Advances in CLINICAL CHEMISTRY

VOLUME 51

Edited by

Gregory S. Makowski

ADVANCES IN CLINICAL CHEMISTRY

VOLUME 51

This page intentionally left blank

Advances in CLINICAL CHEMISTRY

Edited by

GREGORY S. MAKOWSKI

Clinical Laboratory Partners
Newington, CT
Hartford Hospital
Hartford, CT

VOLUME 51

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier

32 Jamestown Road, London NW1 7BY, UK
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

This book is printed on acid-free paper. ♾

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-380981-0
ISSN: 0065-2423

For information on all Academic Press publications
visit our website at www.elsevierdirect.com

Printed and bound in USA
10 11 12 10 9 8 7 6 5 4 3 2 1

CONTENTS

Cocaine in Acute Myocardial Infarction

GIUSEPPE LIPPI, MARIO PLEBANI, AND GIANFRANCO CERVELLIN

EGFR and KRAS in Colorectal Cancer

BEN MARKMAN, FRANCISCO JAVIER RAMOS, JAUME CAPDEVILA, AND JOSEP TABERNERO

Prostaglandin EP Receptors and Their Roles in Mucosal Protection and Ulcer Healing in the Gastrointestinal Tract

KOJI TAKEUCHI

This page intentionally left blank

CONTRIBUTORS

Numbers in parentheses indicate the pages on which the authors' contributions begin.

RADU ALBULESCU (145), *'Victor Babes' National Institute of Pathology, Splaiul Independentei; and National Institute for Chemical Pharmaceutical Research and Development, Calea Vitan, Bucharest, Romania*

EKREM ALTUNEL (25), *Ege University Medical School, Izmir, Turkey*

EUGENE O. APOSTOLOV (25), *Department of Pharmacology and Toxicology; and Division of Nephrology, Department of Internal Medicine, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA*

ALEXEI G. BASNAKIAN (25), *Department of Pharmacology and Toxicology; Division of Nephrology, Department of Internal Medicine, University of Arkansas for Medical Sciences; and Renal Medicine Service, Central Arkansas Veterans Healthcare System, Little Rock, Arkansas, USA*

JAUME CAPDEVILA (71), *Medical Oncology Department, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital, Barcelona, Spain*

GIANFRANCO CERVELLIN (53), *U.O. Pronto Soccorso e Medicina d'Urgenza, Dipartimento di Emergenza-Urgenza, Azienda Ospedaliero-Universitaria di Parma, Italy*

HARVEY J. COHEN (181), *Department of Pediatrics, Stanford University, Stanford, California, USA*

MIHAIL EUGEN HINESCU (145), *'Victor Babes' National Institute of Pathology, Splaiul Independentei; and 'Carol Davila' University of Medicine and Pharmacy, B-dul Eroilor Sanitari, Bucharest, Romania*

FRANCISCO JAVIER RAMOS (71), *Medical Oncology Department, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital, Barcelona, Spain*

XUEFENG B. LING (181), *Department of Pediatrics, Stanford University, Stanford, California, USA*

GIUSEPPE LIPPI (53), *U.O. Diagnostica Ematochimica, Dipartimento di Patologiae Medicina di Laboratorio, Azienda Ospedaliero-Universitaria di Parma, Italy*

BEN MARKMAN (71), *Medical Oncology Department, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital, Barcelona, Spain*

ELIZABETH D. MELLINS (181), *Department of Pediatrics, Stanford University, Stanford, California, USA*

MONICA NEAGU (145), *'Victor Babes' National Institute of Pathology, Splaiul Independentei, Bucharest, Romania*

ERCAN OK (25), *Ege University Medical School, Izmir, Turkey*

J.A. PÁRAMO (1), *Hematology Service, University Clinic, Atherosclerosis Research, CIMA, University of Navarra, Pamplona, Spain*

MARIO PLEBANI (53), *Dipartimento di Medicina di Laboraotorio, Azienda Ospedaliero-Universitaria di Padova and Leonardo Foundation, Abano Terme (PD), Italy*

SUDHIR V. SHAH (25), *Division of Nephrology, Department of Internal Medicine, University of Arkansas for Medical Sciences; and Renal Medicine Service, Central Arkansas Veterans Healthcare System, Little Rock, Arkansas, USA*

KARL G. SYLVESTER (181), *Department of Pediatrics; and Department of Surgery, Stanford University, Stanford, California, USA*

JOSEP TABERNERO (71), *Medical Oncology Department, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital, Barcelona, Spain*

KOJI TAKEUCHI (121), *Divison of Pathological Sciences, Department of Pharmacology and Experimental Therapeutics, Kyoto Pharmaceutical University, Kyoto, Japan*

CRISTIANA PISTOL TANASE (145), *'Victor Babes' National Institute of Pathology, Splaiul Independentei, Bucharest, Romania*

PREFACE

I am pleased to present Volume 51 of Advances in Clinical Chemistry series for the year 2010.

In the second volume for this year, a number of diverse topics are reviewed. This volume leads off with a review on the importance of prothrombin fragments in pathophysiologic processes such as thrombosis and cardiovascular disease, the leading causes of death in the Western world. This chapter is followed by an interesting review on the role of carbamylation, the nonenzymatic modification of protein by cyanate, in atherosclerosis and its potential exacerbation by end-stage kidney disease. The role of cocaine in cardiac disorders ranging from arrhythmias to myocardial infarction is next presented. The next chapter presents a comprehensive review on the molecular mechanisms of EGFR and KRAS in the initiation and progression of colorectal cancer, the third most common cancer worldwide. The next review highlights the interesting importance of endogenous prostaglandins and their receptors in mucosal protection and ulcer healing in the gastrointestinal tract. The following review explores the critical need for development of accurate diagnostic and therapeutic biomarkers for detection of pancreatic cancer, an insidious and complex pathophysiologic process. The volume concludes with an exploration of the role of urine peptidomics as a novel analytical approach to biomarker discovery for both systemic and renal diseases.

I extend my appreciation to each contributor of Volume 51 and thank colleagues who found time to contribute to the peer review process. I also extend a thank you to my editorial liaison at Elsevier, Gayathri Venkatasamy, for continued professionalism.

I hope the second volume for 2010 will be enjoyed by our readership. As always, your comments and suggestions for up-to-date review articles for the Advances in Clinical Chemistry series are always appreciated.

In keeping with the tradition of the series, I would like to dedicate Volume 51 to my nephew Steven on the occasion of this 30th birthday.

GREGORY S. MAKOWSKI

This page intentionally left blank

URINE PEPTIDOMICS FOR CLINICAL BIOMARKER DISCOVERY

Xuefeng B. Ling,*[,1] Elizabeth D. Mellins,* Karl G. Sylvester,*[,†] and Harvey J. Cohen*

***Department of Pediatrics, Stanford University, Stanford, California, USA**
[†]Department of Surgery, Stanford University, Stanford, California, USA

1. Abstract

Urine-based proteomic profiling is a novel approach that may result in the discovery of noninvasive biomarkers for diagnosing patients with different diseases, with the aim to ultimately improve clinical outcomes. Given new and emerging analytical technologies and data mining algorithms, the urine peptidome has become a rich resource to uncover naturally occurring peptide biomarkers for both systemic and renal diseases. However, significant analytical hurdles remain in sample collection and storage, experimental design,

[1]Corresponding author: Xuefeng B. Ling, e-mail: ling.xuefeng@gmail.com

0065-2423/10 $35.00
DOI: 10.1016/S0065-2423(10)51007-2

data analysis, and statistical inference. This study summarizes, focusing on our experiences and perspectives, the progress in addressing these challenges to enable high-throughput urine peptidomics-based biomarker discovery.

2. Introduction

Since samples can be collected noninvasively and in large amounts, urine is a desirable choice of biological fluids for proteomic discovery of disease related biomarkers. Urine biomarker translation to clinical practice may lend itself to long-term disease monitoring, response to therapy, and potentially home assays. Urine peptidome, a diversified pool of the naturally occurring peptides, is emerging as a rich source for clinical biomarkers reflecting patients' pathophysiological status.

However, there are several barriers to the success of the field. Operationally, a consistently performed protocol for urine peptidome handling and storage needs to be developed to reduce analytical bias. Mass spectrometric urine peptidome profilings generate vast amounts of peptide peak spectra, but a high-confidence database of urine peptide sequences is yet to be established. Another issue is that the large-scale mass spectrometry-based urine peptidomics creates experimental and analytical bottlenecks. Is pooling a viable data reduction strategy such that significant amounts of discovery efforts can be saved? Although quantitative isotope labeling based methods, for example, iTRAQ^{TM}, is capable of significantly reducing mass spectrometric time, urine peptidomics biomarker discovery may be better addressed by using label-free MS techniques to achieve sufficient statistical power. Do urine "housekeeping" peptides exist such that the common variations caused by biological and analytical issues can be corrected? Statistically, urine peptidomics analysis involves large-scale simultaneous hypothesis testing. Robust statistical methods are clearly needed to extract important patterns and trends, and guide us away from the false discoveries. Here we describe our attempt to bring together many of the important new perspectives in urine peptidomics and explain them in a generalized framework for urine-based biomarker discovery.

3. Urine Peptidome is a Rich Source of Peptides of Diversified Protein Origins

A normal adult human excretes 30–130 mg of protein and 22 mg of peptides per day in urine [1, 2]. Naturally occurring urine peptides have certain advantages over urine proteins as biomarkers. The roughly equal

masses of protein and peptide in urine represent at least a 10-fold greater molar excess of peptides. The urine proteome contains a number of abundant proteins that obscure the lower abundance proteins, which are more likely to be biomarkers.

The most popular approaches [3] for urine composition analyses include hyphenated mass spectrometry-based techniques: 2D gel electrophoresis (2DE)-MS, LC–MS, SELDITOF-MS, and capillary electrophoresis (CE)-MS. However, only CE-MS and LC–MS, capable to directly interface with MS/MS instruments for biomarker peptide sequencing, allow the study of urine peptidome (in general, peptide/proteins < 10 kDa) with required depth of analysis, dynamic range, and enhanced accuracy of quantization. Initial peptidomic studies conducted in normal urine [4] and clinical samples [5] provided proof of principle of the effective use of prefractionation techniques in urinary peptidome profiling. Toward high mass accuracy and resolution, the evolution of mass spectrometric technologies in ion source (e.g., electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI)), mass analyzer (e.g., time-of-flight (TOF), quadrupole, quadrupole ion trap, linear quadrupole ion trap (LTQ), Fourier transform ion cyclotron resonance (FTICR), orbitrap), and detector is critical to urine peptidome in depth profiling and characterization. To determine the biomarker sequences, tandem mass spectrometry consisting of a combination of two or more mass analyzers (e.g., triple quadrupole, Q-TOF, ion trap TOF, LTQ Orbitrap, LTQ-FTICR, MALDI-TOF/TOF spectrometers) generate peptide sequence tags, which subsequently can be utilized to identify a peptide in a protein database [6–8]. Although each implementation is different, tandem MS search algorithms, including SEQUEST, MASCOT, Spectrum Mill, X!TANDEM, etc., operate under the same general principles and there should be little difference in the output of the algorithms so long as consistent scoring procedures are applied [9, 10].

