Minireview

Urine Collection and Processing for Protein Biomarker Discovery and Quantification

C. Eric Thomas¹, Wade Sexton², Kaaron Benson³, Rebecca Sutphen⁴, and John Koomen¹,⁵

Abstract

**Background:** Urine is a useful source of protein for biomarker discovery and assessment because it is readily available, can be obtained by noninvasive collection methods, and enables monitoring of a wide range of physiologic processes and diseases. Urine aliquots provide enough protein for multiple analyses, combining current protocols with new techniques.

**Conclusions:** Standardized collection and processing protocols are now being established and new methods for protein detection and quantification are emerging to complement traditional immunoassays. The current state of urine collection, specimen processing, and storage is reviewed with regard to discovery and quantification of protein biomarkers for cancer. *Cancer Epidemiol Biomarkers Prev; 19(4); 953–9. ©2010 AACR.*

Urine as a Source of Biomarkers

Protein in urine originates from glomerular filtration of plasma, excretion from epithelial cells in the urinary tract, sloughing of epithelial cells and casts, and formation of urinary exosomes (1). Changes in urine protein components and concentrations, therefore, may report directly on dysfunction of cells within the urinary tract, whereas other diseases may be detectable through the transmission of analytes from blood into urine. Although blood plasma is strictly governed by homeostatic mechanisms, protein accumulation in urine is not. As a consequence, the proteome of urine seems to be significantly different from blood plasma in terms of protein composition; for example, albumin comprises ~50% (by weight) of the total plasma proteome but only 7% of the protein excreted in urine (1). Urine is a complex and diverse source of candidate protein biomarkers; >1,500 proteins were identified in a study using liquid chromatography coupled to tandem mass spectrometry peptide sequencing (LC-MS/MS) for the identification of urinary proteins separated by SDS-PAGE and reverse-phase liquid chromatography before tryptic digestion (2). Another study using capillary electrophoresis coupled to mass spectrometry resulted in the identification of as many as 4,094 peptides in unfractionated and undigested samples (3).

Potential urinary protein biomarkers have been reported for several types of cancer (Table 1). Assays for bladder-tumor–associated antigen (4) and nuclear matrix protein, NMP22 (5-7), in urine have already been approved by the Food and Drug Administration (FDA) for bladder cancer screening. Urinary cathepsin D has been reported as a potential prognostic biomarker for renal cell carcinoma, correlating well with survival in studies using mass spectrometry and immunoassays (8). Additional candidate biomarkers (9-23) for diagnosis, prognosis, and selection of treatment listed in Table 1 are in the development pipeline. Biomarker panels are also under investigation to improve positive predictive value with the goal of using molecular information to tailor treatment regimens to each patient (24).

Challenges in Urine Collection and Processing

Urine specimens show a high degree of variability in volume; protein concentration (particularly in the case of kidney damage or dysfunction); total protein excreted pH (ranges from 4-8); as well as variability in urine components due to age, health, diet, or other factors; proteolysis while the urine is stored in the bladder; and degradation of collected urine samples upon storage (25). The range of values observed for these properties is high not only from person to person (26, 27) but also within each individual. As an example, urine was collected in our laboratory over the course of 10 days from a single subject; protein concentrations were measured by Bradford assay for each individual sample and pooled 24-hour collections.

**Authors’ Affiliations:** ¹Proteomics, ²Genitourinary Oncology, ³Malignant Hematology, and ⁴Molecular Oncology, Moffitt Cancer Center; and ⁵Genetics, College of Medicine, University of South Florida, Tampa, Florida

**Corresponding Author:** John Koomen, SRB3 Molecular Oncology and Proteomics, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Drive, Tampa, FL 33612. Phone: 813-745-8524; Fax: 813-745-3829. E-mail: john.koomen@moffitt.org

doi: 10.1158/1055-9965.EPI-10-0069

©2010 American Association for Cancer Research.
Individual and average protein concentration values are plotted in Fig. 1. Relative SDs were lowest for the 24-hour collection (39%) and first morning collections (41%), whereas the second urine samples of the day (54%) and spot collection (61%) exhibited higher variability. Average concentrations were similar, ranging from 26 mg/mL for 24 h, random spot, and second morning urine to 34 mg/mL for first morning urine. The purpose of this comparison is to illustrate that high variability is observed with each collection method; practical considerations often limit studies to random spot collection.

