CLINICAL PROTEOMICS: THE POTENTIALITY OF URINE ANALYSIS FOR UNDERSTANDING DIABETIC NEPHROPATHY

Massimo Papale,1 Maria Teresa Rocchetti,1,2 Loreto Gesualdo2

1. Core Facility of Proteomics and Mass Spectrometry, Department of Surgery and Medical Sciences, University of Foggia, Italy
2. Department of Emergency and Organ Transplantation, University of Bari, Italy

Disclosure: No potential conflict of interest.
Citation: EMJ Neph. 2013;1:32-39.

ABSTRACT

The incidence of diabetic nephropathy (DN) is constantly rising in parallel with the prevalence of type 2 diabetes and has been predicted to double within the next 15 years. Albuminuria is considered the earliest putative diagnostic sign of diabetic renal damage but it is poorly associated to the complex histopathological picture of glomerular and tubular damage hence, up to now, the accurate diagnosis of the DN requires renal biopsy. The identification of new biomarkers of DN is an urgent need since the proper management of the DN patients requires early and unbiased diagnosis. The Proteomics approach to the study of the human disease allows a large-scale characterisation of the protein content of a biological sample, and its application to urine may be a challenging but powerful strategy to identify new DN biomarkers. In this review we discuss the main results of a decade of proteomic studies focused on the urinary investigation of diabetic nephropathy.

Keywords: Diabetic nephropathy, urinary proteome, proteomics, urine, biomarkers.

THE PATHOPHYSIOLOGY OF DIABETES AND DIABETIC NEPHROPATHY

Diabetic nephropathy (DN) is the most common chronic kidney disease (CKD) in developed countries1 and the most frequent cause of end-stage renal disease (ESRD) worldwide. It has been estimated that 40% of the patients undergoing renal dysfunction and that require renal replacement therapy are affected by DN.2 DN is a severe complication of both type 2 diabetes mellitus (T2DM) and type 1 diabetes (T1D), but the incidence of nephropathy is more prevalent in T1D primarily due to the fact that, in T2DM, death as a result of cardiovascular causes is more common than death from renal failure.3,4 The use of renin–angiotensin system inhibitors and strict glycemic control is contributing to slow the incidence of ESRD in T2DM patients.5 However, between 2000 and 2030, the prevalence of T2DM has been predicted to increase by 20% in developed countries and about 50-70% in developing ones.6 This will lead to an increase of the incidence of ESRD,7,8 concomitantly with the progressively declining rate of mortality due to cardiovascular causes.9,10 The primum movens of T2DM complication is chronic hyperglycaemia, which initiates specific modifications of the electron transport proteins by advanced glycation end-products (AGEs) and alters normal metabolism by increasing production of reactive oxygen species (ROS).11 Hyperglycaemia and increased ROS production alter cell homeostasis in endothelium and renal cells and impair endothelial nitric oxide synthase and prostacyclin synthase, that, in turn, contribute to defective angiogenesis and persistent expression of pro-inflammatory genes, also after glycaemia normalisation.12 These factors, together with genetic background and lifestyle, may predispose a considerable number of T2DM patients to develop DN.

The pathogenesis of DN involves structural changes, including glomerular and tubular hypertrophy, with
progressive accumulation of extracellular matrix components in the glomerular mesangium and tubulointerstitium, and changes in podocytes.\textsuperscript{13-15} According to the most recent pathologic classification of DN,\textsuperscript{16} the severity of the glomerular lesions correlates with the progression of the DN and may allow four classes to be distinguished, namely: class I (glomerular membrane basement thickening); class II (mesangial expansion without Kimmelstiel-Wilson lesions); class III (presence of at least one glomerulus presenting Kimmelstiel-Wilson lesions) and class IV (Kimmelstiel-Wilson lesions in at least 50% of the glomeruli).