As a benchmark, currently a 1D HPLC (this study) or CE separation [11] with MS is adequate for the analysis of greater than 25,000 or 100,000 urine peptides. Nevertheless, catalogs of precise and comprehensive quantification of urinary polypeptides in either normal or disease subjects are yet to be constructed. Proteomic analyses have identified 1543 different urinary proteins, including a large proportion of membrane proteins [12]. Recent reviews [11, 13] showed CE-MS (capillary electrophoresis coupled with mass spectrometry) urine peptidomics analysis defined 116,869 different peptides features, clustered by molecular mass and CE-migration time. Further noise filtering reduced the number to 5010 peptide features with robust signals across samples, of which 444 different peptide sequences from a total of 60 unique protein precursors that were determined by MSMS (''Human urinary peptide sequences v2.0'' on the Mosaiques Diagnostics' corporate

webpage http://mosaiques-diagnostics.de/diapatpcms/mosaiquescms/front_content.php?idcat=257/). Of these 444 identified protein sequences, the deviation between the calculated and the observed masses ranged from −42.5 to 25.72 ppm (median: 2.59 ppm, standard deviation: 9.37 ppm). Therefore, the urine peptidome is indeed a rich source of naturally occurring peptides. However, the identities of the majority of these urine peptides remain to be determined.

To further explore the identities of the urine peptidome contents, urine samples from systemic juvenile idiopathic arthritis (SJIA), Kawasaki disease (KD), febrile illnesses (FI), necrotizing enterocolitis (NEC), and normal volunteer (V) subjects were collected for mass spectrometric analysis. Informed consent was obtained from all patients and healthy controls. The extraction protocol of the urine peptidome and proteome is as previously described [14]. Second morning void mid-stream urine samples (1–10 ml) were collected in sterile containers and were centrifuged at $2000 \times g$ for 20 min at room temperature (RT) within 1 h of collection. The supernatant was transferred, adjusted to pH 7.0, and stored frozen at −80 °C until further use. Urinary samples were processed by centrifugal filtration at 3000 × *g* for 20 min at 10 °C through Amicon Ultra centrifugal filtration devices (10 kDa cutoff) (Millipore, Bedford, MA) preequilibrated with 10 ml Milli-Q water. The retentate (urine proteome) was washed twice, brought to the final volume of 400 µl with 20 mM Tris–HCl (pH 7.5), and quantitated by the bicinchoninic acid (BCA) protein assay (Pierce, Rockford, IL). The filtrate (urine peptidome) containing the low MW naturally occurring peptides was processed with Waters Oasis HLB Extraction Cartridges (Waters Corporation, Milford, MA), and extracted with ethyl acetate. The resulting urine peptide samples were quantified by the 2,4,6-trinitrobenzenesulfonic acid (TNBS) assay, as previously described [15]. Three nanomoles of peptides were fractionated by 2D chromatography—a strong cation exchange (SCX) column as the first and a reversed phase (RP) column as the second dimension, and then subjected to extensive MSMS sequence identification involving a Thermo Finnigan LTQ-FTICR spectrometer.

MS/MS spectra were searched by SEQUEST (BioWorks™ rev.3.3.1 SP1) against the International Protein Index (IPI) human database version 3.5.7 restricted to human entries (76,541 sequences). mMASS, an open source mass spectrometry tool (http://mmass.biographics.cz/), was used for manual review of the protein identification and MS/MS ion pattern analysis for additional validation. Different fragmentation techniques were used for the validation of a peptide sequence, as well as for the detection, localization, and characterization of posttranslational modifications. Peptide identifications were considered acceptable if they passed the thresholds and additionally if the XCorr (the cross-correlation value from the search) was greater

than 2.0 and the deviation between calculated and observed masses was less than 10 ppm. This in-depth 2D MS/MSMS analysis led to the identification of 11,988 different urine peptide sequences from 8519 unique protein precursors.

The protein IDs of the protein precursors of the urine peptides were uploaded to PANTHER 7.0 (http://www.pantherdb.org/) to explore the molecular function, and to gain insight to the biological processes, and cellular components that these urine naturally occurring peptides might involve (Fig. 1). The PANTHER (Protein ANalysis THrough Evolutionary

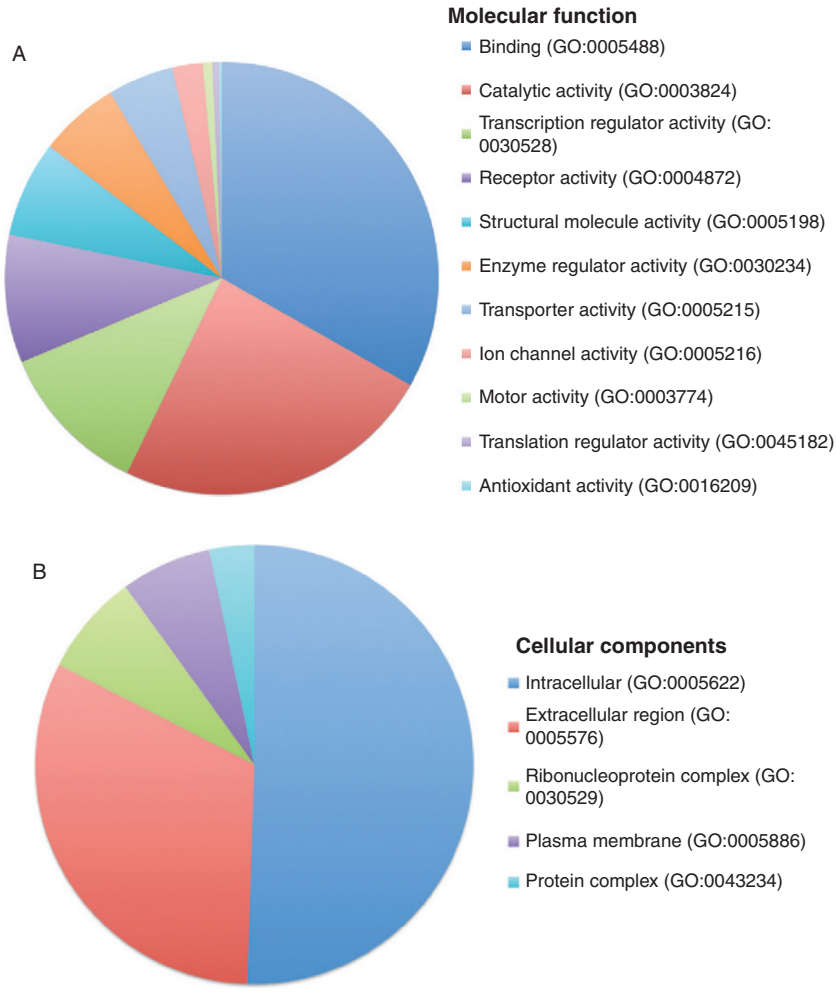


FIG. 1. (Continued)

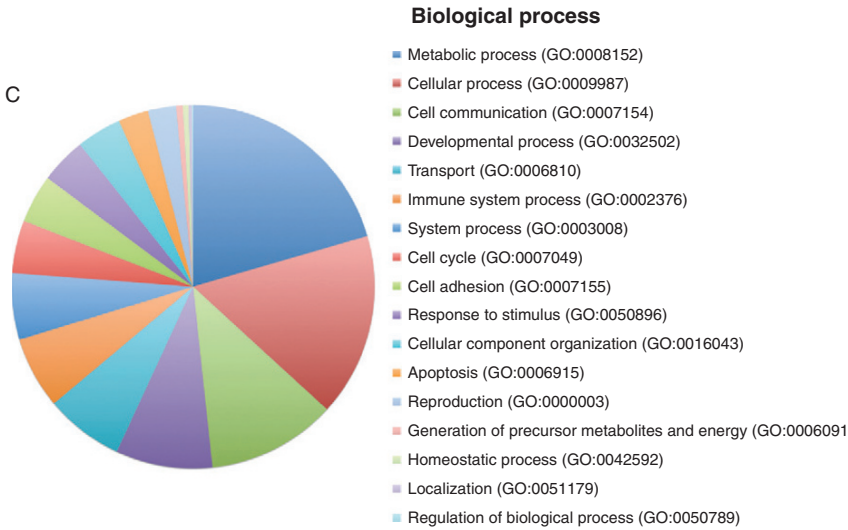


FIG. 1. Characterization of the urine peptides' protein precursors via PANTHER Gene Ontology annotation.

Relationships) Classification System [16] is a unique resource that classifies proteins according to Gene Ontology (GO) using published scientific experimental evidence and evolutionary relationships. GO cellular component analysis revealed the following GO terms as overrepresented: 50.50% intracellular (GO:0005622), 32.00% extracellular region (GO:0005576), 0.9% ribonucleoprotein complex (GO:0030529), 0.8% plasma membrane (GO:0005886), and 0.4% protein complex (GO:0043234). In the molecular function category, these GO terms are overrepresented: 33.30% binding (GO:0005488), 24% catalytic activity (GO:0003824), 11.50% transcription regulator activity (GO:0030528), 9.6% receptor activity (GO:0004872), 7.2% structural molecule activity (GO:0005198), 6% enzyme regulator activity (GO:0030234), 4.9% transporter activity (GO:0005215), 2.3% ion channel activity (GO:0005216), 0.7% motor activity (GO:0003774), 0.5% translation regulator activity (GO:0045182), and 0.2% antioxidant activity (GO:0016209). In the GO biological process category, these GO terms are overrepresented: 20.50% metabolic process (GO:0008152), 16.30% cellular process (GO:0009987), 11.50% cell communication (GO:0007154), 8.60% developmental process (GO:0032502), 7.0% transport (GO:0006810), 6.40% immune system process (GO:0002376), 5.90% system process (GO:0003008), 4.70% cell cycle (GO:0007049), 4.30% cell adhesion (GO:0007155),

4.10% response to stimulus (GO:0050896), 4.0% cellular component organization (GO:0016043), 2.70% apoptosis (GO:0006915), 2.50% reproduction (GO:0000003), 0.60% generation of precursor metabolites and energy (GO:0006091), 0.50% homeostatic process (GO:0042592), 0.30% localization (GO:0051179), and 0/10% regulation of biological process (GO:0050789).

Sequence determination of some urine peptides by mass spectrometry may be difficult. This may therefore limit the number of biomarker candidates that can be moved forward successfully for further evaluation. Based on our current experience, there are three reasons for failure to obtain a peptide sequence by MS/MS analysis. (1) Some peptides are too low in abundance in the original samples for successful MS/MS. This can be overcome by increasing the sample load and/or the purity of the peptide since peptide ionization efficiency in MALDI is related to the purity of the sample. To increase sample load and purity, a 2D or 3D HPLC purification may be required prior to MS/MS analysis. (2) Some peptides appear to have adequate signals in MS mode but do not produce a sufficient number of product ions in MS/MS to allow identification. (3) Many urine peptides have posttranslational modifications. Although some modifications, for example, hydroxylation of proline, are recognized by the database software, we have observed other modifications that are not normally considered by these same methods. In these instances it is necessary to analyze the data manually. Our overall experience is that we are able to confidently identify approximately 75% of the urine peptides we have analyzed in our laboratory. Additionally, we have also observed that different mass spectrometers, for example, Thermo Finnigan LTQ-FTICR and ABI MALDI-TOF, can complement each other and can therefore analyze peptides unable to be identified by only one mass spectrometric platform MS/MS. It is possible that certain critical biomarker peptides will not be identified by mass spectrometry. In these instances, Edman chemistry based amino acid sequence analysis, for example, through an Applied Biosystems 494 protein sequencer, may be a good alternative method. We have encountered very few instances of N-terminal blocked urine peptides and therefore expect Edman sequencing to be successful. However, for this technique, it is necessary to purify picomole quantities of peptides. This will likely require 2–4 steps of ion exchange and reverse phase chromatographic fractionation of 200–500 ml of urine. Purification of the peptide of interest can be monitored by MALDI-TOF mass spectrometry.