For proteomics and protein biomarker experiments, normalization of the amount or concentration of protein is critical. One method for standardization uses the ratio of each protein's expression to the total protein in the sample (28). However, more precise normalization of data may be performed by calculating ratios to other excreted small molecules or proteins. Creatinine (28), collagen (28), albumin, cystatin C (29-32), and N-acetyl-β-D-glucosaminidase (33) have all been suggested as potential standards for normalization of protein quantification (28). Each of these potential normalization factors should be carefully examined regarding their stability under different disease conditions.

### Table 1. Protein biomarkers in urine investigated for detection and staging cancer

<table>
<thead>
<tr>
<th>Cancer</th>
<th>Potential biomarkers</th>
<th>Comment (Reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>Bladder-tumor–associated antigen</td>
<td>FDA-approved immunoassay (4)</td>
</tr>
<tr>
<td></td>
<td>NMP22</td>
<td>FDA-approved immunoassay (5)</td>
</tr>
<tr>
<td></td>
<td>Calreticulin</td>
<td>Partially Validated through Western blot (15)</td>
</tr>
<tr>
<td></td>
<td>Clusterin</td>
<td>Diagnostic/prognostic (13)</td>
</tr>
<tr>
<td></td>
<td>Cystatin B</td>
<td>Partially Validated through IHC and Western blot (11)</td>
</tr>
<tr>
<td></td>
<td>Proepithelin</td>
<td>Partially validated through IHC and ELISA (18)</td>
</tr>
<tr>
<td></td>
<td>UHRF1</td>
<td>Observed through IHC (22)</td>
</tr>
<tr>
<td></td>
<td>α-1B-Glycoprotein</td>
<td>Discovered using lectin affinity chromatography (17)</td>
</tr>
<tr>
<td>Renal</td>
<td>Cathepsin D</td>
<td>Correlated with survival (8)</td>
</tr>
<tr>
<td></td>
<td>NMP22</td>
<td>FDA-approved for bladder cancer (16)</td>
</tr>
<tr>
<td>Prostate</td>
<td>Collagen α-1(III) peptide Collagen α-1(I) peptide</td>
<td>MMP substrate (21)</td>
</tr>
<tr>
<td></td>
<td>Psoriasis susceptibility 1 candidate gene 2 protein peptide</td>
<td>CE-MS detection, decreased in cancer (21)</td>
</tr>
<tr>
<td></td>
<td>Sodium/potassium-transporting ATPase γ peptide</td>
<td>CE-MS detection (21)</td>
</tr>
<tr>
<td></td>
<td>PCA3</td>
<td>mRNA detection in urine (14)</td>
</tr>
<tr>
<td>Ovarian</td>
<td>Eosinophil-derived neurotoxin COOH-terminal osteopontin fragments</td>
<td>Glycosylated fragment (62)</td>
</tr>
<tr>
<td></td>
<td>Bcl-2</td>
<td>ELISA (9)</td>
</tr>
<tr>
<td>Breast</td>
<td>MMP9</td>
<td>Detected by zymography (56)</td>
</tr>
<tr>
<td></td>
<td>ADAM12</td>
<td>Detected by Western blot (56)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>Cystatin SN</td>
<td>Tissue IHC and RT-PCR: ELISA with Urinary Protein (23)</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>IL-6, MMP9</td>
<td>Index for Bone Disease (19)</td>
</tr>
<tr>
<td></td>
<td>Antibodies</td>
<td>Diagnostic/Prognostic (10, 12, 20)</td>
</tr>
</tbody>
</table>

NOTE: This table includes two protein immunoassays have FDA approval for detection of bladder cancer. Additional entries describe other candidates that are in various stages of the development pipeline. Several methods can be used for protein assessment, including ELISA, immunohistochemistry of tumor tissue (IHC), real-time PCR (RT-PCR), or capillary electrophoresis coupled to mass spectrometry (CE-MS).

Abbreviations: MMP, matrix metalloproteinase; IL-6, interleukin-6.

