedly lower levels of at least 35 metabolites, including indoxyl sulfate and p-cresol sulfate. Importantly, the levels of these colon-derived metabolites were substantially higher in the patients with ESKD than in individuals with normal kidney function, and the majority were not effectively removed by conventional hemodialysis. Abe and colleagues (36) have raised specific interest in phenyl sulfate, a tyrosine metabolite derived from gut microbes, as a potential causal factor in diabetic kidney disease. In patients with diabetes, they found that phenyl sulfate levels correlate with albuminuria, and, further, that phenyl sulfate administration and inhibition of the bacterial enzyme responsible for its synthesis induce and ameliorate albuminuria in mice, respectively.

Hazen and colleagues have focused on a specific gut-derived metabolite, trimethylamine-N-oxide (TMAO), as a potential causal factor in cardiovascular disease. Initial interest in TMAO was motivated by an LC-MS metabolomics analysis of blood samples obtained from individuals with prevalent cardiovascular disease, with subsequent studies showing that elevated TMAO is associated with new-onset cardiovascular events (37,38). The requirement for gut flora on blood TMAO was demonstrated with elegant feeding studies before and after gut flora suppression using antibiotics (37,38), and a functional role for TMAO in atherogenesis has been posited on the basis of its ability to inhibit reverse cholesterol transport, increase platelet hyperresponsiveness, and promote macrophage foam cell formation (39,40). In mice, dietary TMAO loading can also cause tubular injury and kidney fibrosis (41). Because TMAO is primarily excreted via the kidney and levels increase with CKD, several groups have examined its association with cardiovascular disease among individuals with CKD and ESKD, with mixed results (41–44). Whether or not molecules such as phenyl sulfate and TMAO prove to be important causal factors in CKD, these studies underscore the ability of metabolomics to integrate both endogenous and exogenous outputs, with the latter including contributions from diet and the microbiome.

**Application of Metabolomics for Translational AKI Research**

Parikh and colleagues (45) applied LC-MS–based metabolomics to examine endogenous kidney metabolism in mice with loss- and gain-of-function of Peroxisome proliferator-activated receptor γ coactivator 1-α, which they have shown are sensitized and protected from AKI, respectively. They found that niacinamide levels were lower in kidneys
from PGC1α-deficient animals and that niacinamide levels fell in response to AKI. By contrast, niacinamide levels were higher in transgenic PGC1α animals and niacinamide supplementation was renoprotective in different injury models, in part by rescuing AKI-induced impairment in lipid metabolism. Niacinamide is a precursor in the salvage pathway of NAD+, an essential cofactor for energy metabolism that is a key determinant of cellular health across a range of tissues. Parikh and colleagues (46) have used LC-MS to demonstrate that urinary metabolites indicative of reduced de novo NAD+ biosynthesis are associated with AKI risk in humans, and, further, have generated pilot data raising the possibility that NAD+ boosting with oral niacinamide supplementation may attenuate kidney injury in the context of cardiac surgery. These promising initial findings underscore an important advantage of metabolomics-based discovery, the potential for rapid translation. More specifically, many metabolites are readily delivered in vivo, and their cognate enzymes, transporters, and receptors are all potential targets for small-molecule agonism or antagonism.

**Future Directions**

Although progress has been made, more work is required to harness the full potential of proteomics and metabolomics in the study of kidney disease. In one example, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) CKD Biomarkers Consortium (www.ckdbiomarkersconsortium.org) is actively generating blood proteomic and metabolomic data in >5000 individuals with CKD, spanning several different study cohorts, including both adults and children (47). In addition to enhancing statistical power compared with prior CKD studies, this coordinated effort will enable examination of different outcomes such as CKD progression, cardiovascular events, and cognitive development, as well as different “subphenotypes,” i.e., looking at patient subsets on the basis of diagnosis (48), comorbidity, or other differentiating features. Because some proteomics and metabolomics data will be generated in a common sample set that has been genotyped, there will also be an opportunity for integrative analyses; paradigms exist for genome-wide association studies of the blood proteome or metabolome, whereas efforts to integrate the proteome and metabolome are less developed.

Even as proteomic and metabolomic studies in kidney disease continue to expand, important opportunities remain. To date, there is a relative paucity of data in large cohorts with repeated measurements over time that show the longitudinal trajectory of the proteome and metabolome. Further, the spatial heterogeneity of the proteome and metabolome mean that an examination of blood or urine provides only limited information about intraorgan biology. The NIDDK Kidney Precision Medicine Project (www.kpmp.org) is anticipated to provide much-needed human kidney tissues for biomarker research, and this program and more mechanistic investigation in model systems will be critical to determine both why select proteins or metabolites are altered in kidney disease and whether they play a causal role in disease pathogenesis. Recent advances in identifying post-translational modifications of proteins will enhance our understanding of the proteome in kidney disease (49). In parallel, the integration of genomics with proteomics and metabolomics may permit Mendelian randomization–based approaches to test whether select markers lie in causal pathways (50). These efforts will require vigorous collaboration across disciplines, and, taken together, will build on the momentum of genomic discovery in nephrology research to further advance our understanding and therapeutic approaches to kidney disease.

**Disclosures**

Dr. Dubin and Dr. Rhee have nothing to disclose.

**Funding**

This work was supported by National Institute of Diabetes and Digestive and Kidney Diseases grants U01DK108809 and U01DK106981 and National Institute of Nursing Research grant R01NR017399.

**References**


Published online ahead of print. Publication date available at www.cjasn.org.