Together our urine peptidome sequence identification and subsequent comprehensive GO analysis indicate that the urine peptidome contents are derived from proteins representing diverse molecular functions, as well as cellular processes and biological processes. Since 70% of the urinary proteome/peptidome originates from the kidney and urinary tract, with the remaining 30% from the circulation in healthy individuals, analysis of the

urinary proteome/peptidome can be highly informative for both renal and systemic disease diagnosis and prognosis [3]. In fact, urinary peptidomics is emerging as a powerful noninvasive tool for diagnosis and monitoring both systemic and renal diseases: coronary artery disease [17], acute renal tubulointerstitial rejection [18], chronic renal allograft dysfunction [19], diabetic nephropathy, and chronic kidney disease [20–22], congenital unilateral ureteropelvic junction obstruction in newborns [23], urothelial cancer [24], and prostate cancer [25, 26]. We have applied HPLC coupled with MALDI-TOF analysis to profile the urine peptidome and discovered urine peptide biomarker candidates that can aid in the diagnosis and prognosis of various pediatric diseases including acute rejection (AR) following renal transplantation [27], pediatric renal dysfunction, KD, SJIA, and NEC.

4. Quantitative Urine Peptidomics for Biomarker Discovery

Although LC–MS/MSMS techniques are instrumental in characterizing the urine peptidome, quantitative urine peptidomics based biomarker discovery still remains challenging due to several technological limitations. Among the emerging quantitative technologies, iTRAQ (isobaric tags for relative and absolute quantification) allows the concurrent protein sequence identification and relative quantification of those peptides with known protein sequences in up to eight different biological samples in a single experiments [28]. However, due to its limited throughput and current cost, iTRAQ is not feasible to simultaneously compare large sample sizes of disease subjects to achieve the discovery of differential features of sufficient statistical power. In addition, the success of iTRAQ efforts depends on the peptide sequence determination. Despite our increased understanding of the urine peptidome composition, urine peptidome sequence characterization by Mosaiques Diagnostics [11, 13] and our database (Stanford University), combined, could only determine a small portion (1/10) of the $\sim$ 100,000 peptide features revealed by the HPLC or CE coupled mass spectrometric analysis. Therefore, iTRAQ leads to undersampling and incomplete analytic coverage of the urine peptidome. We currently discourage urine peptide biomarker screening employing the isotope labeling approach.

As an alternative, label-free LC–MS-based approach has been applied as a quantitative biomarker discovery method for the experiment design of large sample size to statistically validate the results. The label-free LC–MS approach can compare and quantify peptides with precision and accuracy comparable to those based on isotope labeling [29]. Utilizing a strategy of ion mapping to uncouple the MS and MSMS processes, the label-free approach is an unbiased approach to identify differential peptide features in

which the peptides are selected on the basis of discriminant analysis of MS signal intensities and then subjected to extensive MSMS sequence identifications. Thus, the label-free quantitative LC–MS more robustly analyzes the full potential of the urine peptidome as a source of disease biomarkers.

The label-free LC–MS approach involves the comparison of urine peptidomes of different samples, and thus, multiple LC–MS spectra. However, comparing multiple LC–MS spectra in a label-free analysis is computationally intensive, demanding robust detection of LC–MS peaks, alignment of all LC–MS peaks, and determination of the common peak indices across all assayed samples. One analytical strategy to generate the peptide indices across all assayed samples is to determine them experimentally. One method of this approach [19] is to identify multiply charged peptide ions that are reported repeatedly by LC–MS/MS analysis of the pooled samples of the same disease category. Defined as the peptide indices across samples, targeted acquisition and quantification of these peptide ions will then be performed for individual samples and compared. Another method of this approach [20] is to use the FT-ICR spectrometry to survey samples to reveal peptide MS peaks with high accurate masses (mass deviation < 1 ppm), and then define these peptide peaks as the common peptide indices across all assay samples. Peptide peaks within different spectra across all assayed samples are assumed identical if mass deviations were within 50 ppm error for monoisotopic and within 75 ppm error for unresolved peaks.

We have employed an algorithm MASS-Conductor© (Copyright © 2008, Ling) to computationally detect MALDI peaks from raw MS datasets, align all sample spectra, and define common peaks as peptide peak indices for comparative analysis. Give the large amount of raw spectrometric data, for example, a 40 sample peptidomics study raw data encompasses 241.5 GB, robust automatic high-throughput data management and data reduction methods are critically needed. In LC–MALDI urine peptidomics analyses, the m/z (mass-to-charge ratio) ranges were from 800 to 4000 with peak density of maximum 30 peaks per 200 Da, minimal S/N ratio of 5, minimal area of 10, minimal intensity of 150, and 200 maximum peaks per LC fraction. The MS peaks are located in the raw spectra of the MALDI data by an algorithm [30] that identifies sites (mass-to-charge ratio, m/z values) whose intensity is higher than the estimated average background and the $\sim$100 surrounding sites, with peak widths $\sim$0.5% of the corresponding m/z value. To align peaks from the set of spectra of the assayed samples, we applied linkage hierarchical clustering to the collection of all peaks from the individual spectra [31]. The clustering, computed on a 24 node LINUX cluster, was 2D, using both the distance along the m/z axis and the HPLC fractionation time, with the concept that tight clusters represent the same biological peak that has been slightly shifted in different spectra. We then

extracted the centroid (mean position) of each cluster, to represent the ''consensus'' position as the peak index (bin) across all spectra. As an example of the data reduction via peak detection, MS data points within one subject's LC–MALDI spot/fraction 13 were compared before and after the peak finding and indexing processes which reduced 2530 data points (top panel) to 62 (bottom panel) peak points (Fig. 2A; *m*/*z* 1200–1500) and from 118,142 data points (left panel) to 1690 (right panel) peak points (not shown, *m*/*z* 900–4000). Despite massive data reduction, the overall LC–MS peak profiles were accurately captured. Using either the raw data points

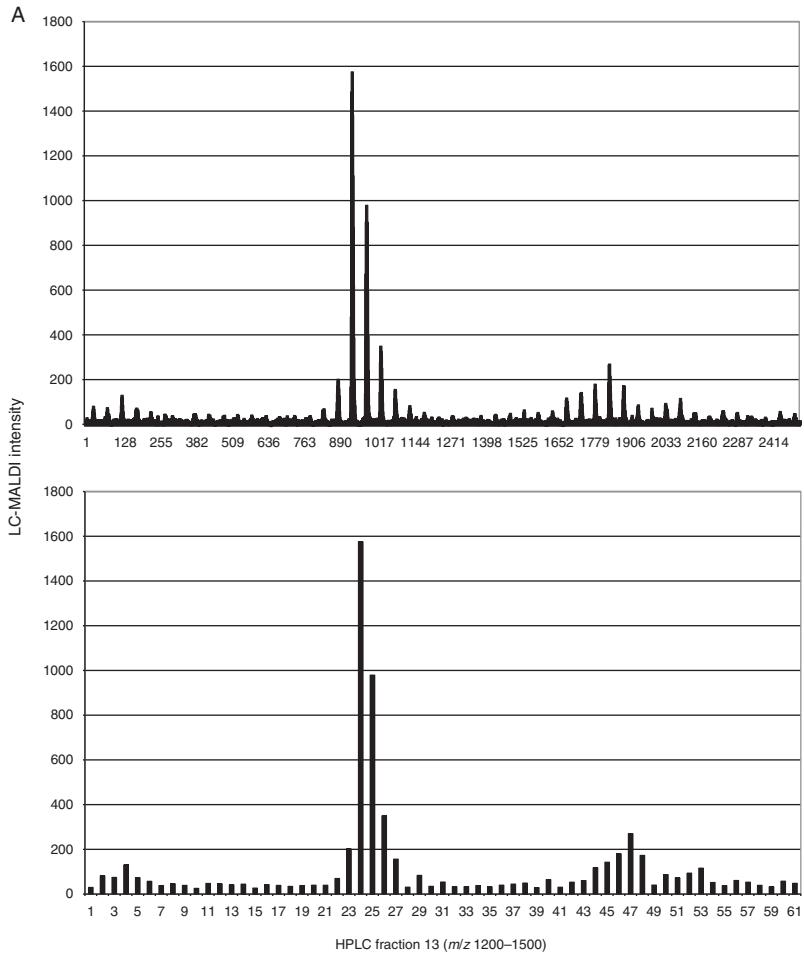


FIG. 2. (Continued)

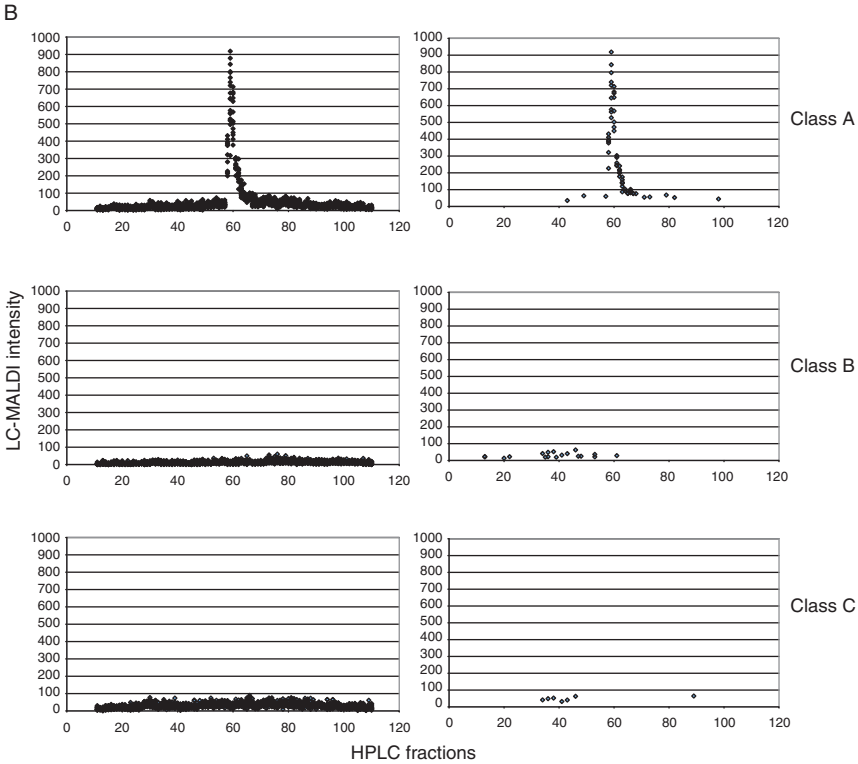


FIG. 2. Data reduction examples after peak finding and peak indexing in the LC–MALDI-based urine peptidome analysis. (A) LC–MS profiles, before (left) and after (right) the data reduction, of one urine sample in HPLC fraction 13 with m/z between 1200 and 1500. (B) Three different class urine samples' LC–MS profiles of a candidate differential peptide before (left) and after (right) the data reduction across all HPLC fractions.

(left panel) or the peak indices (right panel), Fig. 2B plotted LC–MALDI intensity against HPLC fractions of a candidate differential peptide peak across three subjects belonging to three different patient populations. Together these illustrate the necessity of the data reduction and the effectiveness of data processing algorithm. The output of data processing is essentially a P $\times$ N table in which each of P peptides has been quantified in each of the N study sample. As outlined in Fig. 3, this table, reduced from LC–MS spectra of all samples, can be subjected to downstream statistical learning including transformation, normalization, and unsupervised/supervised analyses suited to the experimental design to mine for a differential subset of the P peptides, which will then be subjected to MSMS protein sequence identification and future quantitative prospective MRM validations [32, 33].

