9-1-2009

Urinary Protein Profiles in a Rat Model for Diabetic Complications

Daniela M. Schlatzer
Case Western Reserve University, daniela.schlatzer@case.edu

Jean-Eudes Dazard
Case Western Reserve University, jean-eudes.dazard@case.edu

Rob M. Ewing
Case Western Reserve University

Serguei Ilchenko
Case Western Reserve University

Mark R. Chance
Case Western Reserve University, mark.chance@case.edu

Author(s) ORCID Identifier:

 Mark R. Chance

Recommended Citation

Urinary Protein Profiles in a Rat Model for Diabetic Complications*⑤

Daniela M. Schlatzer‡, Jean-Eudes Dazard‡, Moyez Dharsee§, Rob M. Ewing‡§, Serguei Ilchenko‡, Ian Stewart§, George Christ¶, and Mark R. Chance‡‖

Diabetes mellitus is estimated to affect ~24 million people in the United States and more than 150 million people worldwide. There are numerous end organ complications of diabetes, the onset of which can be delayed by early diagnosis and treatment. Although assays for diabetes are well founded, tests for its complications lack sufficient specificity and sensitivity to adequately guide these treatment options. In our study, we employed a streptozotocin-induced rat model of diabetes to determine changes in urinary protein profiles that occur during the initial response to the attendant hyperglycemia (*e.g.* the first two months) with the goal of developing a reliable and reproducible method of analyzing multiple urine samples as well as providing clues to early markers of disease progression. After filtration and buffer exchange, urinary proteins were digested with a specific protease, and the relative amounts of several thousand peptides were compared across rat urine samples representing various times after administration of drug or sham control. Extensive data analysis, including imputation of missing values and normalization of all data was followed by ANOVA analysis to discover peptides that were significantly changing as a function of time, treatment and interaction of the two variables. The data demonstrated significant differences in protein abundance in urine before observable pathophysiological changes occur in this animal model and as function of the measured variables. These included decreases in relative abundance of major urinary protein precursor and increases in pro-alpha collagen, the expression of which is known to be regulated by circulating levels of insulin and/or glucose. Peptides from these proteins represent potential biomarkers, which can be used to stage urogenital complications from diabetes. The expression changes of a pro-alpha 1 collagen peptide was also confirmed via selected reaction monitoring. *Molecular & Cellular Proteomics 8: 2145–2158, 2009.*

Diabetic nephropathy (DNP)[1] accounts for ~44% of new cases of end stage renal disease (ESRD) (1). This high morbidity is the result of the impact of a growing population and longer life expectancy. With an increase in the prevalence of DM and a corresponding reduction in the mortality associated with both type 1 and type 2 DM, patients are living longer and are therefore at higher risk to develop complications such as nephropathy (2). Moreover, Type 1 DM patients who progress to ESRD have a substantial risk of mortality with estimated annual health care costs in the United States to be approximately $1.9 billion (3, 4). Two key therapies for the prevention and management of ESRD are aggressive glycemic control and blood pressure regulation (5, 6). Early intervention is essential in reducing the severity and course of this complication (6), and changes in urine biomarkers have historically been used to diagnose and monitor disease progression. In addition, urine represents a desirable matrix in which to detect biomarkers of nephropathy as urinary protein excretion profiles are reflective of functional changes within the kidney, such as glomerular filtration rate. Clinical determinations of urinary total protein and urinary albumin excretion are commonly used measurements to monitor and/or determine the onset of diabetic nephropathy. Unfortunately, these measurements often lead to improper diagnoses for at risk DM patients (7, 8). Therefore, new prognostic indicators are required to accurately target these patients for therapeutic intervention earlier in the course of the disease as well as identify patients who are unlikely to progress, as therapy may be of little or no benefit to them.

Utilizing experimental models to study the pathophysiological changes that occur as function of disease progression has provided an approach for biomarker discovery. In diabetes, animal models have been widely used in the investigation of the progression of diabetes complications such as nephropathy. Research conducted on the association between hyperglycemia and microvascular disease in diabetes as well as the study of the effect of extracellular matrix protein ex-

---

[1] The abbreviations used are: DNP, diabetic nephropathy; ESRD, end stage renal disease; DM, diabetes mellitus; STZ, streptozotocin; Diffset, differential set; ANOVA, analysis of variance; FC, -fold change; PCA, principle component analysis; SRM, single reaction monitoring; LC/MS, liquid chromatography/mass spectrometry; MS/MS, mass spectrometry/mass spectrometry; ^{18}O, $^{18}Oxygen$; pFDR, expected positive false discovery rate; QC, quality control; MIDD, monoclonal immunoglobulin deposition disease.

---

pression on changes in morphology in the diabetic kidney are two such examples (9). In addition, these models have assisted in developing appropriate clinical trials for the prevention and treatment of these complications. One such example is the use of anti-hypertensive treatment regimes in genetically hypertensive rats; these have examined whether early intervention may be renoprotective and therefore delay or prevent the onset of diabetic nephropathy (10–12).

STZ-induced hyperglycemia in rodents is the most extensively studied model of diabetic nephropathy and associated complications (9, 13). Hyperglycemia occurs in this model because of the toxin's destruction of pancreatic Beta-islet cells, which are essential to the production of insulin. STZ-induced hyperglycemia is associated with reliable and consistent structural and functional deficits in specific urogenital organ function (i.e. kidney and bladder). Increased glomerular filtration and hypertrophy, as well as increased urgency and morphology changes, are structural and functional abnormalities that have been observed in the kidney and bladder, respectively, in both in humans and the STZ rat model (14–19).

Currently, there are two primary methods used to monitor disease progression in diabetes. The measurement of urinary albumin excretion rates and total protein concentration are routinely used to monitor disease progression as they reflect structural and functional changes in the kidney. Measurements of albumin by immunochemical assays and size exclusion high performance liquid chromatography are routinely employed (7). However, urine consists of a multitude of proteins, many of which are also reflective of pathophysiological changes because of DM urogenital complications (20–23). Proteomics provides a powerful approach for the detection of urinary protein changes as a result of disease, and multiple proteomic techniques are available in large scale protein profiling to discover new biomarkers (24–26). To date, proteomic strategies for biomarker discovery in urine have primarily included top-down approaches, for example two-dimensional gel electrophoresis coupled with mass spectrometry and/or surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS) analyses (27–30). In addition a number of studies using capillary electrophoresis mass spectrometry, which have a number of advantages for analysis of urine, have been successfully carried out (31–33). While these approaches can easily detect and quantify a variety of proteinaceous species including isoforms, post-translational modifications, or degradation products, other methods could be used to expand the number of proteins that are both quantified and identified providing an expanded set of biological targets to understand the complications of disease and its progression. Recent advances in both chromatography and mass spectrometry have enabled bottom-up approaches that identify and quantify at the peptide level (34–36). One advantage of bottom-up proteomics is increased overall proteome coverage. Moreover, most bot-

tom-up methods provide both qualitative and quantitative data in a single run, and quantifying at the peptide level leads directly into a bottom-up confirmation/validation analysis thereby avoiding the peptide selection step in this procedure. Approaches to bottom-up proteomics include specific peptide labeling or label-free analysis. Specific labeling approaches such as isobaric tag for relative and absolute quantification (iTRAQ) and ^{18}O employ differential stable isotope labeling strategies that create specific mass tags for different samples, which are mixed and then identified and quantified using mass spectrometry (37). The utility of these techniques is that they accommodate a wide range of pre-fractionation strategies thereby improving proteome coverage.

The label-free approach capitalizes on the highly reproducible chromatography and high mass accuracy available in current LC/MS systems. This method observes all detectable peptides and if interrogated by MS/MS their corresponding fragment ions. This approach quantifies a peptide by its intensity and groups each peptide across individual samples based on its accurate mass and retention time (38, 39). These intensities associated with specific mass and retention time values are organized into peptide array tables that may be further processed using statistical techniques that accommodate high-dimensional data. As with other bottom-up approaches, this method is also amenable to pre-fractionation strategies, but unlike labeled approaches, the removal of chemical or metabolic labeling steps simplifies the overall approach.

Here we use a comparative label-free LC/MS/MS approach to identify and rank candidate biomarkers of urogenital complications from an STZ rat model of diabetes. We describe further technical validation of our approach by confirming the changes observed with the putative biomarker, pro-alpha (2) with an alternative method: selected reaction monitoring (SRM).

EXPERIMENTAL PROCEDURES
STZ-induced Hyperglycemia

Hyperglycemia was induced in male F-344 rats by a single intraperitoneal injection of STZ (35 mg/kg). All procedures were approved by the Animal Institute Care and Use Committee at Wake Forest University. STZ was dissolved in citrate buffer (1:1 mixture of 0.1 M citric acid and 0.2 M Na_2HPO_4). Rats were fasted for a period of 24 h prior to STZ injection and presented with a 10% dextrose solution for a 24-hour period immediately following injection. The rat becomes hyperglycemic and glycosuric within 24 h. The extent of diabetes was confirmed via serum glucose evaluation from tail vein puncture. Blood was withdrawn from the tail vein whereas the animal was placed in a custom made restraining device to expose the tail. Blood glucose levels were determined with a glucometer (Ascensia Elite XL by Bayer system). Blood glucose levels were subsequently measured once a week for the first month and in the absence of signs of distress, monthly thereafter. Biochemical and electrophysiological deficits associated with the induced hyperglycemia are well characterized and generally occur early (within weeks of the onset of glucose elevation) (13, 40). Control and DM urine was collected at 3 day, 1 month, and 2 months time points. To

collect urine, rats were placed in a metabolic cage and urine was collected over a 6 h time period. The samples were then frozen in a −80 °C freezer until proteomic analysis.

Experimental Design

Factorial Arrangement—The categorical factors under study consist of a treatment factor (two levels) crossed with a time factor (three levels). The two levels of the treatment group were the drug-induced disease group (T), while the control group (C) was the one that received a sham injection. The three levels of the time factor were the time points of 3 days, 1 month, and 2 months (3 d, 1 m, 2 m) from the start of the treatment (day 0). The experimental groups consisted of the 2×3 possible combinations of levels of the factors, further denoted 'C.3d′, 'T.3d′, 'C.1m', 'T.1m', 'C.2m', 'T.2m'. This is a *two-way factorial arrangement*, which allows evaluation of the two main effects of the factors and their interactions.