At urinary level, microalbuminuria (urine albumin excretion 30-300 mg/24h) is considered the earliest putative diagnostic sign of diabetic renal damage even if it may not correlate with the complex histopathological picture of glomerular and tubular damage in T2DM.\textsuperscript{17} In fact, it is not always associated with the presence of Kimmelstiel-Wilson nodules when renal biopsies are examined,\textsuperscript{18} thus representing a better predictor of cardiovascular disease than of renal damage progression.\textsuperscript{19} Further to this, urine contains more than 60 forms or fragments of albumin,\textsuperscript{20} which are not all recognised by the routinely immunoassay-based methods that ultimately may underestimate the correlation between the albuminuria and the renal damage. Due to the complexity of DN pathophysiology it is necessary to set up unbiased methods that can simultaneously detect new sets of biomarkers for earlier diagnosis and prognosis of DN.\textsuperscript{21}

The development of renal damage in T2DM patients is antedated and/or accompanied by a number of molecular changes that may be now identified by a number of high-throughput strategies. These include the next generation sequencing (NGS) approaches for complete sequencing of whole genomes,\textsuperscript{22} transcriptomes,\textsuperscript{23} and epigenetic DNA modifications,\textsuperscript{24} and also proteomic and metabolomics strategies for accurate measurement of the entire content of proteins and metabolites of biological samples. The aim of the present review is to provide a concise overview of the main contributions of the proteome science to the identification of a set of new urinary biomarkers that could help in achieving early diagnosis and better management of DN.

![Figure 1. Workflow of the biomarker discovery strategy by hightroughput proteomic analysis.](image)

The complex datasets generated by the high-throughput analysis may allow identification of thousands of exploratory biomarkers. The bioinformatics management is critically required to select, among the exploratory biomarkers, the disease-correlated ones (qualified biomarkers). The management of the qualified biomarkers by means of supervised statistical methods is then essential to setup new classificatory models useful for the diagnosis and prognosis of the diseases. Finally this multi-markers pattern should be validated, in multicentric cohorts of patients, by routinely immunoassays in order to verify their usefulness in clinical practice.
The Proteomic Approach To The Study of Renal Diseases

The term 'proteomics' indicates a complex and interdisciplinary matter requiring expertise spanning from chemistry to biology and bioinformatics, in order to reveal the meaning of complex protein datasets of a biological sample in physiological and pathological conditions. The completion of the human genome sequencing together with the exponential development of ionisation sources (i.e. matrix-assisted laser desorption/ionisation [MALDI]) and electrospray ionisation [ESI]) and bioinformatics tools have rapidly provided new technological platforms for the analysis of complex protein datasets and the interpretation of the cross-linked relationship among the differently expressed proteins. Starting from the last decade, proteomics has been exponentially applied to nephrology leading to the identification of a number of putative biomarkers that are expected to enter shortly into the clinical practice, making proteomics a science of key interest not only for researchers but also for clinicians.

The proteomics analysis of biological samples may be pursued by distinct and complementary strategies that allow separating the protein mixtures and identifying the key disease-related molecules by mass spectrometry analysis. Two-dimensional gel electrophoresis (2-DE), the most popular gel-based approach, allows double protein separation according to the isoelectric point (pI) and the molecular mass (MW) and provides, for each sample, a characteristic proteomic map showing the separated proteins as protein spots or spot trains due to the presence of protein post-translational modifications (PTMs).

Comparative software analysis of the 2-DE maps between pathological samples and matching controls may allow identifying differently expressed protein spots that are excised from the gel, trypsin digested to obtain small peptides mixtures, and analysed by mass spectrometry (MALDI-TOF MS, nanoHPLC-ESI-MS/MS) to obtain the protein ID. Even if highly informative, 2-DE proteome underestimates the protein complexity of the sample since, for example, less expressed proteins, proteins having a molecular weight lower than 10 kDa and higher than 250 kDa, and transmembrane (hydrophobic) proteins are difficult to visualise. Although 2-DE is the only tool to depict protein isoforms (train spot), this approach may be laborious and expensive without providing satisfactory results. Usually, 2-DE is appropriate to study a restricted and well-characterised cohort of patients in order to identify putative disease-associated biomarkers, but they need to be further validated in larger cohorts of patients to ascertain their usefulness as disease biomarkers.