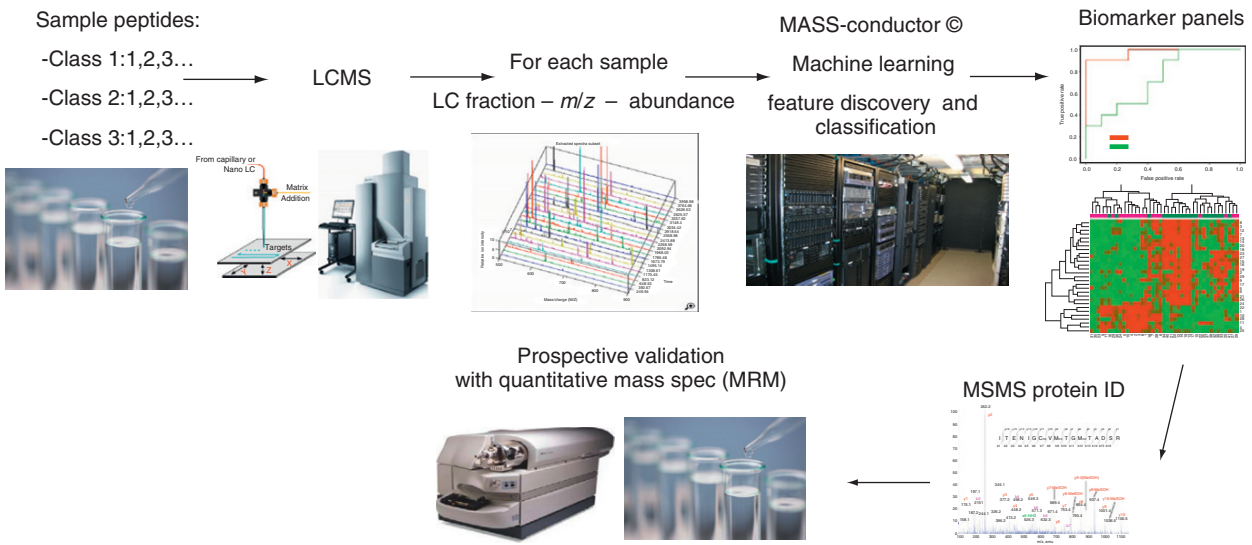


FIG. 3. Diagram of a typical LC–MS-based label-free peptidomics biomarker discovery. Class, different disease or disease condition; *m*/*z*, mass-to-charge ratio; MRM, multiple reaction monitoring; LC, liquid chromatography.

5. Urine Sample Handling and Storage

To minimize potential bias or confounding factors, it is essential that urine samples are collected and handled in standardized ways. However, significant variations exist between different researchers' proposed collection methods, and different clinical labs' practice of the same protocol may not be consistent as required. All of these sample-handling differences, even the minor ones, can have profound impact on the outcomes of urine peptidome discovery. To address concerns that centrifugation of urine samples\prior to freezing at some study sites might present an obstacle, we performed an experiment to determine the effect on the urine peptidome of freezing urine prior to centrifugation. Five healthy volunteers' urines were pooled and divided into 20 equal aliquots. Five aliquots were used for each of the four experimental arms (Fig. 4A): spin/freeze/thaw/process (S/F/T), freeze/thaw/spin/process (F/T/S), RT (room temperature storage) 4 h/spin/freeze/thaw/process (RT/S/F/T), and RT 4 h/freeze/thraw/spin/process (RT/F/T/S). Each aliquot was profiled by LC–MALDI and analyzed by MASS-Conductor© algorithm as described. The results of the analysis are shown in Fig. 4B. Urine peptidomic features between S/F/T and F/T/S samples were compared by Student's *t*-test analysis. Thresholds of *t*-value of 3 or -3 (P value $= 0.01$) were chosen to select potentially differential features between the two classes. This leads to 0.81% (red star) and 0.55% (blue star) potentially "differential" features, totaling 1% of peptide features. To evaluate these potential differential features in the context of multiple hypothesis testing, the combined dataset of S/F/T and F/T/S was permuted 20 times. Student's *t*-test was applied to each of the permuted, therefore, random datasets with the same thresholds of *t*-values to identify "falsely discovered" differential features. The 20 permuted data results were plotted (Fig. 4B) as box–whisker graphs with the originally identified "differential" features (red and blue stars). The originally identified "differential" features are observed to fall in the range of the false discovered ones. Therefore, we conclude there are no significant differences between S/F/T and F/T/S samples. A similar analysis led to the same conclusion in regard to RT/S/F/T and RT/F/T/S. From these analyses, we conclude that the order of freezing and centrifugation does not significantly affect the urine peptidome and therefore urine can be frozen at the sites within 4 h of collection without prior centrifugation. This simplified collection protocol will allow more consistent sample collection and handling.

In another explorative study, we focused on the mass spectrometric analyses of the impact of varied durations of RT storage on the analytical reliability and reproducibility for the urine peptidome contents. Five healthy volunteers' urines were pooled and divided into 30 equal aliquots. Five

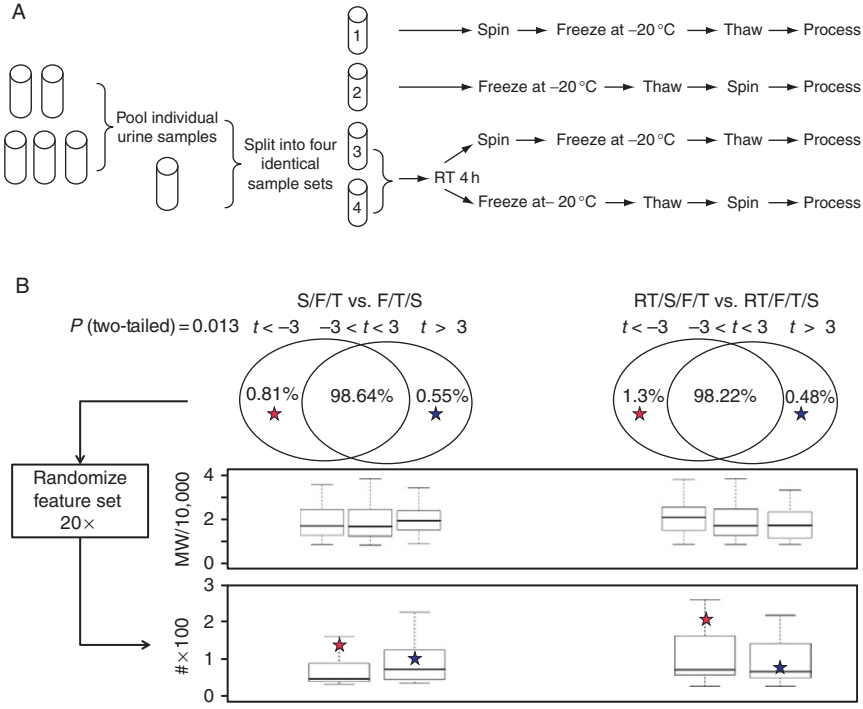
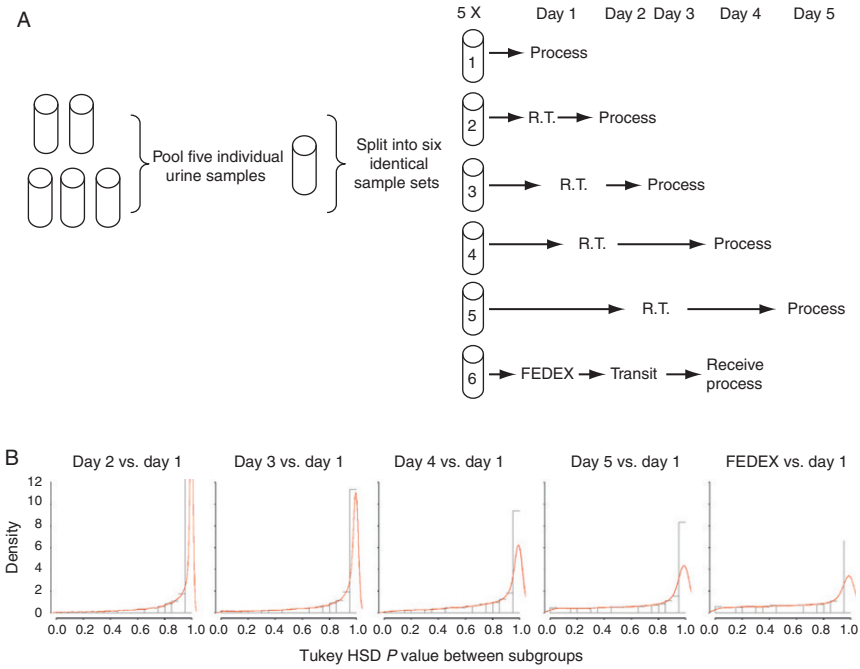


FIG. 4. The order of freezing and centrifugation in urine sample processing does not significantly affect the urine peptidome. (A) Experimental design to evaluate the impact of different urine handling protocols. (B) Comparative analysis of samples through different processing combinations including spin (S), freeze (F), thaw (T), and room temperature storage for 4 h (RT). Potential "differential" features: red, blue stars.

aliquots were used for each of the six experimental arms (Fig. 5A): process immediately (day 1 urine), RT storage durations from 1 day (day 2 urine) up to 4 days (day 5 urine), and urine sample transported using Federal Express at RT leaving Stanford at day 1 and arriving at day 3 at Palo Alto (California USA, 2 miles away) (FEDEX urine). Each aliquot was profiled by LC–MALDI and analyzed by MASS-Conductor© algorithm as described. Analyses of samples RT stored for the same length of time yielded consistent urine peptidomes (data not shown). When we compared day 1 urine peptidome with those stored with different RT storage days, the chance of finding a "significant" difference just by serendipity increases, therefore, Tukey HSD (Honestly Significant Differences) test was used. Shown in Fig. 5B, the Tukey HSD *P* values of the urine peptide features, comparing day 1 and other five

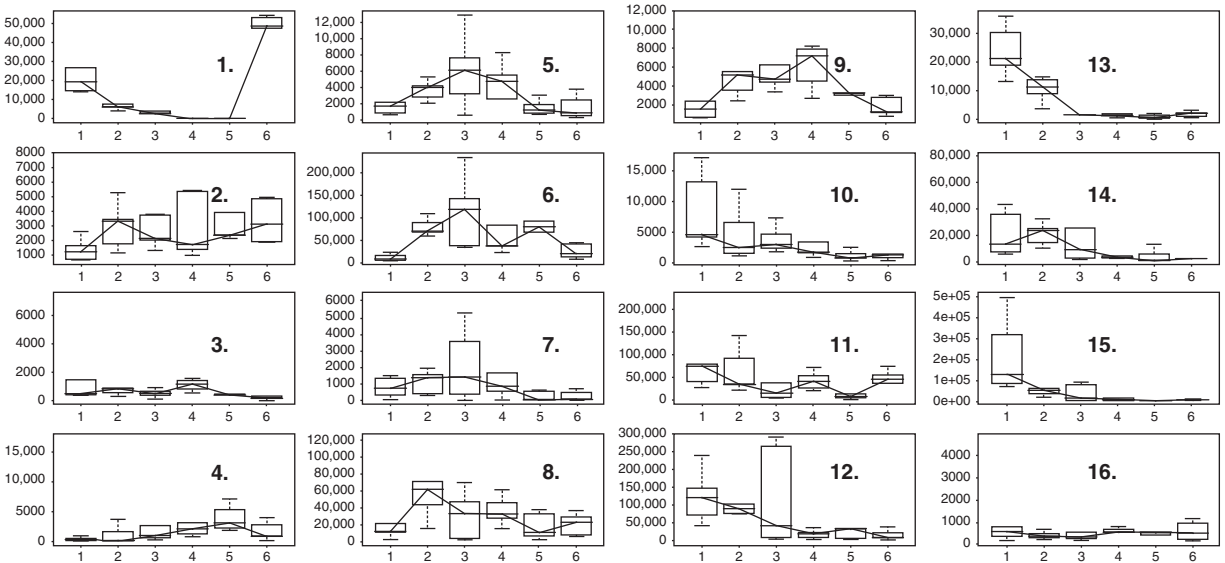


C

#	Peptide	MH+	Sequence
1.	UMOD	982.59	VLNLGPITR
2.	UMOD	1047.48	SGSVIDQSRV
3.	UMOD	1211.66	DQSRVLNLGPI
4.	UMOD	1225.69	SRVLNLGPITR
5.	UMOD	1324.76	IDQSRVLNLGPI
6.	UMOD	1423.83	VIDQSRVLNLGPI
7.	UMOD	1468.82	DQSRVLNLGPITR
8.	UMOD	1510.87	SVIDQSRVLNLGPI
9.	UMOD	1567.91	GSVIDQSRVLNLGPI
10.	UMOD	1581.91	IDQSRVLNLGPITR
11.	UMOD	1654.91	SGSVIDQSRVLNLGPI
12.	UMOD	1680.98	VIDQSRVLNLGPITR
13.	UMOD	1755.96	SGSVIDQSRVLNLGPIT
14.	UMOD	1768.01	SVIDQSRVLNLGPITR
15.	UMOD	1912.07	SGSVIDQSRVLNLGPITR
16.	UMOD	2040.16	SGSVIDQSRVLNLGPITRK