Design Layout, Experimental Units, Pairing, Replications, and Pooling—Supplemental data Table I highlights the experimental units under study ($n = 16$). Experimental units were 6 and 10 rats, respectively, in control and treated groups at all time points; and of 4, 6, and 6 rats at 3 days, 1 month, and 2 months to both control and treated. As rats were sacrificed at all time points, all experimental groups were independent (unpaired). In addition, treatments were assigned completely at random to independent experimental units. Therefore, this is an arrangement of treatments within an unbalanced and complete randomized design. Technical replicates were not performed, samples were not pooled, and no common reference sample was used.

Urine Sample Preparation

Four hundred microliters of each rat urine sample was concentrated to 100 μl using a Microcon 3,000 molecular weight cut off filter (Millipore, Bedford, MA). Each sample is subsequently buffer-exchanged three times using 300 μl of 50 mM Tris buffer to a final volume of ~100 μl, and protein concentrations were determined by 2D Quant kit as described by the manufacturer (GE Healthcare). Twenty micrograms of each sample was run on a one-dimensional SDS gel for quality control. Subsequent to one-dimensional SDS gel analysis, aliquots of each sample were adjusted to 40 μg in 57 μl in 50 mM Tris. Twenty microliters of 0.2% Rapigest (Waters, Milford, MA) and dithiothreitol to a final concentration of 5 mM were added. The samples were reduced at 80 °C for 15 min, cooled to room temperature, and alkylated with iodoacetamide at a final concentration of 10 mM for 30 minutes. Proteolytic digestion was performed with Lys-C (Wako Chemicals, Richmond, VA) with final enzyme to protein ratio of 1:6 (w/w), and the final digestion volume was adjusted to 200 μl with 13 μl 1.5 M Tris. The samples were then incubated for 18 h at 37 °C. A primary stock solution of MassPREP™ Digestion Standard Mix (Waters) peptides containing yeast enolase was prepared to a concentration of 1 pmol/μl each. The standard peptide mixture was added such that the final concentration was 0.4 pmol/μl (final volume of 150 μl).

Label-free Expression

Liquid Chromatography and Mass Spectrometry—Sixty nanograms of each sample were analyzed by LC/MS/MS and the order of sample injections randomized over all samples. Separation of peptides via capillary liquid chromatography was performed using a Dionex Ultimate 3000 capillary LC system (Dionex Sunnyvale, CA). Mobile phase A (aqueous) contained 0.1% formic acid in 5% acetonitrile, and mobile phase B (organic) contained 0.1% formic acid in 85% acetonitrile. Samples were trapped and desalted on-line in mobile phase A at 10 μl/min for 10 min using a Dionex PepMap 100, (300

μm x 5 mm). The sample was subsequently loaded onto a Dionex C18 PepMap (75 μm x 15 cm) reversed phase column with 5% mobile phase B. Separation was obtained by employing a gradient of 6% to 28% mobile B at 0.300 μl/min over 100 min. The column was washed at 99% mobile phase B for 10 min, followed by a re-equilibration at 100% mobile phase A for 17 min. Mass spectrometry analyses of samples were performed using a hybrid linear ion trap Fourier-transformation cyclotron resonance mass spectrometer (FT-LTQ; Thermo, Waltham, MA). Positive mode electrospray was conducted using a nanospray source, and the mass spectrometer(s) was operated at a resolution of 25,000. Quantitative and qualitative data were acquired using alternating full MS scan and MS/MS scans in normal mode. Survey data were acquired from *m/z* of 400–1600 and up to three precursors, based on intensity, were interrogated by MS/MS per switch. Two micro scans were acquired for every precursor interrogated, and MS/MS was acquired as centroid data. The FT and LTQ were mass calibrated immediately before the analysis using the instrument protocol. Raw LC/MS/MS data was processed via Proteomarker software (Infochromics, Toronto, Canada).

Data Processing–Qualitative—The raw data for each run were first extracted to provide MS/MS peak lists for identification and intensity-based profile peak lists for quantification. The MS/MS peak lists were subsequently searched by Mascot version 2.2.0 (Matrix Science, London, UK). The database used was a compilation of both the rat and mouse International Protein Index (IPI) (July 2007; 97811) plus the sequence for a dietary protein from rat chow, GY5 which was soybean protein we expected to be recovered in urine. Search settings were as follows: no enzyme specificity, mass accuracy window for precursor ion, 10 ppm; mass accuracy window for fragment ions, 0.8 Daltons; variable modification, including only carbamidomethylation of cysteines and oxidation of methionine. The criteria for peptide identification were a mass accuracy of ≤10 ppm and an expectation value of $p \leq 0.05$. Proteins that had more than two peptides matching the above criteria were considered confirmed assignments whereas proteins identified with one peptide regardless of the Mascot score were highlighted as tentative assignments.

Data Processing–Quantitative—Automated differential quantification of peptides in a set of samples was accomplished with Proteomarker. For each profile MS spectrum of an LC/MS/MS run, isotopic envelopes were resolved, charge states determined, and monoisotopic masses assigned to generate a monoisotopic peak list. This peak list was further consolidated by grouping masses observed at different charge states and appearing in contiguous elution time points, resulting in a set of chromatographic peaks that corresponded to distinct putative peptides. Mascot protein identity data were integrated at this point by matching peptide sequences to chromatographic peaks on the basis of mass and retention time. Chromatographic peaks were aligned across samples using the attributes of monoisotopic mass, retention time, charge state distribution, and peptide sequence (when available). To effectively align peptides, Proteomarker corrects for retention time drift by warping elution times across samples using endogenous peaks. Alignment between sample datasets resulted in a peptide abundance matrix in which each row (or differential set) represented a putative peptide with sample-specific attributes such as observed mass, starting and ending chromatographic retention times, peptide sequence (if assigned), and calculated chromatographic area used as a measure of intensity for the given peptide.

Data Quality Control–Prefiltering—Subsequent to raw data acquisition and processing, data QC and statistical analysis were performed. Fig. 1 describes the protocol for data QC and statistical analysis that was conducted for this study. Two major peptide pre-filtering steps were performed to remove very poor quality peptide identifications as well as those differential sets (diffset) in the
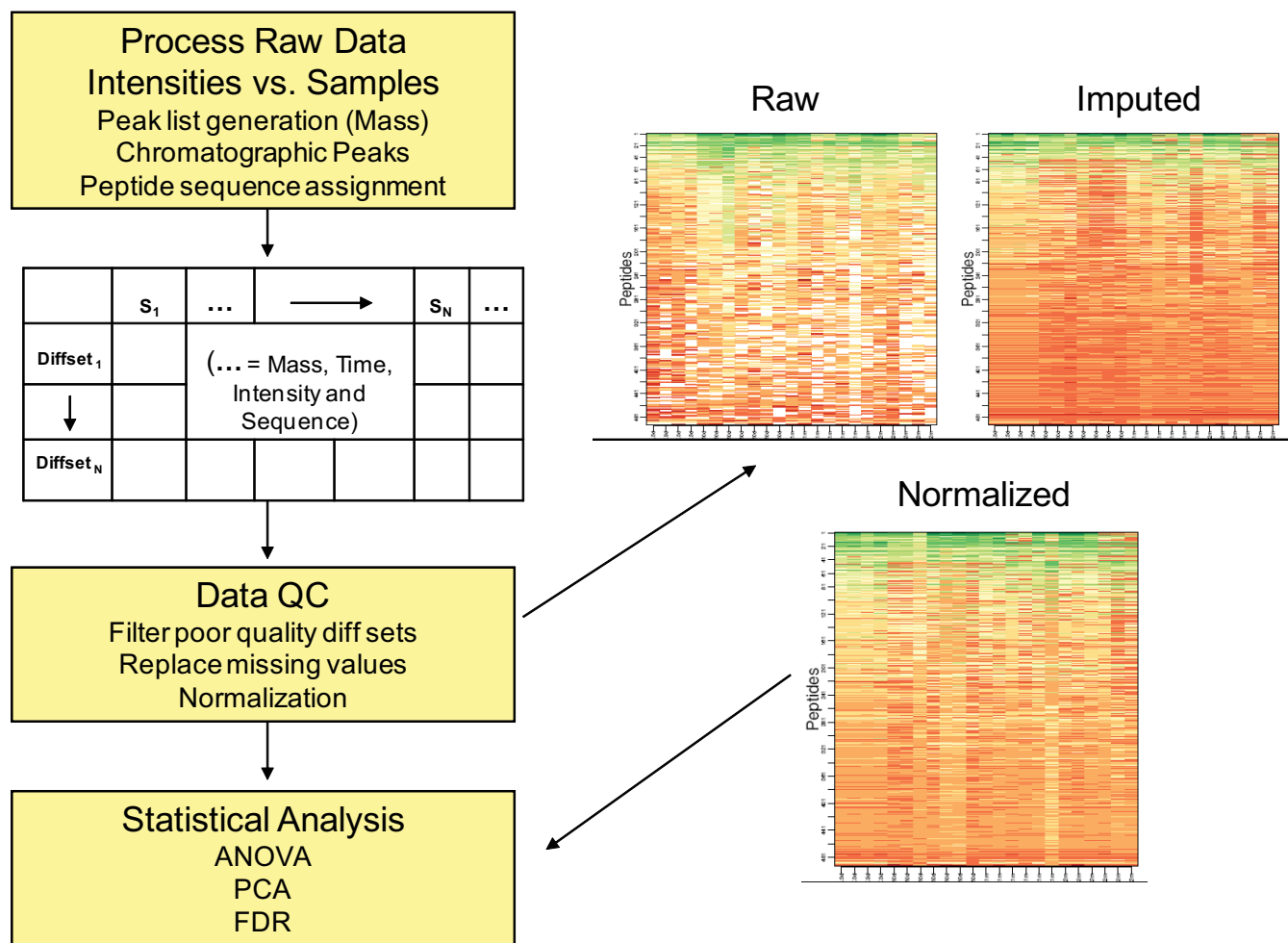
FIG. 1. **Workflow for quality control and statistical analysis of label-free expression data.**

abundance matrix that did not receive a qualitative identification. A diffset for analysis was rejected if (i) a peptide sequence was not assigned or (ii) the consensus peptide sequence score was below the 81st percentile (<9.96), which represents the lowest mode of distribution of Mascot scores. To further reduce the dimensionality of the data, we carried out peptide specific intensity summaries for diffsets that contained peptides of identical sequence. When several diffsets could be summarized, the diffset with the least number of missing values was retained with all its annotations including retention time, m/z ratio, sequence, score, protein annotation; intensity values from other filtered diffsets were transferred into the retained diffset. A filtering step was added to retain those diffsets; (i) for which the number of *missing* intensity values by experimental group (3 day, 1 month, or 2 months) was less than 50% in *at least one* experimental group, and (ii) for which a certain threshold of missing intensity values per diffset was not exceeded. Given these criteria, the number of missing events per diffset was bounded to 15 as the control 3 day experimental group contained only one sample (see under "Results").