The development of a number of so-called profiling technologies has permitted high-throughput analysis of thousands of biological samples and appears to be more appropriate for clinical proteomics studies since they may combine the multicentre collection of numerous samples with their rapid analysis in order to identify a new set of biomarkers applicable to the general population. The profiling technologies include a number of complementary strategies, namely liquid chromatography (LC), capillary electrophoresis (CE), and thin-layer chromatography (TLC) coupled to mass spectrometry (MS). These strategies can identify, in a shortened time, many putative biomarkers ready to be validated. However, the complex datasets generated by these approaches must be properly managed by means of statistical and bioinformatics tools to finally allow the recognition of reliable disease-specific biomarkers before proceeding with their validation.

Recently, the biomarker task force of the National Cancer Institute has developed the guidelines for biomarkers studies that can be extended to any kind of disease. In general, a qualified biomarker must have a clear clinical significance for the disease or a consistent scientific body of evidence must support its probable implication in the pathophysiology of the disease. On the contrary, the disease-associated proteins may be defined as exploratory biomarkers. In order to select the qualified biomarkers among the exploratory, specific bioinformatics tools must be used to select functionally correlated subsets and to evaluate their diagnostic power. The use of bioinformatics software, such as String and Ingenuity, permits a search for the known interactions of any well-characterised protein, and to define a large number of potentially interacting molecules for each protein. This approach may lead to an ever-expanding network of molecular correlations, thus, clinicians having a specific knowledge of the pathophysiology of the disease should always check the appropriateness of each possible interactome in order to restrict the further validation to a sub-set of disease correlated biomarkers. The lack of this essential contribution may prevent the identification of the qualified biomarkers and their use in diagnostics. After the identification, the qualified
biomarkers should be managed through supervised statistical analysis in order to verify if their combined evaluation may allow the creation of proteome-based models useful for improving the diagnostic and prognostic power of each of them.

Briefly, this kind of data analysis uses specific algorithms\textsuperscript{34,35} that verify the best association among the identified biomarkers to recognise the pathologic phenotypes in a “training set” of control and disease samples. The optimal pattern is then validated against an independent “validation set” to confirm its diagnostic utility. The main focus of proteome analysis in nephrology is the identification of biomarkers useful for the prediction of a pathologic phenotype in still asymptomatic patients or for an early and accurate diagnosis to permit rapid and personalised renoprotective treatment. Among biological samples, urine appears the most eligible for identifying kidney biomarkers and therefore most of the clinical proteomics studies in nephrology have been focused on this biological sample. In the next paragraphs we will briefly discuss the main contribution of the urine proteomics to the understanding of DN.

The Urine Proteome: Potentialities and Pitfalls

Many published studies have discussed and emphasised the potentiality of urine as the most appropriate biological fluid for biomarkers discovery in kidney diseases.\textsuperscript{36-42} Some of the well-known and recognisable urine characteristics include: easy, non-invasive accessibility, allowing for multiple and abundant collection; the presence of both kidney-derived (about 70%) and plasma-derived (about 30%) proteins, useful for the identification of both systemic and kidney-specific biomarkers; the lower complexity and increased stability of the urine proteome when compared to that of other biological fluids such as serum and plasma, ensuring the possibility of analysis, and also samples can be collected and subsequently stored for long periods.\textsuperscript{43-45} However, the use of urine for proteomic analysis also has some pitfalls such as the presence of salts and other interfering agents, the higher intra and inter-subject variability,\textsuperscript{19} and in nephropathic patients, the predominant presence of serum proteins like albumin that interfere with the recognition of the lower expressed proteins and may prevent the identification of more sensitive and specific biomarkers.