FIG. 5. (Continued)

D


1.
50,000
40,000
30,000
20,000
10,000
0
1 2 3 4 5 6
5.
12,000
10,000
8000
6000
4000
2000
0
9.
12,000
10,000
8000
6000
4000
2000
13.
30,000
20,000
10,000
0
2.
8000
7000
6000
5000
4000
3000
2000
1000
6.
200,000
150,000
100,000
50,000
0
10.
15,000
10,000
5000
0
14.
80,000
60,000
40,000
20,000
0
3.
6000
4000
2000
0
7.
6000
5000
4000
3000
2000
1000
0
11.
200,000
150,000
100,000
50,000
0
15.
5e+05
4e+05
3e+05
2e+05
1e+05
0e+00
4.
15,000
10,000
5000
0
8.
120,000
100,000
80,000
60,000
40,000
20,000
0
12.
300,000
250,000
200,000
150,000
100,000
50,000
0
16.
4000
3000
2000
1000
0

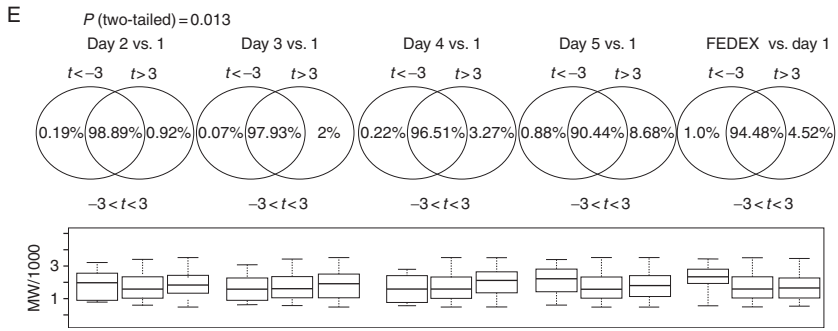


FIG. 5. Exploration of the impact of durations of the room temperature (RT) storage on the urine peptidome contents. FEDEX: at day 1, one tube of urine was fedexed out, from Stanford University and to Palo Alto (2 miles away) in California USA, and was received at day 3. (A) Experimental design. (B) Tukey HSD (Honestly Significant Differences) analysis of samples of different RT storage durations. (C) Uromodulin (UMOD) C-terminal urine peptides. (D) UMOD peptide abundance quantified by LC–MALDI signals in samples of different RT storage durations. (E) Comparative analysis of samples through different RT storage durations to the immediate processed samples.

urine peptidomes respectively, became progressively more spread between 0 and 1. This result indicated that the longer the storage time at RT, the peptide features became progressively more differential in general. Among the urines, the FEDEX urine contents changed the most, which may be due to the uncontrolled transit environment the urine samples were exposed to. To survey the urine peptidomes, we have analyzed 16 different uromodulin (UMOD) C-terminal peptides (Fig. 5C) cleaved from the UMOD protein precursor after shedding from the apical plasma membrane into the tubule lumen [34]. Shown in Fig. 5D, there are three groups of kinetic patterns of peptide signal change: Group 1 peptides #1, 10, 11, 12, 13, 14, 15, 16, of which the peptide LC–MS signal decreased from day 1 to day 5; Group 2 peptide #4, of which the peptide LC–MS signal increased from day 1 to day 5; Group 3 peptide #2, 3, 5, 6, 7, 8, 9, of which the peptide LC–MS signal increased from day 1 to day 2 or 3 then kept decreasing afterward. This survey of peptides derived from the same origin showed that most of the peptides' signal decayed over time indicating time-dependent degradation. Some of the peptides' signal peaked at day 2 or 3 RT storage time suggesting that prolonged peptide degradation caused the accumulation of the smaller degradation intermediates, which in turn were subject to further downstream degradation. Shown in Fig. 5E, all urine peptidomic features between day 1 and other RT storage times or FEDEX samples were compared by Student's

t-test analysis. Thresholds of *t*-value of 3 or -3 (P value $= 0.01$) were chosen to select potentially "differential" features between the day 1 urine and others. The Venn diagram analysis demonstrates the progressive increase in number of differential peptide features with longer storage at RT. However, over longer RT storage, there was no obvious change of the distribution of the molecular weight (MW; Fig. 5E bottom panel) with differential (*t*-value less or greater than 3) or nondifferential (*t*-value between -3 and 3) peptide features, indicating the degradation of peptides was global across different sizes. The observation of the nondifferential peptide features remained 90.44% of all peptide features even after 4 days of RT storage, suggesting a significant pool of urine peptidomes remained largely undifferentiated across different RT storage time points. Our results indicate that, if unable to be frozen immediately after collection, short-term RT storage urine samples can be allowed until freezing. However, when considering RT storage duration as a variable in urine biomarker analyses, we recommend that the case and control samples be handled in the exact same manner throughout the study processes to avoid preanalytical bias.

6. Do "Housekeeping Peptides" Exist in Urine Peptidome?

Urine peptidomics analyses suffer two major different origins of variance [20]: analytical issues including mass spectrometric ion suppression; biological issues including dilution of urine by different hydration states of the urine donors. In gene expression analysis, housekeeping genes, for example, actin, GAPDH, and ubiquitin, are genes, typically needed for maintenance of the cell, therefore, constitutively transcribed at a relatively constant level across many or all known conditions. Given that their expression is unaffected by experimental conditions, housekeeping genes are commonly used for expression normalization to correct biological and analytical variances. Previous CE-MS analysis [20] of the urine peptidome found 29 endogenous collagen-derived peptides, with mass evenly distributed between 1000 and 2900 Da, capable of serving as "housekeeping" peptides that can sufficiently address both analytical and biological (mainly the urine dilution) variance during the biomarker analyses of macroalbuminuria, normoalbuminuria, and nondiabetic subjects. To explore whether this 29-collagen-peptide panel can be utilized as the "housekeeping" peptides to normalize other disease samples, we have applied this panel to normalize the LC–MALDI profiled urine peptidomes from 130 subjects of SAF (SJIA with both systemic and arthritis flare, $n = 36$), AF (SJIA with arthritis flare, $n = 17$), QOM (SJIA quiescence but still on medicine, $n = 20$), V (healthy volunteer, $n = 10$), KD ($n = 24$), and FI (febrile illness, $n = 23$) diseases.

As a simulation estimation, the mass region between 900 to 4000 Da was divided into 29 intervals, and from each interval we selected one peptide randomly which led to a panel of 29 peptides. This random panel construction process was repeated 500 times resulting 500 random 29-peptide panels. These random 29-peptide panels were also utilized to normalize the 130 urine peptidomes. To gauge the effectiveness of the normalization, coefficient of variation (CV) was calculated for all of the urine peptide features before and after the normalization. A successful normalization is expected to reduce the signal variations across samples of different disease states, and the CV distribution density peak should consequently shift to less CV values. As expected, all 500 random 29-peptide panels (Fig. 6), revealed by 10, 50, and 90 percentile density plots, increased the global variations as CV density peaks shift to larger CV values. However, the 29-"housekeeping"-peptide panel shift, beyond the random panels, the CV density peak to higher values, indicating that the previously [20] described 29-"housekeeping"-peptide panel increased global variations of our assayed urine peptidomes. Our results demonstrate that the previously described 29-collagen peptides, at least, cannot be utilized as the "housekeeping" peptides in the systemic pediatric disease subject normalization process. Whether universal housekeeping peptides exist or not in urine peptidomes still remained to be answered and yet to be explored.

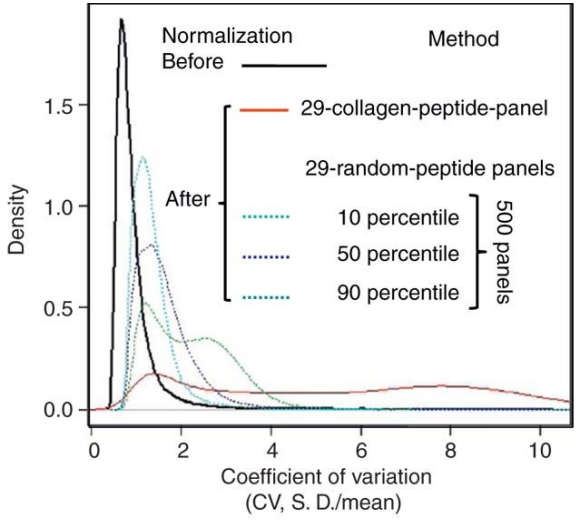
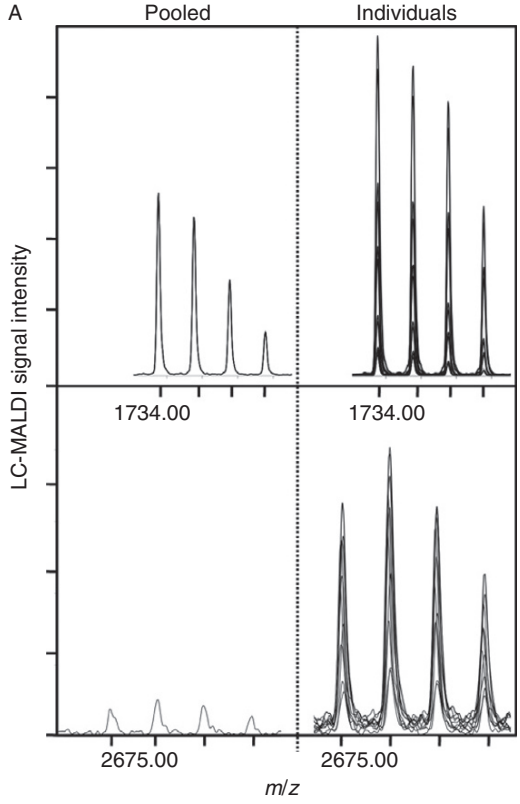


FIG. 6. Normalization of 130 urine peptidomes of six class categories using either previously described 29 endogenous collagen "housekeeping" peptide panel or 500 panels of randomly selected 29-peptide sets.