Data Quality Control–Probability Model for Missing Value Imputation—Missing values in LC/MS data arise because of imperfect detection and alignment of peak intensities or by true absence of expression. The probability model and iterative algorithm described in Wang *et al.* (41) was used to account for the *missingness mechanism* at play and its extent for these data. This model substitutes a missing

measurement of intensity with its expected value of the true intensity given that it is unobservable. Estimation of the imputation parameters was done in a simulation type of study in order to minimize the percentage of remaining missing values.

Data Quality Control–Normalization, Variance Stabilization, and Normality—The purpose of normalization is to identify and remove sources of systematic variation because of experimental artifacts in the measured intensities. In addition, it is important to ensure that the imputed intensities are corrected for variance stabilization and normality. Various transformations were applied on the data features to ensure that the above assumptions were met. We used a common normalization method known as natural cubic smoothing splines as described in Workman *et al.* (42). This normalization procedure uses sample quantiles from the data features to fit natural cubic smoothing B-splines. The splines are then used as signal-dependent normalization functions on the original data.

Statistical Analysis–Grouping Effect, Linear Model, and Empirical Bayes Estimators for Statistical Inference—Potential groups and outliers among the samples were checked by a Principal Component Analysis (PCA) (43). Subsequent to PCA, we fitted the same statistical model individually to each univariate response variable (single peptide expression). A standard analysis method in modeling of high dimensional data is to fit the same statistical model (usually linear) individually to each outcome variable (diffset or peptide) and test for the contrast or effect of interest using the hypothesis testing framework.

A drawback of this univariate approach is that the correlation structure or dependence between the variables is ignored. However, some compensating possibilities exist by borrowing information across similar variables, resulting in more stable variance estimates, which in turn assist in inference about each variable individually. Thanks to the parallel nature of the high-throughput data, this can easily be done by application of empirical Bayes methods and estimators derived from them, resulting in greater statistical power (moderated F-, t-, and B statistics) (44–48). The B statistic is the empirical Bayes $\log_2$ of the posterior odds that the peptide is differentially expressed and represents a measure of statistical significance. In addition, a number of authors have noted in gene expression studies that such approach of "borrowing strength" across variables is more reliable and does not reduce the power available to detect changes in expression for individual variables (49). In addition, the posterior odds statistics have proven to be a useful means of ranking variables in terms of evidence for differential expression, especially when the sample size are small, as is the case here (44–53).

One common application of this approach is in the linear model setting (48). When multiple sources of variation and correlation are at play, a linear *mixed* effects model of analysis of variance is usually an appropriate and powerful approach. To fit a mixed effects model, where LC/MS runs or subjects random effects would enter into the model would usually require having technical replicates, which we do not have. Nevertheless, because in the present study the biological samples were all independent, and the LC/MS runs were randomized, we technically could ignore the corresponding random effect terms and use a simple *fixed* effects model of analysis of variance (two-way ANOVA). The fixed effects to be estimated are the treatment effect (T_l), the time effect (D_k), and an interaction effect (DT_{kl}). Letting y be the intensity signal on the original scale for the *j*-th peptide, and letting $z_{jkl} = t(y_{jkl})$ be the outcome on the transformed scale, where $t(y_{jkl})$ is the appropriate transformation described above (54), a linear fixed ANOVA model for *each* individual peptide *j* is fitted as follows:

$$z_{jkl} = t(y_{jkl}) = \mu_j + D_{jk} + T_{jl} + DT_{jkl} + \varepsilon_{jkl} \qquad \text{(Eq. 1)}$$

where μ_j represents the average signal intensity for that peptide across all factors and observations, and the error term $\varepsilon_{jkl} \sim N(0, \Sigma)$ is assumed to be normally distributed with mean 0 and some (unknown) variance component.

In this experimental design, contrasts were built for each of the fixed effects of interest, coefficients were estimated, and variables were ranked in order of evidence of differential expression. Corresponding p values were adjusted for multiple testing using a recent extension of the standard Benjamini-Hochberg procedure, which controls the expected False Discovery Rate (FDR) under the usual assumption of general variable dependence (55). This error rate, called the *positive* FDR (denoted pFDR) (56–58), results in a procedure much less conservative than the FDR and therefore typically more powerful and more appropriate for large datasets. In our study, the threshold was set arbitrarily to pFDR = 0.05. Whenever available, implementations and algorithms of our methods were from the freely available consortium CRAN (Comprehensive R Archive Network). All other R codes written in our group can be provided upon request. For image analysis, exploratory data analysis, normalization procedures, and quality assessment, we used the packages "vsn" (59), "affy" (42, 60), "PreprocessCore" (61). In addition, for linear modeling and supervised inferences, we used the package "limma" (48). Finally, for the control of the positive FDR, we used the package "qvalue" (53).

Confirmation of Peptide GEPGSVGAQGPPGPSGEEGK Selected Reaction Monitoring Mass Spectrometry—The pro-alpha (2) 1 collagen peptide GEPGSVGAQGPPGPSGEEGK was synthesized (Sigma-Aldrich, St. Louis, MO) and stock solutions were prepared in 10% acetonitrile at 1 nmol/μl. A 1 pmol/μl stock was infused into the mass

spectrometer to determine ideal analysis conditions for the peptide as well as to select the best fragment ion which to monitor in selected reaction monitoring. Based on the MS/MS of this peptide, the y^{18} fragment was chosen as it was both the most stable and intense fragment in the spectrum (see under "Results"). The stock solution was diluted in 10% acentonitrile/0.1% formic acid to various concentrations between 50 amol/μl and 25 fmol/μl. Digested samples were prepared such that their total concentration was 5 ng/μl. Separation of target peptide via capillary liquid chromatography was performed using a Dionex Ultimate 3000 nanoscale LC system. Mobile phase A (aqueous) contained 0.1% formic acid in 5% acetonitrile, and mobile phase B (organic) contained 0.1% formic acid in 85% acetonitrile. Samples were trapped and desalted on-line in mobile phase A at 10 μl/min for 10 min using a Dionex PepMap 100 (300 μm × 5 mm) column. Twenty microliters of standard as well as individual urine samples were subsequently loaded onto a Dionex C18 PepMap 75 μm × 15 cm reversed phase column with 5% mobile phase B. Separation was obtained by employing a gradient of 6% to 28% mobile B at 0.300 μl/min over 30 min. The column was washed at 99% mobile phase B for 10 min, followed by a re-equilibration at 100% A for 10 min. Mass spectrometry analyses of samples were performed using an LTQ mass spectrometer (Thermo). Positive mode electrospray was conducted using a nanoflow sprayer, and quantitative data were acquired via selected reaction monitoring mode. The transition that was monitored was a normal scan of 897.40 m/z to 804.60 m/z, which corresponded to the 2+ charge state of the intact peptide and its y^{18} fragment. On the LTQ, the product ion was isolated at 1.6 m/z width, the ion injection time was limited to 150 ms per microscan with one microscan, an AGC setting for the MS/MS of 1×10^4, and the data were acquired in centroid mode. Raw chromatograms were subsequently processed and analyzed using Xcalibur Quan View software version 2.0.5 (Thermo). Masses corresponding to the transition state of 897.40–804.60 m/z, the retention time of which fell within a window ± 2.5 min of the target retention time, were extracted, and the peak area was integrated by Quan View. The peak areas generated by the software were also manually inspected to ensure proper integration and adjusted when necessary. Following peak area integration, the software generated concentration curves for the standard samples via linear regression and the endogenous peptide from rat urine sample were plotted against this curve to determine concentration per 100 ng of total protein.

RESULTS

Label-free Expression Analysis—The sample preparation protocol provided sufficient protein concentrations from 400 μl of urine for label-free comparative analysis. These concentrations ranged from 1 μg/μl to 9.5 μg/μl. These variations are attributed to the wide range of sample collection volumes (1–15 ml). Reproducible protein patterns via one-dimensional SDS-PAGE were observed across time and treatment points (data not shown). These samples were subsequently analyzed by LC/MS/MS. The optimized LC/MS/MS analysis described under the "Experimental Procedures" section provided excellent chromatographic reproducibility across injections. Fig. 2 highlights this reproducibility with retention times on average deviating on the order of ± 1 min.