Since proteomics was firstly applied to the analysis of urine samples, it has been realised that the initial aim of any clinical proteomics study must be the definition of standardised procedures to reduce the effect of confounding factors on the reproducibility of the proteomic data. Our group and other authors have contributed to the realisation of this objective through the publication of a number of methodological works,\textsuperscript{39,45-49} which have allowed for the planning of more accurate biomarker discovery studies in following years. The importance of this aspect is considered a central issue for the nephrology community at national, European (European Kidney and Urine Proteomics (EuroKUP) and international level (Human Kidney and Urine Proteome Project (HKUP) through the creation of groups of study or consortia involved in the standardisation of consensus procedures for collection, storage and analysis of urine by proteomics approaches. It is expected that this attempt to spread a growing awareness of the importance of adopting standardised and comparable protocols among clinicians, nursing staff, and researchers will contribute to set clinical studies of major impact for the identification of reliable biomarkers.

Milestones In Urine Proteomics Applied To Diabetic Nephropathy

Since 2004, when Mischak and coworkers\textsuperscript{50} described three polypeptide patterns able to recognise ‘normal’, ‘diabetic’, and ‘diabetic patients with renal damage’, about 15 original works dealing with the identification of urinary biomarkers of DN have been published. Even if this proof-of-concept work lacked some details on the criteria that are now considered essential for the definition of a qualified set of biomarkers (i.e. the validation in an independent test set or the bioinformatics analysis to establish the functional association between the biomarker and the disease), it has been successful in showing, for the first time, that urine proteomics could provide new important information about kidney disease in T2DM patients.

In the following years, several well-designed works, based primarily on urine screening by CE-MS and SELDI-TOF/MS, have allowed for the identification of new promising biomarkers for early diagnosis and prognosis of DN. Rossing et al.\textsuperscript{51} applied CE-MS analysis to T1D patients, describing a panel of 65 urine biomarkers able to recognise DN with 97% sensitivity and specificity. Their results were further validated in a multicentre independent cohort\textsuperscript{52} of T2DM patients, providing the first evidence that CE-MS urine proteome profiling may adequately identify subjects with DN in the general population. About
half of the polypeptides included in the proteomic pattern were identified as collagen fragments, thus suggesting that changes in the collagen metabolism may be closely linked to the renal damage in T2DM. Furthermore, Good and coworkers\(^\text{53}\) reported a CE-MS based classifier including 273 urinary small peptides (namely ‘Classifier273’) that seem to be highly specific and sensitive for CKD, irrespective of the underlying pathology. In a very recent work, Zurbrig et al.\(^\text{54}\) demonstrated that this classifier was more specific and sensible than urine albumin excretion rate (UAER) in predicting the occurrence of the microalbuminuria in T1D and T2DM normoalbuminuric patients. These data, even if limited to a restricted number of diabetic patients, would suggest that the urine proteome might allow the identification of DN risk patients, thus permitting early onset of renoprotective treatments to slow the progression of the renal damage.

SELDI-TOF/MS analysis has also been extensively used for identifying urine biomarkers of DN. For example, Dihazi et al.\(^\text{55}\) identified and validated among 100 differently excreted SELDI peaks, two mass peaks corresponding to B2-microglobulin and ubiquitin ribosomal fusion protein, which were selectively and differently excreted in nephropathic diabetic patients. More recently, Wu et al.\(^\text{56}\) reported 300 differently excreted urine mass peaks among T2DM patients with normo, micro and macroalbuminuria, and described a four-peak pattern useful for recognising DN with 88% and 97% sensitivity and specificity, respectively. Interestingly, in these studies the progression of renal damage in T2DM was expressed only according to the albumin excretion rate.

Our group also performed a comparative SELDI analysis of the urine proteome,\(^\text{57}\) taking into account a more accurate selection of the T2DM patients since only diabetic patients with biopsy-proven Kimmelstiel-Wilson lesions were included in the DN group. We confirmed the data of Dihazi, concerning the increased excretion of B2-microglobulin in DN, and found significant deregulated excretion of the ubiquitin as potential biomarkers of DN. Further, we confirmed the specificity of the identified biomarkers in an independent test set of T2DM patients having biopsy-proven non-diabetic chronic kidney disease (CKD). It is worth noting that both CE-MS and SELDI profiling are able to specifically analyse low molecular weight proteins while being ineffective to cover the medium and high size proteome.