Protocols for Urine Collection and Processing

Myriad protocols exist for urine collection and processing (selected examples are given in Table 2) and there are advantages to 24-h, first morning urine, and spot urine collection. Twenty-four-hours collection, which is our current standard clinical protocol for antibody measurements in myeloma, is awkward for the patient and may lead to degradation and contamination of urine protein, particularly through the lysis of suspended cells, because the samples are stored at 4°C and transported under ambient conditions. Single sample collection is more convenient for patients, is more easily standardized, and enables quicker processing and storage of samples. Among single-sample collections, first morning urine provides the least variability in protein concentration; second morning and random spot urine collection display somewhat higher variability, but such collection minimizes the amount of time spent in the bladder, in which increased proteolysis may occur. Additionally, random spot collection facilitates coordination between patients, clinicians, and researchers.

Midstream collection of urine minimizes problems with bacterial contamination and preservatives such as

---

Cancer Epidemiol Biomarkers Prev; 19(4) April 2010

Cancer Epidemiology, Biomarkers & Prevention
sodium azide and boric acid have been used to prevent bacterial growth in stored urine (34). Although protease inhibitors are often added to biological samples to preserve intact protein forms, their value in urine samples is unknown due to pH, dilution, and denaturation of proteases in the urine (35, 36). Experiments to quantify the activity of different proteases would add significantly to the existing literature. Despite significant variability in urine pH (37), most publications do not report pH modification for proteomic analyses. Centrifugation performed within 20 to 30 minutes of collection minimizes contamination of the urine due to lysis of suspended cells (38). Ultrafiltration (34, 39-45) has been reported to be the best method for concentration and cleanup of peptide and protein components from urine (45) and facilitates collection of lower molecular weight analytes and buffer exchange for downstream processing. Alternatively, protein precipitation using organic solvents (34, 40, 44-47), dialysis (34, 39, 46, 48), lyophilization (34, 39, 46), and ultracentrifugation (34) may be used.

After initial processing, samples must be stored in a way that preserves the analytes for downstream processing. Typically, clinical specimens are frozen as soon as possible following collection. Although freezing and thawing should be avoided, minimal loss has been observed, through SDS-PAGE and LC-MS/MS, with up to five freeze-thaw cycles (49). However, one cannot conclude that a specific individual protein will be preserved.

The Human Kidney and Urine Proteome Project, associated with the Human Proteome Organisation (HUPO),

Table 2. Selected strategies for urine protein collection and analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>NMP-22 immunoassay (7)</th>
<th>Ovarian cancer biomarker discovery (62)</th>
<th>Urine protein cataloguing (2)</th>
<th>HUPO Kidney and Urine Proteome Project (50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Single/pooled samples</td>
<td></td>
<td>Midstream 2nd or spot urine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sodium azide/boric acid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Prostate inhibitor</td>
<td>Centrifuge 2,000 × g (10 min)</td>
</tr>
<tr>
<td>Additive(s)</td>
<td>Prostate inhibitor</td>
<td>Centrifuge 2,000 × g (10 min)</td>
<td></td>
<td>Centrifuge 10,000 × g (10 min)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ultrasound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>Separation</td>
<td>NA</td>
<td>Two-dimensional Gel RPLC, SDS-PAGE</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>Detection</td>
<td>Immunoassay</td>
<td>LC-MS/MS</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>Result</td>
<td>Detection of bladder cancer 81% sensitivity, 87% specificity</td>
<td>Targets identified by MS and verified with ELISA</td>
<td>1,500 proteins identified in urine</td>
<td>NA</td>
</tr>
</tbody>
</table>

NOTE: Workflows are presented for antibody-based quantification techniques as well as proteome cataloging experiments. The recommendations of the HUPO Human Kidney and Urine Proteome Project are also included.

Abbreviations: NA, not applicable; RPLC, reverse-phase liquid chromatography; MALDI-TOF-MS, matrix assisted laser desorption ionization time-of-flight mass spectrometry; IP, immunoprecipitation.
Figure 2. Translation of a peptide candidate for detection and quantification of osteopontin in urine from an LC-MS/MS catalog to an LC-MRM assay. An example base peak chromatogram from LC-MS/MS is shown to illustrate the complexity of the data (A); individual ion signals can be extracted for initial quantification (inset). The tandem mass spectra (B) are used to identify the peptide sequence (and consequently the protein of origin) and they can also be useful for indicating which fragment ions should be selected for monitoring in LC-MRM assays (C).
has tentatively recommended the following steps for processing urine (50): midstream collection of random or second morning urine, freezing within 4 hours of collection or addition of sodium azide or boric acid to sample before freezing, centrifugation at 10,000 × g for 10 minutes to remove cells and debris, and minimization of freezing and thawing. They have also suggested the collection of 19 sample identifiers describing the patient, processing conditions, storage, and data generated from other methods of urinalysis.