Data Quality Control–Prefiltering—As described under the "Experimental Procedures" section, quality control filters were applied to the dataset before statistical analysis. The first filter removed diffsets for which a sequence was not obtained in the label-free LC/MS/MS analysis. A second qualitative filter
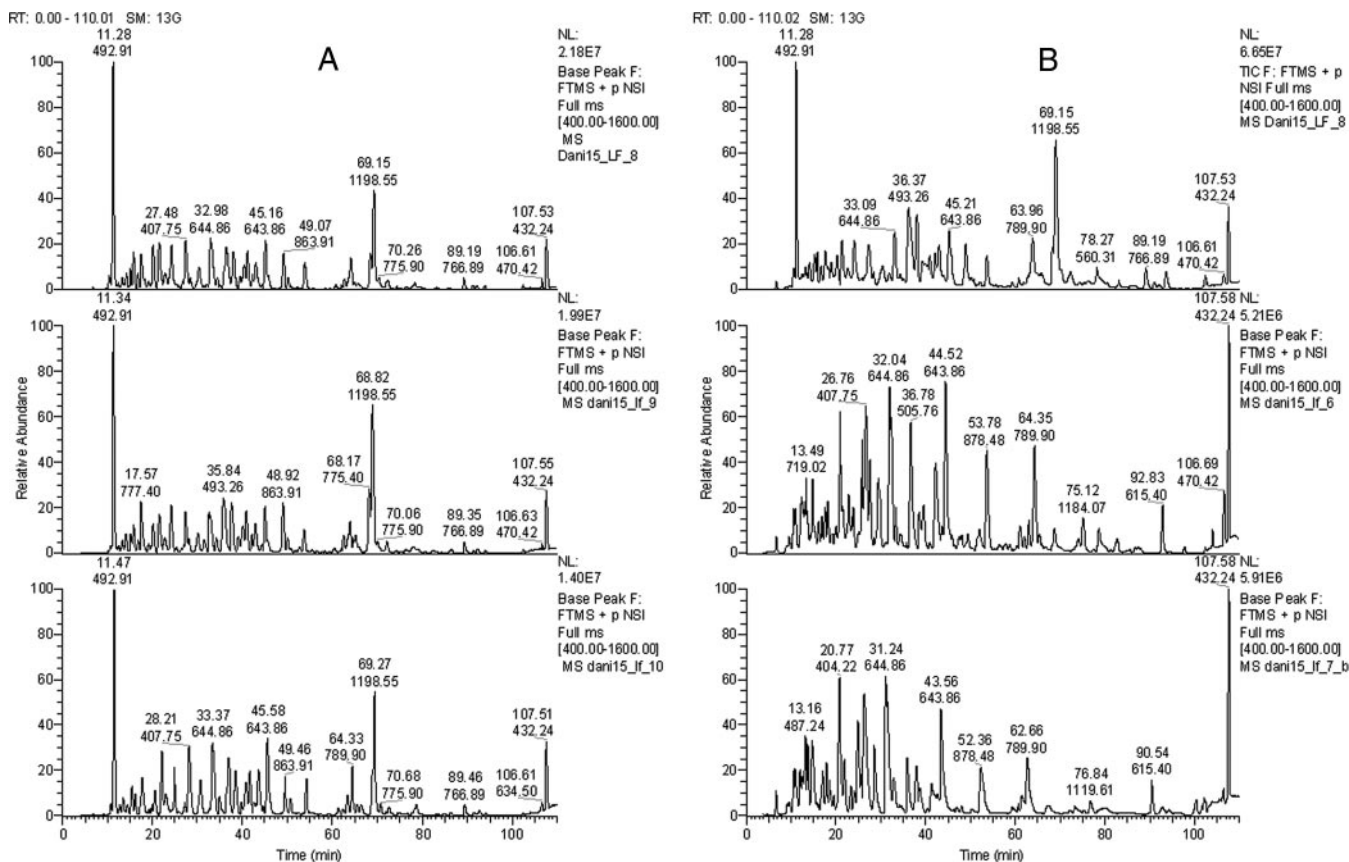
FIG. 2. **Base peak chromatograms for label-free expression analysis.** Groups presented are 2 months Control (*A*) and 2 months STZ-DM (*B*). Excellent chromatographic reproducibility is observed (± 2 min). Consistent peptide profiles in terms of relative intensities are also observed for both groups.
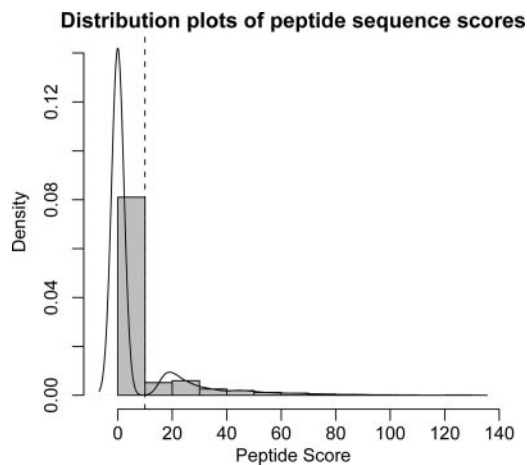

FIG. 3. **Scoring distribution plot for filtering of peptide scores.** Mascot peptide score (*Peptide Score*) *versus* frequency of score (*Density*) is plotted. The 81 percentile cutoff, which was used to filter peptides is highlighted by *dotted line*.

was applied using the peptide identification probability score. These filters conserved those diffsets that were assigned a peptide sequence and a Mascot score with a good likelihood of validation in an independent assay. Fig. 3 highlights the

threshold for Mascot score that was chosen. This threshold separates the lower mode in the distribution of scores from the higher mode. Peptides with scores that fell below this cutoff were removed from the dataset. The prefiltering steps conserved 1931 diffsets, and these were progressed to additional quality control and summarization measures. We then summarized and collapsed those diffsets for which the same peptide sequence was identified. One hundred and forty eight diffsets, sharing identical sequences but separately clustered, were summarized. This left 1783 diffsets with unique and reliable peptide annotations and adjusted intensities that were carried forward for further analysis. At this stage, we removed diffsets that had more than 15 missing intensity values reducing the total number of diffsets to 1429.

Data Quality Control–Missing Values and Imputation—Despite these QC filtering steps, a significant number of missing values remained in the data. Analysis of this dataset showed that the frequency of a missing value was not dependent on sequence length or composition as no observable correlation was discovered when analyzing these variables (data not shown). On the other hand, Fig. 4 highlights an inverse correlation of the average peptide intensity to the number of missing values for a peptide. A consistent trend is observed in

which an increase in average peptide intensity yields fewer missing data events. This was anticipated as low intensity peptides ions are closer to the limit of detection and are difficult to quantify from background noise. This type of *missingness* is referred to as Missing Not At Random (MNAR). The total number of missing values for the 1429 diffsets was 42.3%. The imputation model described in the "Experimental Procedures" section reduced the missing data to 31.6% (~25% reduction). Because the remaining missing values typically represent undetectable true absent expression levels, these were substituted by the minimum of the imputed values. Subsequent to imputation, we carried out normalization procedure that was applied to the intensities of the entire dataset across all the samples, where each sample represented an independent LC/MS/MS run. Among a variety of transformation methods tested, the cubic smoothing splines normalization pro-

cedure achieved the best combination of normalization, variance stabilization, and normality across all samples (Fig. 5). The normalized data were then subjected to statistical analysis.

Principal Component Analysis, Statistical Inferences, and FDR Corrections—Potential groups and outliers among the samples were evaluated by a PCA. Based on the PCA, three groups stand out: the control and treatment group at 3 days (C.3d′, 'T.3d′ '); the two treatment groups at 1 and 2 months ('T.1m', 'T.2m'); the two control groups at 1 and 2 months ('C.1m', 'C.2m'). The treatment groups behave similarly at 1 month and 2 months with little overlap with the other control groups (Fig. 6). This indicates that there is probably no treatment effect at 3 days but there is a clear treatment effect with a potential interaction effect at 1 and 2 months time points. No outliers were observed.

Statistical inference of the dataset was performed using a fixed two-way ANOVA model. Diffsets were ranked by *p* value. Corrected *p* values, adjusted for multiple comparisons were reported with a positive $pFDR_{max}$ threshold of up to 5%. In this design, *treatment* and *time* main effects as well as their *interaction* effect can be evaluated. supplemental Tables 2, 3, and 4 highlight the peptides that were determined as having a significant change, including 434 peptides up- or down-regulated across all treatment comparisons, 82 up- or down-regulated in all time comparisons and 54 up- or down-regulated in all interaction comparisons. Note that a cutoff FDR correction of 5% was applied, meaning that no more than 21, 4, and 2 of these selected peptides, respectively, are expected false positives. Overall, the majority of significant peptides observed across comparisons changed with treatment. These represent peptides/proteins the expression of which is sensitive to STZ-induced hypoglycemia and/or drug treatment. Moreover, the 2 months treatment comparison yielded the largest number of significant changes within the treatment groups. The volcano plots in Fig. 7 highlight the 326 peptides observed as significant at 2 months post-STZ treatment in comparison to 19 at 1 month (see also supplemental Tables 5,
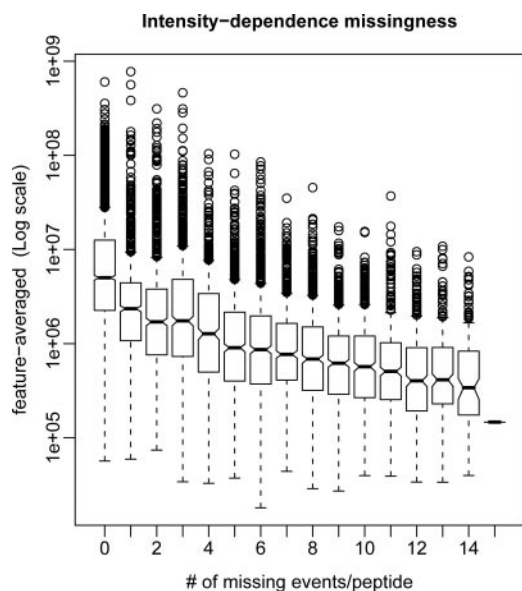