7. To Pool or Not Pool, Practical Considerations of Benefits, Risks and Biases

The biomarker experimental design usually demands sufficiently large sample size to achieve required analytical power. Label-free LC–MS-based urine peptidome profiling generates large amount of data bringing significant analytical challenges to the downstream data mining analysis. Therefore, to pool the samples within the same subject class before LC–MS analysis was proposed as a cost reducing approach to reduce the number of LC–MS runs, making it possible for additional multidimensional separations to detect even lower abundance species, and allowing higher laboratory throughput. The pooling strategy relies on the unsupported assumption that pooling samples averages their contents. Statistical analysis in microarray studies has found pooling to approximate individually run samples at the cost of statistically robust results and a significant loss in overall transcription change discoveries [35–38]. The effect of pooling in proteomics analysis was examined in SELDI-TOF profiling of serum [39], which noted a loss in the number of differential masses after pooling, and low abundance biomarkers were more susceptible to the deleterious effects of pooling than higher abundance biomarkers. We have explored the pooling strategy in the profiling of [27] of urine samples (AR, acute rejection; HC, normal protocol biopsies; STA, stable renal graft function), and found urine peptide biomarker candidates with differential fold of abundance among pooled sample categories. We expected that pooling affects data quality and inference in urine peptidomics, but the exact effects are not yet quantified. As an example, the comparison of urine peptide MH+ 1734's LC–MALDI profiles (Fig. 7A top panel) between the pool and the individual normal urines supports the notion that pooling samples average the individual samples' peptide contents. However, the pooled LC–MALDI profile of urine peptide MH+ 2675 (Fig. 7A, bottom panel) showed a much lower signal than that of the individuals, which most likely was due to the ion suppression effect [40] in mass spectrometry. To systematically qualify and quantify the pooling effects on biological conclusions in the context of urine peptidomic experiments, we designed a study to evaluate the urine peptidomics changes between normal ($n = 10$), microalbuminuria ($n = 10$), and nephrotic syndrome ($n = 7$) subjects. The ultimate aim is to find urine-based biomarkers capable of diagnosing nephritic syndrome or other proteinuric diseases (Drs. Sutherland, Ling, Cohen, Stanford University, ongoing study). To investigate the impact of pooling, differential urine peptides between normal, microalbuminuria, and nephrotic syndrome were identified both the fold change using pooling method and the Student's *t*-test comparing all individuals. Data points outside the two vertical lines (1st and 99th) mark the top 2% of urine peptides selected by the pooling fold

change method. Data points (Fig. 7B) outside the two horizontal lines (1st and 99th) represent the top 2% of urine peptides, to be the most reliably altered between the two contrasting classes. For discussion purposes, data



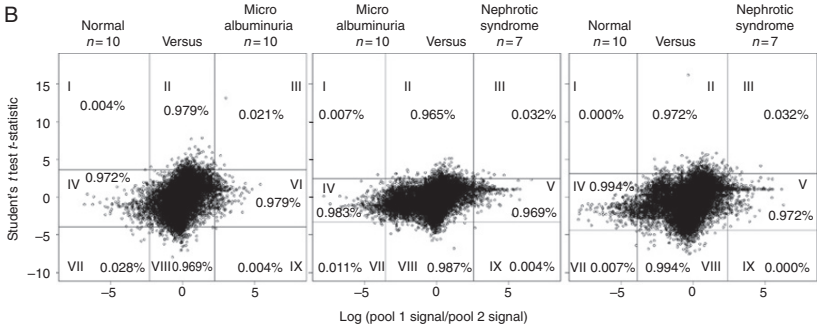


FIG. 7. (Continued)

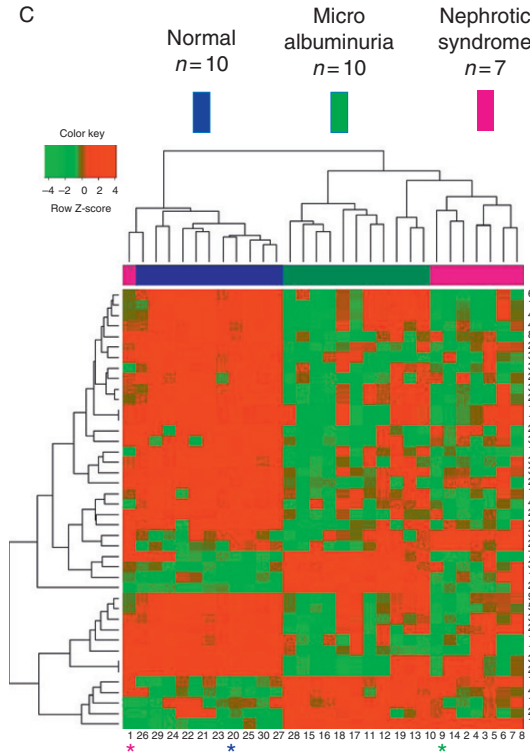


FIG. 7. LC–MALDI profile comparisons of urine pooled samples and individuals. (A) Comparative analysis of the urine peptide's isotopic envelop between the signals of the pool sample and the overlaid signal values of the corresponding individuals. *Top panel*: urine peptide 1734. *Bottom panel*: urine peptide 2675. (B) Selection of urine peptides that are differential between assayed sample categories, based fold changes on pools (the *x*-axis) or by the Student's *t*-test statistics on individuals (the *y*-axis). The two vertical lines denote the 1st and 99th percentiles of the fold differences between the pools of the two compared sample categories. The differential urine selected by fold difference using the pooling method fall outside these two vertical lines. The two horizontal lines denote the 1st and 99th percentiles of the Student's *t*-test *t*-statistics differences between individuals of the two compared sample categories. Nine sectors are designated I through IX on the diagram as follows: (I, III, VII, IX) urine peptides selected by both methods. (IV, VI) urine peptides selected only by the fold difference using the pooling method. (II, VIII) urine peptides selected only by the Student's *t*-test *t*-statistics differences analyzing all individuals of the compared categories. (V, center sector and not labeled, where majority of the peptides cluster) urine peptides selected by neither method. (C). Unsupervised analysis of normal, microalbuminuria, nephrotic syndrome peptidomes together with the three-class sample pools (labeled with matched color stars) using a biomarker panel of 42 urine peptides capable of differentiating these three-class subjects.

points found by the individual profiling-based method are considered as ''true positives'' (sectors I, II, III, VII, VIII, IX). Those found by pooling method only (sectors IV, VI) are considered as ''pooling false positives'' and those found by individual profiling-based method only (sectors II, VIII) are considered as ''pooling false negatives.'' Data points in sectors I, III, VII, and IX, total of 0.057%, 0.054%, and 0.039% of all urine peptides respectively, generally of low number, represent the differential peptides selected by both the pooling- and individual-based methods contrasting normal versus microalbuminuria, microalbuminuria versus nephrotic syndrome, and normal versus nephrotic syndrome categories. In contrast, the pooling false positives and false negatives of all urine peptides are 1.951% and 1.948% respectively contrasting normal versus microalbuminuria, 1.952% and 1.952% respectively contrasting microalbuminuria and nephrotic syndrome, 1.966% and 1.966% respectively contrasting normal and nephrotic syndrome. All of these indicate the ineffectiveness of pooling method resulting in a loss of sensitivity and an increase of false positives. PAM algorithm [41] has been applied to the three-class (normal $n = 10$, microalbuminuria $n = 10$, nephrotic syndrome $n = 7$) peptidomes, leading to a biomarker panel of 42 urine peptides capable of differentiating these class subjects. Upon this urine peptide biomarker panel, an unsupervised heatmap analysis (Fig. 7C) was performed using all individual samples and the class pools. Individual subjects of the same sample category effectively clustered together. However, the pooled nephrotic syndrome sample (red star labeled) obviously averaged the heterogeneous individuals in this disease category to cluster with normal samples, therefore, biomarkers indicative of nephrotic syndrome may not be able to found by the pooling strategy. To conclude, we discourage the pooling strategy as statistically invalid and recommend the use of nonpooled (individual) samples for urine peptidomics analysis to mine for statistically significant urine peptide biomarkers.

8. Multiple Hypothesis Testing, False Discovery, and Bootstrapping Analysis

The process of biomarker discovery can be seen as a concurrent statistical test of thousands of null hypotheses, where each peptide peak in the spectrum is a hypothesis to be evaluated. This leads to the multiple testing problem, demanding that the derived test statistics be adjusted to control the expected proportion of false discoveries among all discoveries. This can be achieved either by the overly conservative Bonferroni correction or an analysis of the global false discovery rate (gFDR) [42]. After determining the gFDR test threshold for significance, the local FDR (lFDR) analysis can compute and

assign significance measures to all features. The lFDR analysis [43] addresses one drawback of the gFDR, statistically distinguishing features that are close to the threshold and therefore more likely to be falsely positive from those that are not. Equipped with high computation power and implemented with a permutation-based method [44], the Stanford FDR server (http://translationalmedicine.stanford.edu/Mass-Conductor/FDR.html) has been setup to analyze for FDR, differential abundance in proteome/peptidome/genome analysis, and for the statistical correlation between molecular data and clinical measurements.

It is unlikely that a single urine peptide, selected from the $\sim$100,000 mass spectrometric peptide features, can fulfill the clinical diagnostic/prognostic needs. Most, if not all, of the peptide features ranked by gFDR and lFDR analyses lack the required sensitivity and selectivity. Collectively as a biomarker panel, markers cherry picked by empirical or machine-learning approaches work in concert yielding much higher discriminating power. However, the current gFDR and lFDR analyses are not tailored to compute and assign significance measures to the final biomarker panel. In addition, having so many peptide features relative to so few samples, creates a high likelihood that a given specific sample not fully representative of the population can easily distort the statistical inference. Therefore, there is a significant need for robust statistical methods to address these analytical concerns and challenges.

In the study of NEC (a major cause of neonatal morbidity and mortality [45]) to discover biomarkers that reliably distinguish infants with NEC (medical group, M) from infants with NEC and most likely to progress to severe disease requiring immediate surgery (surgical group, S), we have identified a panel of 13 urine peptides (Ling and Sylvester, unpublished data; Fig. 8A and B; Fisher exact P value 2.5×10^{-7}). The binned LC–MALDI MS peak data obtained for all 34 urine peptidomes (NEC M $n = 17$, S $n = 17$) were analyzed for discovery of discriminant biomarkers using algorithms [41] of nearest shrunken centroid (NSC) for biomarker feature selection, 10-fold cross-validation analyses, and Gaussian linear discriminant analysis (LDA) for classification analyses. To avoid bias in samples where outliers may distort statistical inference, we utilized a bootstrapping (resampling with replacement) technique that resampled the 34 urine peptidomes 500 times to construct 500 biomarker panel datasets. For each of the bootstrapping set, 500 different LDA classifiers were subsequently built for ROC analysis [46, 47]. To summarize the results, the vertical average of the 500 ROC curves was plotted, and the boxes and whiskers were used to describe the vertical spread around the average (Fig. 8C). The mean of the AUCs of 500 ROC analyses is 98.5%, indicating that the statistic learning to discover and develop biomarker panel classifier has unlikely been distorted by sample outliers.