A high-throughput approach that allows a more accurate coverage of the proteome is the so-called shotgun proteomics analysis.\(^\text{29,58}\) In this approach, the proteins of a given biological sample are proteolytically digested into peptides and separated by bidimensional liquid chromatography prior to mass analysis (LC/MS). The ensuing peptide masses and sequences are then used to identify corresponding proteins by database search.\(^\text{59}\) Recently, Jin et al.\(^\text{60}\) employed the urine LC/MS analysis to search for specific DN biomarkers. Specifically, these authors used isobaric tags for relative and absolute quantitation (iTRAQ)\(^\text{60}\) to select and quantify differentially excreted urinary proteins in pooled urine samples of microalbuminuric versus normoalbuminuric diabetic patients. This analysis allowed the recognition of 196 differentially expressed proteins, including 10 (qualified) biomarkers that were identified by bioinformatics analysis. The application of a multiparametric pattern, encompassing three of the ten qualified biomarkers, allowed identification of microalbuminuric patients with about 92% sensitivity and specificity.

It is interesting to consider that most of the urine proteomic studies have investigated only the soluble urine fraction. Indeed, recently, urinary exosomes have been receiving increasing attention as a new source of potential biomarkers.\(^\text{62}\) Exosomes are 30-100 nm vesicles, derived from the endosomal compartment and released via fusion of multivesicular bodies with the plasma membrane.\(^\text{63}\) They comprise of a ceramide and cholesterol-rich lipid bilayer membrane,\(^\text{64}\) an array of membrane and cytosolic proteins,\(^\text{62}\) and selected RNA species.\(^\text{65}\) These vesicles are a rich source of biomarkers because they are released from every segment of the nephron, including podocytes, and are finally excreted in the urine.

Very recently, Raimondo and coworkers\(^\text{66}\) have published an interesting proof-of-concept work on the proteomic analysis of urine exosomes in Zucker Diabetic Fatty (ZDF) rats. They profiled the urinary exosomal protein content of non-diabetic lean rats and ZDF rats with normo or microalbuminuria. By this approach, 280 differently expressed exosomal proteins were identified and categorised according to the function and subcellular localisation. They demonstrated that incipient renal disease correlated with increased cytoplasmic and cytoskeletal proteins in the urine exosomes, and that the identified proteins were mainly involved in metabolic and immunity processes. The above results demonstrate that the proteomic analysis of the urinary exosomes, together
with the analysis of the soluble urinary proteins, may fruitfully contribute to reveal the pathophysiological alterations occurring in DN progression, and to enlarge the panel of DN biomarker candidates.

**CONCLUSIONS AND PERSPECTIVES**

Proteomics has become one of the most powerful tools for the mass-analysis of urine samples and is yielding a decisive contribution for a better understanding DN pathophysiology. More than a decade of studies has provided significant advances in the management of urine samples to find new sensitive and specific biomarkers of DN. However, the proteomic ability to quickly analyse thousands of urinary proteins has generated the wrong belief that, in few years, novel biomarkers that are able to recognise the onset of kidney damage with 100% accuracy would have been identified. Instead, the lack of consensus protocols for collecting, processing, and analysis of the samples has led to poor reproducible results among different studies, thus making difficult their generalisation.

We are now becoming aware of the need for protocol standardisation to enlarge the collection of comparable samples in different countries, and that the bioinformatics analysis of the complex datasets represent a *conditio sine qua non* for restricting the validation of the identified biomarkers to those specifically related to the pathophysiology of renal damage in T2DM. It is expected that this new way of managing the proteomic datasets will critically favour the identification of reliable biomarkers by reducing the effect of confounding factors. Furthermore, proteins are the players of a complex game, which also includes genes, transcripts, and metabolites, each influencing the others. Indeed, in the forthcoming years, bioinformaticians will have to develop more accurate tools to correlate proteomic datasets with the corresponding genomic, transcriptomic, and metabolomic datasets in order to pursue a global characterisation of the biological systems, and to identify a multi-level panel of molecular players cooperating to the onset of the pathological phenotypes.

**REFERENCES**