FIG. 4. **Number of missing events *versus* intensity for the diffset matrix.**

FIG. 5. **Box plot of raw sample intensities in log scale before and after cubic smoothing splines normalization by group.** The summary consists of the median, the upper and lower quartiles, the range, and individual extreme values. The *central box* in the plot represents the inter-quartile range, which is defined as the difference between the 75th percentile and 25th percentile, *i.e.* the upper and lower quartiles (the two "hinges"). The *line* in the *middle* of the box represents the median; a measure of central location of the data.
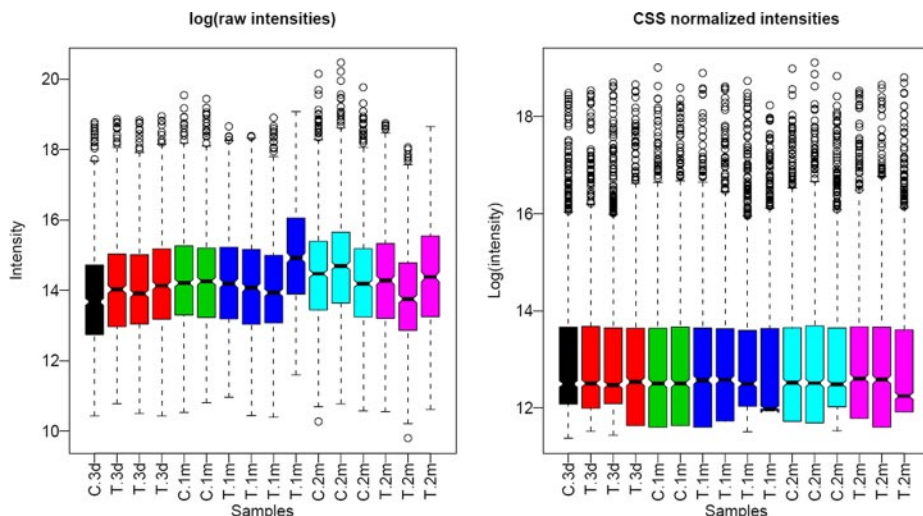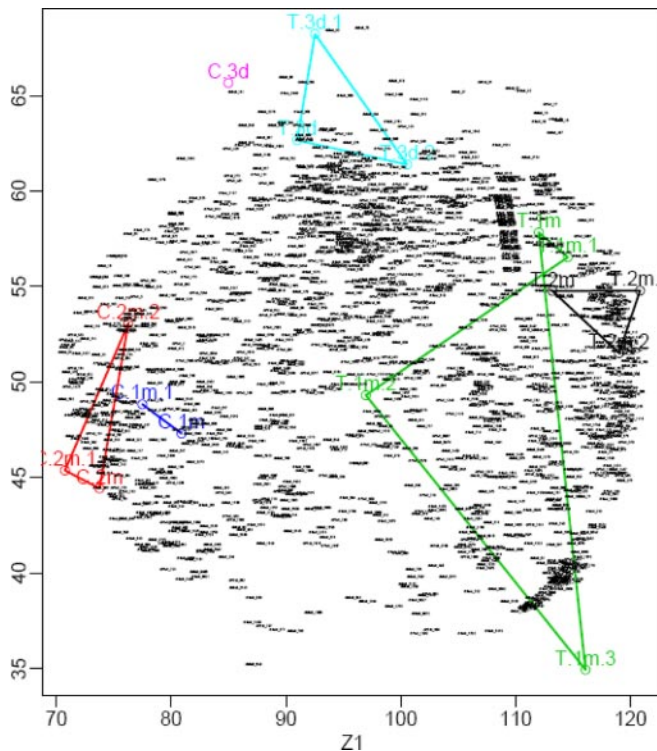
FIG. 6. **PCA plot for all samples analyzed in label-free analysis.** Variables (diffsets) contributing very little to the principle components have small loading values and are plotted around the center of the plot. Variables, which contribute most to distinguishing the sample groups, are plotted around the *borders* of the plot. Distinct separations were observed for the control and treatment groups at 3 days (C.3d′, 'T.3d′); the two treatment groups at 1 and 2 months ('T.1m, 'T.2m'); and the 2 control groups at 1 and 2 months ('C.1m', 'C.2m').

6, and 7). In addition, Table I highlights the proteins that were significant across all treatment comparisons, which met our criteria for confirmation of protein identification as described previously in the "Experimental Procedures" section. A total of 22 proteins were given a confirmed assignment with 5 of those up-regulated and 17 down-regulated across the treatment comparisons. Smaller numbers of peptides were significant with time and/or interaction in comparison to treatment. The 82 peptides that changed with time were related to age in this model as these rats were 12-weeks-old at the time of the study, and we anticipated differences in some protein levels because of maturation. The relative abundance of a few of these proteins also changed as a function of treatment and therefore they are identified as significant in the interaction analysis. The interaction comparison was consistent with observations made in the treatment effect as the majority of peptides identified as significant in the interaction effect were found at the 2 months comparison.

SRM of Pro-Alpha (2) 1 Collagen—A single peptide of rat pro-alpha (2) 1 collagen GEPGSVGAQGPPGPSGEEGK was identified in the label-free expression analysis. It was observed as a lower abundant peptide in the matrix with a tentative sequence assignment and had an average retention time of 14.5 min. Because of the biological interest and association with disease progression as well the existence of a homologous protein in humans, pro-alpha (2) 1 collagen was progressed to validation in the samples via SRM analysis. The peptide was synthetically generated and used for validation of the identification previously made as well as development of the SRM method. The synthetic peptide was infused into an LTQ mass spectrometer and MS/MS conditions optimized. The synthetic peptide fragmentation pattern was an exact match to the endogenous peptide MS/MS acquired in the label-free analysis thereby confirming the previous sequence assignment. In addition, a transition of 897.40 *m/z* to 804.60 *m/z* was chosen which corresponded to the +2 charge state of the y[18] fragment of GEPGSVGAQGPPGPSGEEGK (Fig. 8). This transition was selected for SRM as it was the most reproducible and intense fragment ion, comprising most of the ion signal for the MS/MS. The synthetic peptide was spiked into 0.1% formic acid at various concentrations (50 amol/μl, 125 amol/μl, 250 amol/μl, 500 amol/μl, 1 fmol/μl, and 2.5 fmol/μl) to estimate limit of detection. The limit of detection was defined as 125 amol/μl because below this level the data was not reliable. The curve was linear for this peptide over the concentration range selected with an *r* squared value of 0.996. Potential matrix suppression issues were evaluated by spiking the synthetic peptide standard into rat urine at a concentration of 1 fmol/μl and comparing the intensity response for this concentration to the intensity response of this peptide in standard matrix (0.1% formic). After determining limit of detection and concentration range, the samples were subsequently analyzed. Two curves bracketed the samples with three quality control standards consisting of a pooled 3 day DM sample interspersed in the analysis to ensure instrument reproducibility. The raw data was processed, and concentrations of target peptide were calculated for each sample per 100 μg of total urine protein. The targeted SRM for this peptide confirmed the increase in abundance observed in label-free analysis at the 2 months DM time point. In addition, Fig. 9 shows the targeted assay highlighted changes at the 1 month DM time point that were previously undetected. This is attributed to the low intensity level for this peptide in the label-free analysis, which may have appeared below the signal-to-noise threshold of the peak detection software.

Estimated Log -Fold Change for Label-free Expression Analysis—Supplemental Tables 2–9 contain the estimated log$_2$-fold change (M log-ratios) for individual peptides across effect or contrast of interest. The M log-ratios ($M = \log_2 E_1/E_2$) represent a log$_2$-FC between two or more experimental conditions in the case of a main effect and to a difference in log$_2$-FC in the case of an interaction effect. For example, the range of log$_2$-fold change observed for significant peptides in the 2 months treatment comparison were from 4.94 to −9.00. These -fold changes are to be considered an approximation of the true abundance because of the data
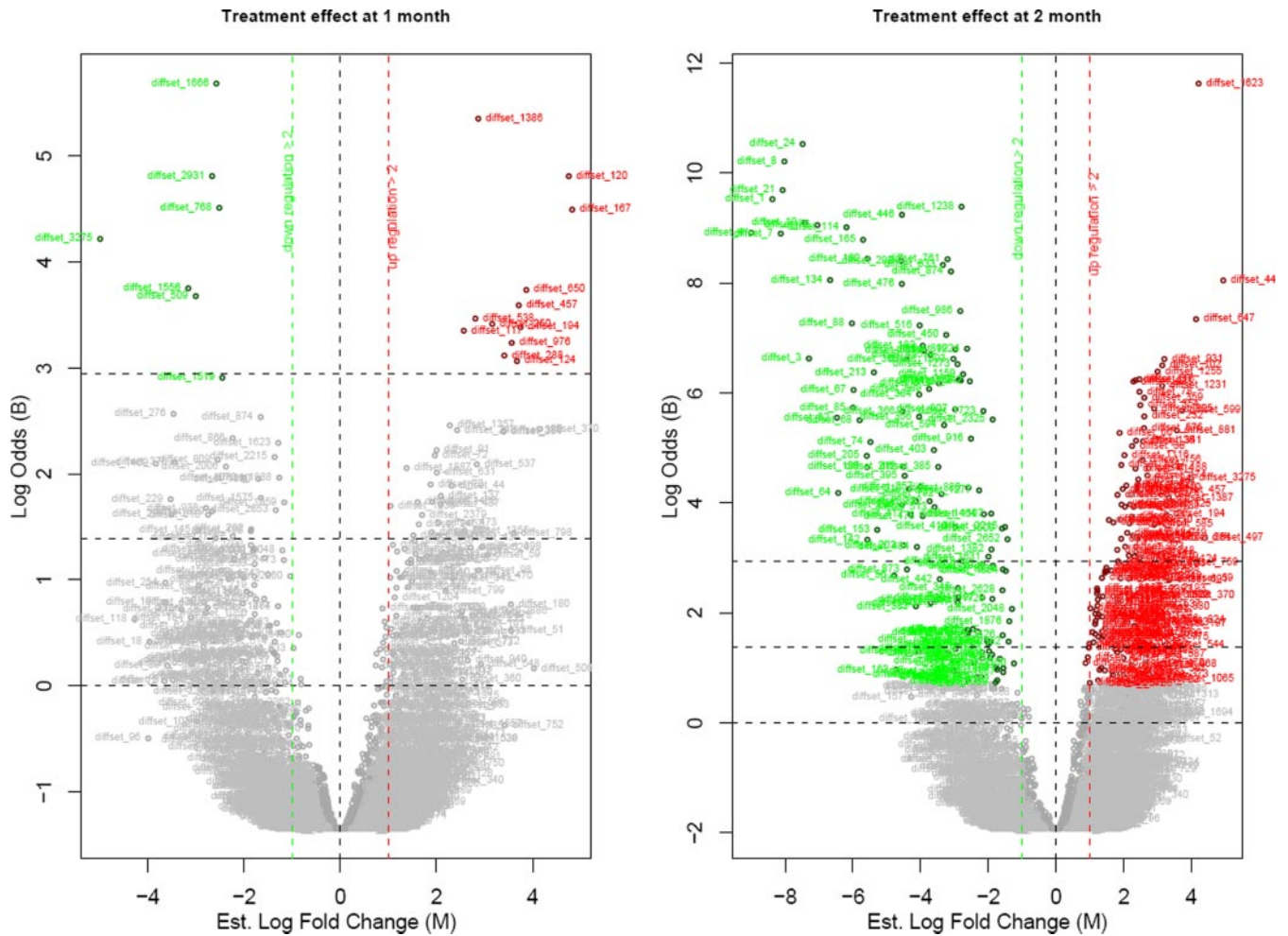
FIG. 7. **Volcano plots of treatment effect at 1 month and 2 months.** The *t* test volcano plots arrange peptides by statistical significance. For each diffset, the *horizontal axis* represents the estimated log -fold change, and the *vertical axis* represents the log odds of differential expression (*B*). Significant peptides are those that have the largest absolute log-fold change and the highest log odds score. Among them, the most significant peptides are those found by ANOVA (highlighted in *red* (*up*) and *green* (*down*)) and distributed in the *top right* or *left* of the plots. *Left panel*, treatment effect at 1 month. *Right panel*, treatment effect at 2 months.

processing required prior to statistical analysis. Pro-alpha (2) 1 collagen is one such example. The estimated average $\log_2$-fold change at the 2 months comparison for this peptide is 1.82, which corresponds to a 3.31-fold change; however an average increase of 3.07 was observed in the absolute quantification for these animals at the same time point.