Based on the existing resources describing sample collection and processing, we have developed the following protocol to accrue samples for diagnostic biomarker discovery and verification experiments in bladder cancer patients and healthy controls. Deidentified random spot urine (~60 mL midstream) is collected into sterile containers for processing within 20 minutes; the samples are placed on ice for transfer to the laboratory and all subsequent processing steps. After transfer to a 50-mL conical centrifuge tube, the urine samples are spun at 1,500 × g for 10 minutes in a refrigerated centrifuge to pellet any cells. Smaller aliquots (10 mL) are spun at 10,000 × g to remove particulates and aliquoted for proteomics and metabolomics. Protein in the urine supernatant is concentrated using 3,000 Dalton molecular weight cutoff filters (Ultra 15, Amicon); the buffer is exchanged to 100 mmol/L ammonium bicarbonate by washing the membrane. All processed samples are frozen at −80°C within 2 hours of collection.

Selected Protein Quantification

Validation of biomarkers is routinely achieved through immunoassays, e.g., ELISA, which recognize an epitope within the protein sequence for detection and quantification. Peptide-based proteomics offers an analogous technology using proteome catalogs generated with LC-MS/MS to supply candidate biomarkers for targeted detection and quantification with liquid chromatography coupled to multiple reaction monitoring mass spectrometry (LC-MRM; ref. 51) An example workflow is illustrated for the development of an assay for osteopontin detected in urine (Fig. 2). Because of the ability to create highly multiplexed assays (52), LC-MRM can be applied to a large number of potential targets informed by the existing literature, gene expression profiles, protein cataloging experiments, among others. Multiplexed LC-MRM assays for several potential protein biomarkers in urine are currently under development in our laboratory; peptide targets have been successfully detected for cathepsins B and D (8, 53, 54), matrix metalloproteases 9 and 13 (55-58), HBGF (59), IGFBP7 (60, 61), and osteopontin (62) from unfraccionated urine protein extracts. Due to the high variability of protein concentration, the inclusion of two different stable isotope-labeled peptide standards at different concentrations may be desirable or necessary for protein quantification with LC-MRM.

Successful LC-MRM assays for 47 abundant proteins in plasma have been reported by Anderson and Hunter (63), with detection down to 0.67 μg/mL without fractionation; similar performance should be expected for urine protein analysis. Increased sensitivity can be achieved using immunodepletion (64), fractionation (65), or antibody-based enrichment in a Stable Isotope Standards and Capture by Antipeptide Antibodies approach. A Stable Isotope Standards and Capture by Antipeptide Antibodies–LC-MS/MS assay for serum thyroglobulin, a marker for thyroid carcinoma that has proven difficult to quantitate by immunoassay, has been developed with a reported limit of detection of 2.6 μg/L (4 pmol/L; ref. 66). Building on these early successes, the generation of standardized assays and reagents for quantification of numerous potential biomarkers has been proposed (67).

Summary

A standardized method for urine collection and processing has yet to be adopted by the proteomics or biomarker assessment community. Thus far, each study uses a protocol devised specifically for the analyze(s) of interest in their experiments. Common themes are found for collection, cell removal, processing, storage, and downstream analysis that can guide the development of a standard method. The HUPO guidelines for data recording offer the best suggestions to date, including details about the patient as well as collection and processing. However, additional background research is necessary to determine the value of additives, such as protease inhibitors, and to understand the effects of different processing steps and time scales.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

The Moffitt Proteomics Facility is supported by the U.S. Army Medical Research and Materiel Command under award no. DAMD17-02-2-0051 for a National Functional Genomics Center, the National Cancer Institute under award no. P50-CA076292 as a Cancer Center Support Grant, and the Moffitt Foundation. This work is also supported by the National Cancer Institute by awards: R21-CA141285 (J. Koomen) and R01-CA106414 (R. Sutphen).

Received 01/15/2010; revised 02/11/2010; accepted 02/17/2010; published OnlineFirst 03/23/2010.

References