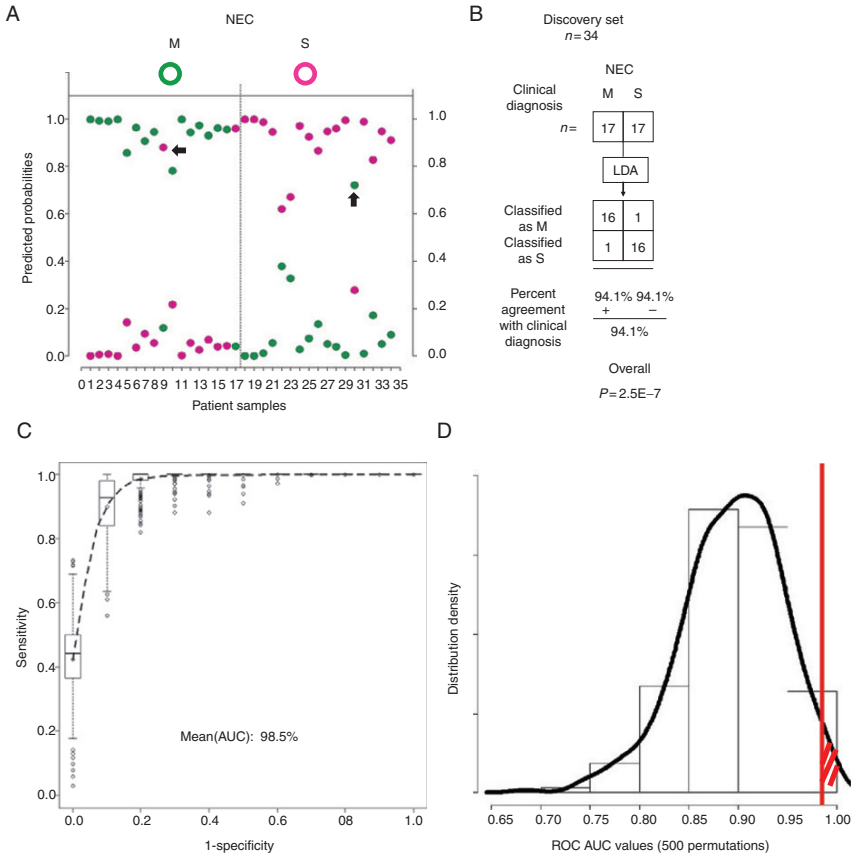


FIG. 8. Significant analysis of the urine peptide biomarker panel differentiating NEC medical (M) and surgical (S) categories. (A) The discriminant probabilities of the NEC urine peptide biomarker panel. The maximum estimated probability for each of the wrongly classified samples is marked with an arrow. (B) Modified 2 × 2 contingency tables were used to calculate the percentage of classification that agreed with clinical diagnosis by the urine peptide biomarker panel. *P*-values were calculated with Fisher's exact test. (C) ROC analysis of the NEC urine peptide biomarker panel in discriminating NEC M and S classes. AUC: area under the curve. The NEC urine peptide biomarker panel dataset was bootstrapped (resample with replacement) 500 times to create 500 datasets. The dotted curves is the vertical average of the 500 bootstrapping ROC curves and the boxes and whiskers plot the vertical spread around the average. (D) Distribution of the standardized ROC AUC values of the 500 falsely discovered panels. Examining all the 500 falsely discovered biomarker panel ROC AUC values, there are only 12 falsely discovered panels that have ROC AUC values greater than that of the original urine biomarker panel (represented by the red vertical line).

Since the 13 peptide NEC biomarker panel was selected from the ~10,000 unique peak features, we set to address the biomarker panels' multiple hypothesis testing problem. In order to estimate the false discovery rate (FDR) in concurrent statistical tests of peptide panels, of the same size as our biomarker panel, the class labels of our training dataset samples were permutated 500 times such that each time every sample would be randomly assigned a new class label (NEC M or S). For each of the 500 simulated "training" sets, NSC algorithm was applied to rank all the peak features upon their discriminating the binary classes. The top 13 peak features were then designated as the "panel" for LDA analysis. ROC analysis subsequently was used to calculate the AUC for this "falsely discovered panel." The AUC values of the 500 falsely discovered panels were standardized, and the density distribution was plotted in Fig. 8D. Examining all the 500 AUC values, there are only 12 falsely discovered panels that have AUC values greater than 98.5% (found for the original 13-peptide biomarker panel biomarker). This method estimates significance measure and compute the targeted biomarker panel's FDR. Therefore, the FDR of our NEC peptide biomarker panel is estimated as 2.4%, supporting the notion that the discovery of our peptide biomarker panel is unlikely to be the outcome of chance.

9. Exploration of Urine Peptide Biomarkers as Predictors of Drug Response

One rationale for our focus on urine is our long-term intent to use urine biomarkers for detection of (subclinical) disease activity and to predict drug responses; such tests would be feasible for frequent determination, especially in children. SJIA is a chronic inflammatory disease of childhood characterized by a combination of systemic features and arthritis [48, 49]. In published reports of clinical observations, a subset of SJIA patients respond to therapies that are effective in polyarticular JIA/RA, for example, methotrexate and TNFα-inhibitors [50, 51], and only ~50% of subjects are persistent responders to IL-1 inhibition [52, 53]. To test the hypothesis that urine peptide biomarkers can predict drug (Enbrel®—TNF inhibition; Anakinra®—IL-1 inhibition) response, we carried out a pilot study including unsupervised clustering analysis (Fig. 9) using pretreatment urine peptide profiles (5 Enbrel CR-"red," 5 Enbrel PR-"green," and 3 Anakinra CR-"blue"; CR = complete responder; PR = partial responder). When comparing pretreatment urine peptide profiles from subjects with CR and PR to TNF inhibition, there are essentially three groups of peptides: Group A and

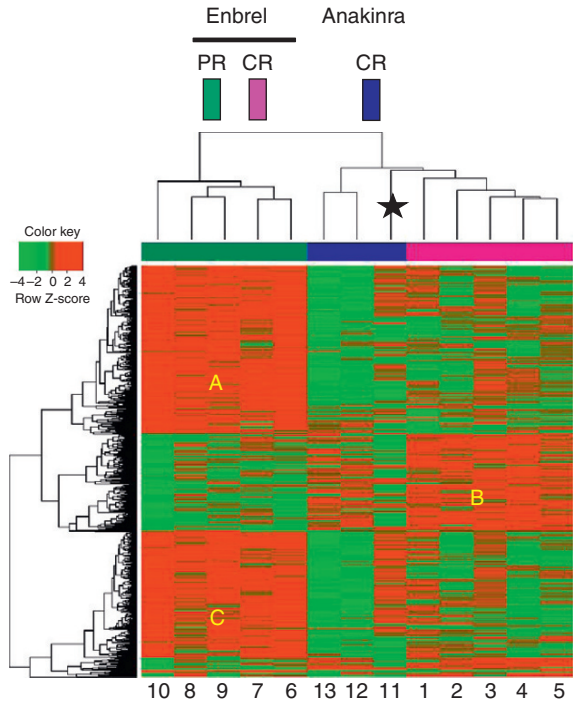


FIG. 9. Unsupervised analysis of the pretreatment urine peptidome profiles revealing urine peptide abundance patterns indicative of SJIA patient drug response. Enbrel CR-"red," $n = 5$; Enbrel PR-"green," $n = 5$; Anakinra CR-"blue," $n = 3$; CR: complete responder; PR, partial responder.

C peptides are abundant in patients later found to be only partially responsive to TNF blockade; Group B peptides are abundant in patients found to respond to TNF blockade. Interestingly, the urine samples from CRs to IL-1 blockade have both common and discriminative profiles with CRs to TNF blockade, such that CRs to IL-1 to blockade form a distinctive class rather than clustering with PRs to anti-TNF. Careful examination of the heatmap revealed that the asterisked sample has a unique urine profile, where its group B and C peptide profiles are similar to samples from PRs to TNF blockade. These preliminary results demonstrate that several different patterns exist in pretreatment urine peptide profiles, which implies that there may be mechanistically distinct subgroups within SJIA patients. The number of profiled samples in the pilot study is small; however, the suggestion is that these profiles may predict the response to treatment with anti-TNF or IL-1

inhibition. Future prospective studies with more study subjects will test the robustness of urine as a source of drug response biomarkers in SJIA and also is likely to provide new insights into the pathogenesis of this disease.

10. Urine Peptidome Proteolytic Degradation Patterns Reflecting Pathophysiology

Our previous integrated analysis [28] of the urine peptidome and the biopsy transcriptome in graft rejection that uncovers that overlapping key gene and peptide pathways can be jointly dysregulated in AR. Disease-specific alterations of gene transcription in the tissue (by array and Q-PCR) and a change in the balance of proteolytic and antiproteolytic activities in urine appear to imply important mechanisms resulting in an altered pattern of a specific panel of urinary peptides in AR. For both systemic and renal diseases, we hypothesize, as diagrammed in Fig. 10, that urine peptide biomarkers are the surrogates of the pathophysiological dysfunctions in signaling, proteolytic and antiproteolytic pathways. The peptide biomarkers can be the derivatives of plasma proteins, disease specific shedding from other organs, and renal specific proteins, and are generated during the proteolysis that occurs in either circulation during systemic diseases or dysfunctional kidneys, and then trimmed down by exoproteases into ladder-like clusters.

In conclusion, urine peptidomic profilings can yield urine peptide biomarkers discriminating both systemic and renal dysfunctions. However, challenges remain to transform the urine peptide biomarkers obtained at the discovery phase into practical clinical utility. Due to the short length of the urine peptides (900–4000 Da), to develop antibodies for each peptide of the biomarker panel may not be feasible. Quantitative mass spectrometry-based approach, that is MRM [33, 34], is an obvious alternative, however, not widely adopted in clinics. With robust experimental design, future prospective studies, either by antibody-based or quantitative mass spectrometry-based approach, are needed to validate the urine peptide biomarkers currently out of the discovery phase in order to optimize them into practical clinical utility for disease diagnosis and prognosis. The integrative analyses of peptidomics, genomics, and clinical information are critical for the understanding of not only mechanisms by which these urine peptide biomarkers are generated but also the pathophysiology of the diseases. In this regard, noninvasive easy to sample urine peptide biomarkers have the potential to greatly advance current diagnostics and therapeutics in both systemic and renal diseases.

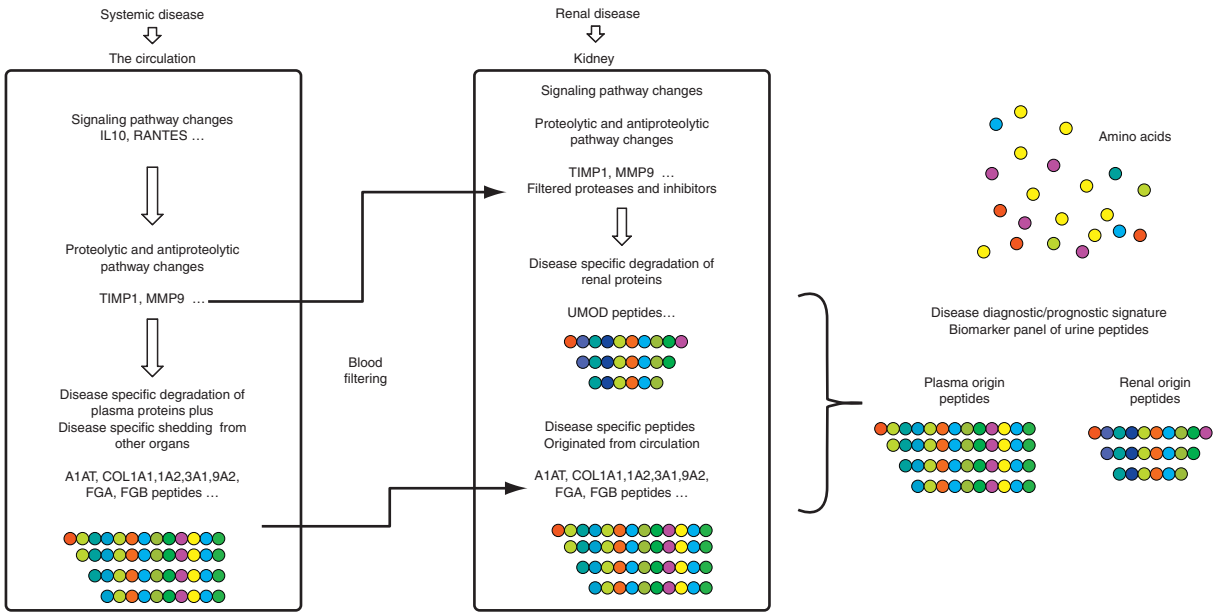


FIG. 10. Urine peptide biomarkers reflect the pathophysiological dysfunctions in signaling, proteolytic and antiproteolytic pathways in systemic or renal diseases. The peptide biomarkers can be the derivatives of plasma proteins, disease specific shedding from other organs, and renal specific proteins, and are generated during the proteolysis that occurs in either circulation during systemic diseases or dysfunctional kidneys, and then trimmed down by exoproteases into ladder-like clusters.

Acknowledgments

The authors thank colleague scientists in Stanford University Pediatric Proteomics Group for critical discussions, and the Stanford University IT group for excellence in Linux cluster support. The authors also thank Edward Chen, Zhen Zhu, David Protter, Chris Xiao, and Roger Lu for data analysis assistance.