DISCUSSION

Diabetes nephropathy accounts for the majority of cases of ESRD diagnosed each year. Diabetic patients who progress to ESRD have high risk of mortality and therefore early intervention is important in the prevention of this disease. Unfortunately, treatment is initiated only after DNP is observed clinically with consistent protein urea in diabetic patients. Thus, new biomarkers for diabetic nephropathy that detect molecular and cellular changes within the kidney prior to the onset of nephropathy could significantly impact the manage-

ment of this disease. To evaluate the utility of label-free comparative analysis as a proteomic tool for biomarker discovery in DNP, we analyzed urine from the STZ experimental model of diabetes. The expected pathophysiology for altered end organ function because of diabetic complications in this model typically begins post 2 months STZ treatment (63, 64). Interestingly, we observed large expression changes at 2 months post-drug treatment as well some changes at 1 month post-drug treatment suggesting this methodology is a sensitive technique for early detection of DNP.

As mentioned above, of the 1429 peptides that progressed to statistical analysis, 633 had adjusted *p* values of 0.05 for the treatment and/or interaction (treatment + time) comparison at one or two months post-STZ treatment. Of these, 285 down-regulated peptides were observed. The hypoinsulinemic state of the STZ rat may also contribute to decrease in peptide/protein abundance observed in our study because insulin regulates gene expression for many proteins (65). One

| Protein Identification	Protein description	Number of peptides	Adju P value range (pFDR)	Direction of change
IPI00191711.1*	Major urinary protein precursor *	23	2.07E-05-4.91E-02	DOWN
gi 736002 emb CAA55977.1	Gy 5	10	5.56E-05-4.72E-02	UP
IPI00206823.1*	Alpha-2u globulin*	8	8.94E-05-1.60E-02	DOWN
IPI00188967.1	Ig kappa chain C region, A allele	3	2.35E-04-4.06E-02	UP
IPI00198021.3	Uromodulin precursor	2	6.43E-04-7.77E-03	DOWN
IPI00208410.1	Isoform 1 of Probasin precursor	3	6.78E-04-2.87E-02	DOWN
IPI00191773.1	Psbpc2 prostatic steroid-binding protein C2 precursor	4	7.67E-04-3.02E-02	DOWN
IPI00191771.1	Scgb2a2 secretoglobin family 2A member 2 precursor	2	1.23E-02-2.42E-02	DOWN
IPI00191774.1	Scgb1d2 prostatic steroid-binding protein C1 precursor	3	9.94E-03-1.22E-02	DOWN
IPI00201262.1	Mug2 Alpha-1-inhibitor 3 precursor	2	3.49E-03-9.08E-03	DOWN
IPI00205208.1	Andpro cystatin-related protein 1 precursor	6	1.66E-03-1.34E-02	DOWN
IPI00205210.1	Cystatin-related protein 2 precursor	2	4.40E-03-2.42E-02	DOWN
IPI00205275.1	Urinary protein 2 precursor	18	2.67E-03-4.56E-02	DOWN
IPI00213958.1	Prostatic spermine-binding protein precursor	2	3.95E-02-4.92E-02	DOWN
IPI00231193.3	Nerve growth factor, gamma	2	1.10E-02-1.55E-02	DOWN
IPI00392216.1	Liver regeneration protein	3	2.36E-02-3.79E-02	DOWN
IPI00779821.1	Serpin A1	3	2.32E-02-3.37E-02	UP
IPI00781465.1	Klk7_predicted 29 kDa protein	2	1.50E-02-4.87E-02	DOWN
IPI00782222.1	Pregnancy zone protein	2	1.41E-02-2.28E-02	UP
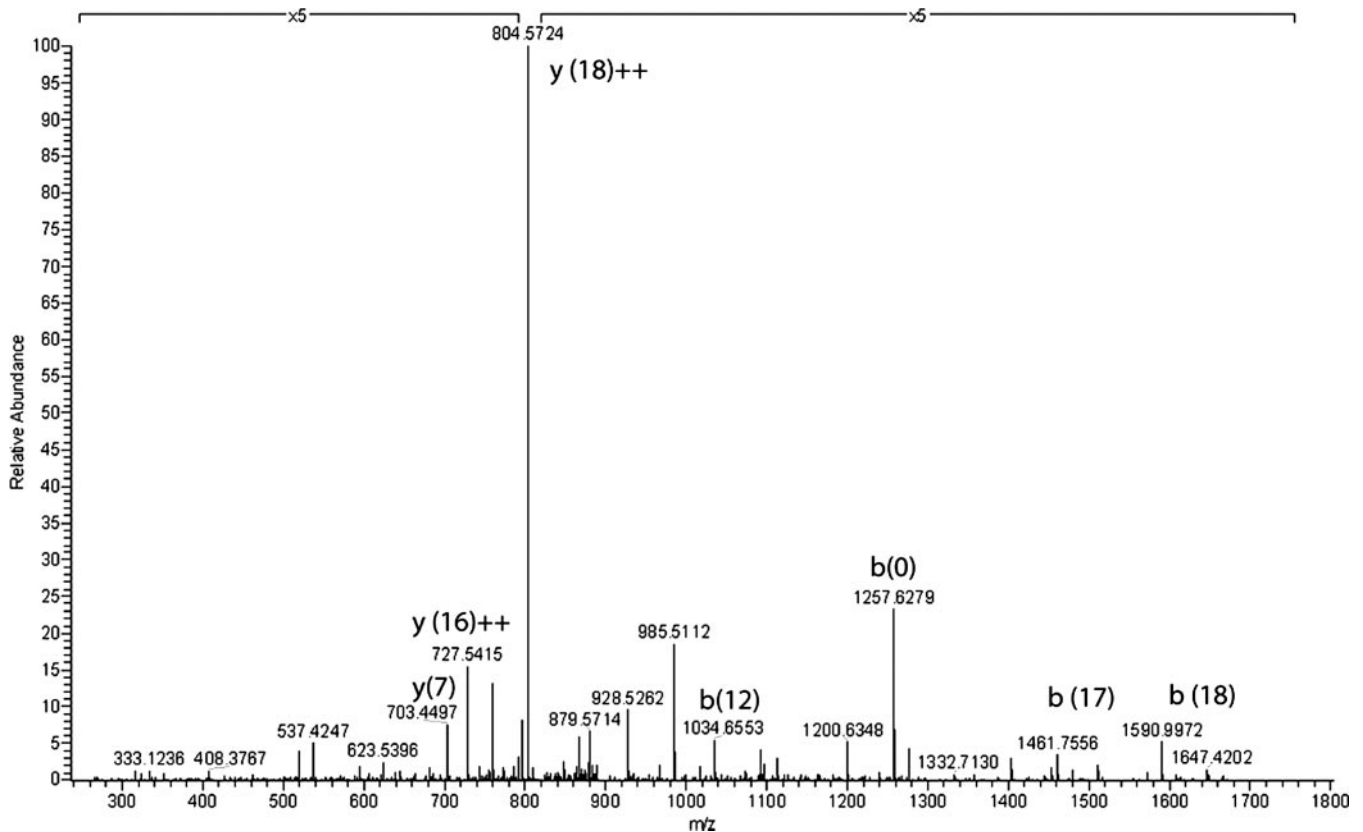
FIG. 8. **MS/MS of infused GEPGSVGAQGPPGPSGEEGK pro-alpha (2) type 1 peptide.** The *x* axis is *m/z*, and the y is relative abundance. $y(18)++$ denotes the +2 charge state of the y[18] fragment at a *m/z* value of 804.50. This fragment corresponds to PGSVGAQGPPGPSGEEGK.

FIG. 9. Box plots of illustrating abundance of pro-alpha (2) 1 collagen peptide GEPGSVGAQGPPGPSGEEGK in targeted SRM analysis (*A*) and label free expression analysis (*B*). Maximum value with a treatment group is represented by *blue box*, median value for a treatment group by *red box*, and minimum value for a treatment group by *green box*. This peptide was not detected in C 3 day, DM 3 day, or C 1 month time points in label-free analysis therefore; *red box* for these time points represents baseline imputation values. The mean and standard deviations (STDEV) for the SRM analysis were the following: C 3 day ($n = 1$) mean NA STDEV NA, DM 3 day ($n = 3$) mean 2.5, STDEV 1.9, C 1 month ($n = 2$) mean 3.1, STDEV 1.0, DM 1 month ($n = 4$) mean 6.2, STDEV 3.6, C 2 months ($n = 3$) mean 5.8, STDEV 0.96 and DM 2 months ($n = 3$) mean 17.8, STDEV 11.9.
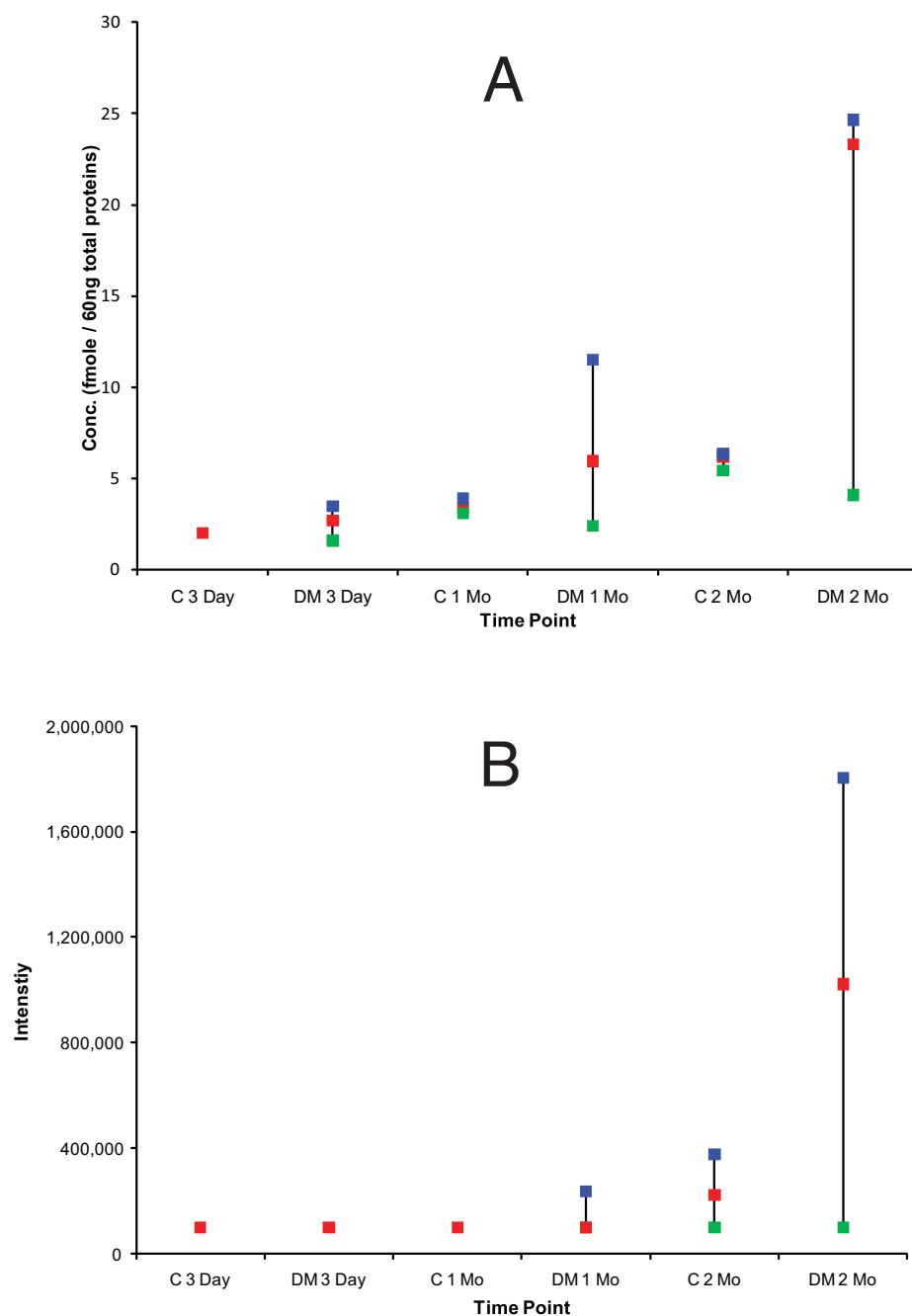
example is major urinary protein precursor, which was one of the most significant proteins observed in both the interaction and treatment analyses, with 23 peptides identified and an average adjusted *p* value of 3.68^{-03}. Major urinary protein in mature male rats accounts for 30–50% of total protein excreted. The primary synthesis site for this protein is the liver, and it is secreted into the blood. The protein is absorbed in the kidney from plasma. It has been shown to bind a number different hydrophobic small molecules, and its physiological role is believed to be pheromone transport in urine (66). Insulin deficiency in diabetic rats has been observed to trigger a

reduction of urinary output for this protein. The primary cause of this reduction is because of the regulatory role of insulin in major urinary protein expression (67, 68). In addition, our urinary results are consistent with Northern blot analysis of mRNA for this protein, which showed a significant decrease in expression at the mRNA level. In this study, the rate of transcription was decreased 10-fold with complete recovery after treatment of insulin in the streptozotocin rat (69).

Just as decreased insulin concentrations can regulate gene expression and corresponding protein abundance, the resulting high levels of circulating glucose because of hypoinsu-

linema can also effect gene expression and protein abundance. Pro-alpha (2) 1 collagen is one such example. Its protein abundance was increased at the 2 months time point and was highlighted as significant in this comparison with an adjusted p value of 0.047. In addition, SRM analysis of the samples analyzed by label-free analysis easily detected this peptide across all samples and confirmed the expression at 2 months post-STZ treatment as well as a potential expression change at one month post-STZ treatment. Pro-alpha (2) 1 collagen is a precursor to type 1 collagen, which is an extracellular matrix protein. It is of particular biological interest in diabetes as expansion of the glomerular mesangial matrix results in diffuse intercapillary sclerosis, which is the most important structural lesion in DNP (15, 16, 70). Studies utilizing animal models suggest that increased synthesis of extracellular matrix components occurs early in the progression of the disease where high levels of glucose stimulate gene expression of these proteins (71–74). Interestingly, another protein which was observed to have increased abundance at 2 months post-STZ treatment was Ig kappa light chain. Six peptides to this protein were identified as significant at this time point with an average adjusted p value of 0.0058. These results are consistent with observations made from a recent plasma proteomic study of STZ-treated rats in which a 7.5-fold decrease of this protein in plasma was observed in STZ-treated rat (75). Moreover, increased urinary excretion of this protein in patients with DM has been observed (76). Increased levels of Ig kappa light chain in plasma and its subsequent increase in urine may be of significant importance in the accumulation of extracellular matrix protein in DNP. Studies of other glomerular-associated diseases suggest a pathophysiological role of immunoglobulins in DNP. One example is myeloma-mediated monoclonal Ig deposition disease (MIDD). Myeloma results in increased plasma levels of Ig kappa, which subsequently is lodged in the kidney and leads to MIDD (77). MIDD is a disease in which kidney deposition of Ig subunits induces an accumulation of extracellular matrix protein resulting in structural lesions that are similar to those found in DNP (78). The similarity of glomerular lesions associated with DNP and those derived from myeloma MIDD suggest a common pathophysiological role of immunoglobulins in the development of DNP.

In this study, we have utilized label-free protein expression to study protein changes in urine in the streptozotocin model of diabetes. To the best of our knowledge, this work represents the first application of label-free expression to detect urinary protein changes from diabetic complications. We have successfully identified a number of protein changes specific to STZ treatment that occur early in the development of DNP. These represent potential markers of DNP that may assist in the diagnosis, treatment, and prevention of this complication. It should be noted that there are limitations of the STZ-rat model of experimental diabetes as no single experimental preparation is an ideal animal model for diabetes in humans.

In addition, some of the proteins identified in this study may not have human orthologs and therefore will not be suitable human targets. However, this analysis can be used to direct additional discovery efforts in human samples to confirm protein observations made in this study as well as identify novel human biomarker candidates as the techniques developed here can be easily translated to a human analysis.

REFERENCES

1. United States Renal Data System SAF: www.usrdsorg/odr/xrender_homeasp (2004) Accessed February 2008
2. Association, A. D. (2002) Diabetic nephropathy. *Diabetes Care* **25,** S85–S88
3. Maahs, D. M., and Rewers, M. (2006) Editorial: Mortality and renal disease in type 1 diabetes mellitus–Progress made, more to be done. *J. Clin. Endocrinol. Metab.* **91,** 3757–3759
4. Andersen, A. R., Christiansen, J. S., Andersen, J. K., Kreiner, S., and Deckert, T. (1983) Diabetic nephropathy in Type 1 (insulin-dependent) diabetes: an epidemiological study. *Diabetologia* **25,** 496–501
5. Remuzzi, G., Macia, M., and Ruggenenti, P. (2006) Prevention and treatment of diabetic renal disease in type 2 diabetes: the BENEDICT study. *J. Am. Soc. Nephrol.* **17,** Suppl. 2, S90–S97
6. Hughes, D. B., and Britton, M. L. (2005) Angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers for prevention and treatment of nephropathy associated with type 2 diabetes mellitus. *Pharmacotherapy* **25,** 1602–1620
7. Peters, T., Jr. (2006) How should we measure the albumin in urine? *Clin. Chem.* **52,** 555–556
8. Perkins, B. A., and Krolewski, A. S. (2005) Early nephropathy in type 1 diabetes: a new perspective on who will and who will not progress. *Curr. Diab. Rep.* **5,** 455–463
9. Kelly, D. J. (2000) Experimental diabetic nephropathy: Is it relevant to the human disease. *Nephrology* **5,** 177–185
10. Cooper, M. E., Allen, T. J., O'brien, R. C., Papazoglou, D., Clarke, B. E., Jerums, G., and Doyle, A. E. (1990) Nephropathy in a model combining genetic hypertension with experimental diabetes: Enalapril versus hydralazine and metoprolol therapy. *Diabetes* **39,** 1575–1579
11. O'Brien, R. C., Cooper, M. E., Jerums, G., and Doyle, A. E. (1993) The effects of perindopril and triple therapy in a normotensive model of diabetic nephropathy. *Diabetes* **42,** 604–609
12. Sassy-Prigent, C., Heudes, D., Jouquey, S., Auberval, D., Belair, M. F., Michel, O., Hamon, G., Bariety, J., and Bruneval, P. (1995) Morphometric detection of incipient glomerular lesions in diabetic nephropathy in rats. Protective effects of ACE inhibition. *Lab. Invest.* **73,** 64–71
13. Rossini, A. A., Like, A. A., Chick, W. L., Appel, M. C., and Cahill, G. F., Jr. (1977) Studies of streptozotocin-induced insulitis and diabetes. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 2485–2489
14. Allen, T. J., Cooper, M. E., O'Brien, R. C., Bach, L. A., Jackson, B., and Jerums, G. (1990) Glomerular filtration rate in streptozocin-induced diabetic rats. Role of exchangeable sodium, vasoactive hormones, and insulin therapy. *Diabetes* **39,** 1182–1190
15. Gilbert, R. E., Cox, A., Dziadek, M., Cooper, M. E., and Jerums, G. (1995) Extracellular matrix and its interactions in the diabetic kidney: a molecular biological approach. *J. Diabetes Complications* **9,** 252–254
16. Hirose, K., Osterby, R., Nozawa, M., and Gundersen, H. J. (1982) Devel-

opment of glomerular lesions in experimental long-term diabetes in the rat. *Kidney Int.* **21,** 689–695