References

[1] B.M. Brenner, F.C. Rector, S.A. Levine, Brenner and Rector's The Kidney, 7 ed., Saunders, Philadelphia, 2004.
[2] K.J. Strong, T.M. Osicka, W.D. Comper, Urinary-peptide excretion by patients with and volunteers without diabetes, J. Lab. Clin. Med. 145 (2005) 239–246.
[3] S. Decramer, A.G. de Peredo, B. Breuil, H. Mischak, B. Monsarrat, J.-L. Bascands, et al., Urine in Clinical Proteomics, Mol. Cell. Proteomics 7 (2008) 1850–1862.
[4] P. Schulz-Knappe, M. Schrader, L. Standker, R. Richter, R. Hess, M. Jurgens, et al., Peptide bank generated by large-scale preparation of circulating human peptides, J. Chromatogr. A 776 (1997) 125–132.
[5] P.R. Cutillas, A.G. Norden, R. Cramer, A.L. Burlingame, R.J. Unwin, Detection and analysis of urinary peptides by on-line liquid chromatography and mass spectrometry: application to patients with renal Fanconi syndrome, Clin. Sci. (Lond.) 104 (2003) 483–490.
[6] J. Hardouin, Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry, Mass Spectrom. Rev. 26 (2007) 672–682.
[7] I. Shadforth, D. Crowther, C. Bessant, Protein and peptide identification algorithms using MS for use in high-throughput, automated pipelines, Proteomics 5 (2005) 4082–4095.
[8] E. Mortz, P.B. O'Connor, P. Roepstorff, N.L. Kelleher, T.D. Wood, F.W. McLafferty, et al., Sequence tag identification of intact proteins by matching tanden mass spectral data against sequence data bases, Proc. Natl. Acad. Sci. USA 93 (1996) 8264–8267.
[9] B.M. Balgley, T. Laudeman, L. Yang, T. Song, C.S. Lee, Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy, Mol. Cell. Proteomics 6 (2007) 1599–1608.
[10] R.G. Sadygov, D. Cociorva, J.R. Yates, 3rd. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book, Nat. Methods 1 (2004) 195–202.
[11] J.J. Coon, P. Zurbig, M. Dakna, A.F. Dominiczak, S. Decramer, D. Fliser, et al., CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics, Proteomics Clin. Appl. 2 (2008) 964–973.
[12] J. Adachi, C. Kumar, Y. Zhang, J.V. Olsen, M. Mann, The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins, Genome Biol. 7 (2006) R80.
[13] S. Decramer, A. Gonzalez de Peredo, B. Breuil, H. Mischak, B. Monsarrat, J.L. Bascands, et al., Urine in clinical proteomics, Mol. Cell. Proteomics 7 (2008) 1850–1862.
[14] X.B. Ling, T.K. Sigdel, K. Lau, L. Ying, I. Lau, J. Schilling, et al., Integrative urinary peptidomics in renal transplantation identifies biomarkers for acute rejection, J. Am. Soc. Nephrol. 21 (2010) 646–653.
[15] S.L. Snyder, P.Z. Sobocinski, An improved 2,4,6-trinitrobenzenesulfonic acid method for the determination of amines, Anal. Biochem. 64 (1975) 284–288.

[16] H. Mi, B. Lazareva-Ulitsky, R. Loo, A. Kejariwal, J. Vandergriff, S. Rabkin, et al., The PANTHER database of protein families, subfamilies, functions and pathways, Nucleic Acids Res. 33 (2005) D284–8.
[17] L.U. Zimmerli, E. Schiffer, P. Zurbig, D.M. Good, M. Kellmann, L. Mouls, et al., Urinary Proteomic Biomarkers in Coronary Artery Disease, Mol. Cell. Proteomics 7 (2008) 290–298.
[18] S. Wittke, M. Haubitz, M. Walden, F. Rohde, A. Schwarz, M. Mengel, et al., Detection of acute tubulointerstitial rejection by proteomic analysis of urinary samples in renal transplant recipients, Am. J. Transplant. 5 (2005) 2479–2488.
[19] L.F. Quintana, J.M. Campistol, M.P. Alcolea, E. Banon-Maneus, A. Sole-Gonzalez, P.R. Cutillas, Application of label-free quantitative peptidomics for the identification of urinary biomarkers of kidney chronic allograft dysfunction, Mol. Cell. Proteomics 8 (2009) 1658–1673.
[20] J. Jantos-Siwy, E. Schiffer, K. Brand, G. Schumann, K. Rossing, C. Delles, et al., Quantitative Urinary Proteome Analysis for Biomarker Evaluation in Chronic Kidney Disease, J. Proteome Res. 8 (2009) 268–281.
[21] K. Rossing, H. Mischak, M. Dakna, P. Zurbig, J. Novak, B.A. Julian, et al., Urinary proteomics in diabetes and CKD, J. Am. Soc. Nephrol. 19 (2008) 1283–1290.
[22] K. Rossing, H. Mischak, H.H. Parving, P.K. Christensen, M. Walden, M. Hillmann, et al., Impact of diabetic nephropathy and angiotensin II receptor blockade on urinary polypeptide patterns, Kidney Int. 68 (2005) 193–205.
[23] S. Decramer, S. Wittke, H. Mischak, P. Zurbig, M. Walden, F. Bouissou, et al., Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis, Nat. Med. 12 (2006) 398–400.
[24] D. Theodorescu, S. Wittke, M.M. Ross, M. Walden, M. Conaway, I. Just, et al., Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis, Lancet Oncol. 7 (2006) 230–240.
[25] D. Theodorescu, E. Schiffer, H.W. Bauer, F. Douwes, F. Eichhorn, R. Polley, et al., Discovery and validation of urinary biomarkers for prostate cancer, Proteomics Clin. Appl. 2 (2008) 556–570.
[26] D. Theodorescu, D. Fliser, S. Wittke, H. Mischak, R. Krebs, M. Walden, et al., Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine, Electrophoresis 26 (2005) 2797–2808.
[27] T. Sigdel, X.B. Ling, K. Lau, L. Li, J. Schilling, M. Sarwal, Urinary peptidomic analysis identifies potential biomarkers for acute rejection of renal transplantation, Clin. Proteomics 5 (2009) 103–113.
[28] S.E. Ong, M. Mann, Mass spectrometry-based proteomics turns quantitative, Nat. Chem. Biol. 1 (2005) 252–262.
[29] C.W. Turck, A.M. Falick, J.A. Kowalak, W.S. Lane, K.S. Lilley, B.S. Phinney, et al., The Association of Biomolecular Resource Facilities Proteomics Research Group 2006 study: relative protein quantitation, Mol. Cell. Proteomics 6 (2007) 1291–1298.
[30] Y. Yasui, M. Pepe, M.L. Thompson, B.L. Adam, G.L. Wright, Jr., Y. Qu, et al., A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection, Biostatistics 4 (2003) 449–463.
[31] R. Tibshirani, T. Hastie, B. Narasimhan, S. Soltys, G. Shi, A. Koong, et al., Sample classification from protein mass spectrometry, by 'peak probability contrasts', Bioinformatics 20 (2004) 3034–3044.

[32] T.A. Addona, S.E. Abbatiello, B. Schilling, S.J. Skates, D.R. Mani, D.M. Bunk, et al., Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma, Nat. Biotechnol. 27 (2009) 633–641.
[33] L. Anderson, C.L. Hunter, Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins, Mol. Cell. Proteomics 5 (2006) 573–588.
[34] S. Santambrogio, A. Cattaneo, I. Bernascone, T. Schwend, L. Jovine, A. Bachi, et al., Urinary uromodulin carries an intact ZP domain generated by a conserved C-terminal proteolytic cleavage, Biochem. Biophys. Res. Commun. 370 (2008) 410–413.
[35] X. Peng, C.L. Wood, E.M. Blalock, K.C. Chen, P.W. Landfield, A.J. Stromberg, Statistical implications of pooling RNA samples for microarray experiments, BMC Bioinform. 4 (2003) 26.
[36] C. Kendziorski, R.A. Irizarry, K.S. Chen, J.D. Haag, M.N. Gould, On the utility of pooling biological samples in microarray experiments, Proc. Natl. Acad. Sci. USA 102 (2005) 4252–4257.
[37] C.M. Kendziorski, Y. Zhang, H. Lan, A.D. Attie, The efficiency of pooling mRNA in microarray experiments, Biostatistics 4 (2003) 465–477.
[38] R.A. Jolly, K.M. Goldstein, T. Wei, H. Gao, P. Chen, S. Huang, et al., Pooling samples within microarray studies: a comparative analysis of rat liver transcription response to prototypical toxicants, Physiol. Genomics 22 (2005) 346–355.
[39] S.T. Sadiq, D. Agranoff, Pooling serum samples may lead to loss of potential biomarkers in SELDI-ToF MS proteomic profiling, Proteome Sci. 6 (2008) 16.
[40] T.M. Annesley, Ion suppression in mass spectrometry, Clin. Chem. 49 (2003) 1041–1044.
[41] R. Tibshirani, T. Hastie, B. Narasimhan, G. Chu, Diagnosis of multiple cancer types by shrunken centroids of gene expression, Proc. Natl. Acad. Sci. USA 99 (2002) 6567–6572.
[42] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, JRSSB 57 (1995) 289–300.
[43] B. Efron, R. Tibshirani, J. Storey, V. Tusher, Empirical bayes analysis of microarray experiment, J. Am. Stat. Assoc. 96 (2001) 1151–1160.
[44] X.B. Ling, H. Cohen, J. Jin, I. Lau, J. Schilling, FDR made easy in differential feature discovery and correlation analyses, Bioinformatics 25 (2009) 1461–1462.
[45] R.C. Holman, B.J. Stoll, A.T. Curns, K.L. Yorita, C.A. Steiner, L.B. Schonberger, Necrotising enterocolitis hospitalisations among neonates in the United States, Paediatr. Perinat. Epidemiol. 20 (2006) 498–506.
[46] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clin. Chem. 39 (1993) 561–577.
[47] T. Sing, O. Sander, N. Beerenwinkel, T. Lengauer, ROCR: visualizing classifier performance in R, Bioinformatics 21 (2005) 3940–3941.
[48] C.A. Wallace, J.E. Levinson, Juvenile rheumatoid arthritis: outcome and treatment for the 1990s, Rheum. Dis. Clin. N. Am. 17 (1991) 891–905.
[49] J. Cassidy, R. Petty, Textbook of Pediatric Rheumatology, 4 ed., W.B. Saunders Company, Philadelphia, 2001.
[50] Y. Kimura, P. Pinho, G. Walco, G. Higgins, D. Hummell, I. Szer, et al., Etanercept treatment in patients with refractory systemic onset juvenile rheumatoid arthritis, J. Rheumatol. 32 (2005) 935–942.
[51] G. Horneff, H. Schmeling, T. Biedermann, I. Foeldvari, G. Ganser, H.J. Girschick, et al., The German etanercept registry for treatment of juvenile idiopathic arthritis, Ann. Rheum. Dis. 63 (2004) 1638–1644.

[52] T. Lequerre, P. Quartier, D. Rosellini, F. Alaoui, M. De Bandt, O. Mejjad, et al., Interleukin-1 receptor antagonist (anakinra) treatment in patients with systemic-onset juvenile idiopathic arthritis or adult onset Still disease: preliminary experience in France, Ann. Rheum. Dis. 67 (2008) 302–308.
[53] M. Gattorno, L. Chicha, A. Gregorio, F. Ferlito, F. Rossi, D. Jarrossay, et al., Distinct expression pattern of IFN-alpha and TNF-alpha in juvenile idiopathic arthritis synovial tissue, Rheumatology (Oxford) 46 (2007) 657–665.

This page intentionally left blank