17. McLennan, S. V., Fisher, E. J., Yue, D. K., and Turtle, J. R. (1994) High glucose concentration causes a decrease in mesangium degradation. A factor in the pathogenesis of diabetic nephropathy. *Diabetes* **43,** 1041–1045

18. Christ, G. J., Hsieh, Y., Zhao, W., Schenk, G., Venkateswarlu, K., Wang, H. Z., Tar, M. T., and Melman, A. (2006) Effects of streptozotocin-induced diabetes on bladder and erectile (dys)function in the same rat *in vivo. BJU Int.* **97,** 1076–1082

19. Yohannes, E., Chang, J., Christ, G. J., Davies, K. P., and Chance, M. R. (2008) Proteomics analysis identifies molecular targets related to diabetes mellitus-associated bladder dysfunction. *Mol. Cell. Proteomics* **7,** 1270–1285

20. Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., and Mann, M. (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol.* **7,** R80

21. Pang, J. X., Ginanni, N., Dongre, A. R., Hefta, S. A., and Opitek, G. J. (2002) Biomarker discovery in urine by proteomics. *J. Proteome Res.* **1,** 161–169

22. Sun, W., Li, F., Wu, S., Wang, X., Zheng, D., Wang, J., and Gao, Y. (2005) Human urine proteome analysis by three separation approaches. *Proteomics* **5,** 4994–5001

23. Zerefos, P. G., Vougas, K., Dimitraki, P., Kossida, S., Petrolekas, A., Stravodimos, K., Giannopoulos, A., Fountoulakis, M., and Vlahou, A. (2006) Characterization of the human urine proteome by preparative electrophoresis in combination with 2-DE. *Proteomics* **6,** 4346–4355

24. Hortin, G. L., Jortani, S. A., Ritchie, J. C., Jr., Valdes, R., Jr., and Chan, D. W. (2006) Proteomics: a new diagnostic frontier. *Clin. Chem.* **52,** 1218–1222

25. Pisitkun, T., Johnstone, R., and Knepper, M. A. (2006) Discovery of urinary biomarkers. *Mol. Cell. Proteomics* **5,** 1760–1771

26. Thongboonkerd, V. (2008) Searching for novel biomarkers and new therapeutic targets of diabetic nephropathy using proteomics approaches. *Contrib. Nephrol.* **160,** 37–52

27. Dihazi, H., Müller, G. A., Lindner, S., Meyer, M., Asif, A. R., Oellerich, M., and Strutz, F. (2007) Characterization of diabetic nephropathy by urinary proteomic analysis: identification of a processed ubiquitin form as a differentially excreted protein in diabetic nephropathy patients. *Clin. Chem.* **53,** 1636–1645

28. Hwang, H. J., Baek, Y. M., Kim, S. W., Kumar, G. S., Cho, E. J., Oh, J. Y., and Yun, J. W. (2007) Differential expression of kidney proteins in streptozotocin-induced diabetic rats in response to hypoglycemic fungal polysaccharides. *J. Microbiol. Biotechnol.* **17,** 2005–2017

29. Out, H. H., Can, H., Spentzos, D., Nelson, R. G., Hanson, R. L., Looker, H. C., Knowler, W. C., Monroy, M., Libermann, T. A., Karumanchi, S. A., and Thadhani, R. (2007) Prediction of diabetic nephropathy using urine proteomic profiling 10 years prior to development of nephropathy. *Diabetes Care* **30,** 638–643

30. Tilton, R. G., Haidacher, S. J., Lejeune, W. S., Zhang, X., Zhao, Y., Kurosky, A., Brasier, A. R., and Denner, L. (2007) Diabetes-induced changes in the renal cortical proteome assessed with two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* **7,** 1729–1742

31. Fliser, D., Novak, J., Thongboonkerd, V., Argilés, A., Jankowski, V., Girolami, M. A., Jankowski, J., and Mischak, H. (2007) Advances in urinary proteome analysis and biomarker discovery. *J. Am. Soc. Nephrol.* **18,** 1057–1071

32. Jantos-Siwy, J., Schiffer, E., Brand, K., Schumann, G., Rossing, K., Delles, C., Mischak, H., and Metzger, J. (2009) Quantitative urinary proteome analysis for biomarker evaluation in chronic kidney disease. *J. Proteome Res.* **8,** 268–281

33. Decramer, S., Gonzalez de Peredo, A., Breuil, B., Mischak, H., Monsarrat, B., Bascands, J. L., and Schanstra, J. P. (2008) Urine in clinical proteomics. *Mol. Cell. Proteomics* **7,** 1850–1862

34. Bogdanov, B., and Smith, R. D. (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* **24,** 168–200

35. Hu, S., Loo, J. A., and Wong, D. T. (2006) Human body fluid proteome analysis. *Proteomics* **6,** 6326–6353

36. Meyer, H. E., and Stuhler, K. (2007) High-performance proteomics as a tool in biomarker discovery. *Proteomics,* **7,** Suppl. 1, 18–26

37. Tao, W. A., and Aebersold, R. (2003) Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Curr. Opin. Biotechnol.* **14,** 110–118

38. Chelius, D., and Bondarenko, P. V. (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **1,** 317–323

39. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **75,** 4818–4826

40. Steffes, M. W., and Mauer, S. M. (1984) Diabetic glomerulopathy in man and experimental animal models. *Int. Rev. Exp. Pathol.* **26,** 147–175

41. Wang, P., Tang, H., Zhang, H., Whiteaker, J., Paulovich, A. G., and McIntosh, M. (2006) Normalization regarding non-random missing values in high-throughput mass spectrometry data. *Pac. Symp. Biocomput.* 315–326

42. Workman, C, Jensen, L. J., Jarmer, H., Berka, R., Gautier, L., Nielser, H. B., Saxild, H. H., Nielsen, C., Brunak, S., and Knudsen, S. (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* **3(9):**research0048

43. Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24,** 417–441

44. Efron, B. (2003) Robbins, empirical Bayes and microarrays. *Ann. Stat.* **31,** 366–378

45. Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001) Empirical Bayes analysis of a microarray experiment. *J. Amer. Stat. Assoc.* **96,** 1151–1160

46. Lönnstedt, I., and Speed, T. P. (2002) Replicated microarray data. *Stat. Sin.* **12,** 31–46

47. Newton, M. A., Kendziorski, C. M., Richmond, C. S., Blattner, F. R., and Tsui, K. W. (2001) On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J. Comput. Biol.* **8,** 37–52

48. Smyth, G. K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3:**Article3

49. Ge, Y., Dudoit, S., and Speed, T. P. (2003) Resampling-based multiple testing for microarray data analysis. *Test* **12,** 1–77

50. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 5116–5121

51. Kendziorski, C. M., Newton, M. A., Lan, H., and Gould, M. N. (2003) On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Stat. Med.* **22,** 3899–3914

52. Newton, M. A., Noueiry, A., Sarkar, D., and Ahlquist, P. (2004) Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5,** 155–176

53. Dabney, A., and Storey, J. D. (2003) Qvalue R package: Q-value estimation for false discovery rate control. *Comprehensive R Archive Network*

54. Cui, X., Kerr, M. K., and Churchill, G. A. (2003) Transformations for cDNA microarray data. *Stat. Appl. Genet. Mol. Biol.* **2:**Article4

55. Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B* **57,** 479–498

56. Storey, J. D. (2002) A direct approach to false discovery rates. *J. R. Stat. Soc.* **64,** 479–498

57. Storey, J. D. (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Stat.* **31,** 2013–2035

58. Storey, J. D., Taylor, J. E., and Siegmund, D. (2003) Strong control, conservative point estimation, and simultaneous conservative consistency of the false discovery rates: A unified approach. *J. R. Stat. Soc.* **66,** 187–205

59. Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18,** Suppl. 1, S96–S104

60. Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30(4):**e15

61. Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19,** 185–193

62. Deleted in proof

63. Chow, F. Y., Nikolic-Paterson, D. J., Atkins, R. C., and Tesch, G. H. (2004) Macrophages in streptozotocin-induced diabetic nephropathy: potential role in renal fibrosis. *Nephrol. Dial. Transplant.* **19,** 2987–2996

64. Tesch, G. H., and Nikolic-Paterson, D. J. (2006) Recent insights into experimental mouse models of diabetic nephropathy. *Nephron Exp. Nephrol.* **104,** e57–62

65. O'Brien, R. M., and Granner, D. K. (1996) Regulation of gene expression by insulin. *Physiol. Rev.* **76,** 1109–1161

66. Cavaggioni, A., and Mucignat-Caretta, C. (2000) Major urinary proteins, alpha(2U)-globulins and aphrodisin. *Biochim. Biophys. Acta* **1482,** 218–228

67. Roy, A. K., Chatterjee, B., Prasad, M. S., and Unakar, N. J. (1980) Role of insulin in the regulation of the hepatic messenger RNA for alpha 2u-globulin in diabetic rats. *J. Biol. Chem.* **255,** 11614–11618

68. Roy, A. K., and Neuhaus, O. W. (1966) Proof of the hepatic synthesis of a sex-dependent protein in the rat. *Biochim. Biophys. Acta* **127,** 82–87

69. Mira, E., and Castaño, J. G. (1989) Insulin short-term control of rat liver alpha 2-microglobulin gene transcription. *J. Biol. Chem.* **264,** 18209–18212

70. Mason, R. M., and Wahab, N. A. (2003) Extracellular matrix metabolism in diabetic nephropathy. *J. Am. Soc. Nephrol.* **14,** 1358–1373

71. Koya, D., Haneda, M., Nakagawa, H, Isshiki, K., Sato, H., Maeda, S., Sugimoto, T., Yasuda, H., Kashiwagi, A., Ways, D. K., King, G. L., and Kikkawa, R. (2000) Amelioration of accelerated diabetic mesangial expansion by treatment with a PKC beta inhibitor in diabetic db/db mice, a rodent model for type 2 diabetes. *FASEB J.* **14,** 439–447

72. Park, I. S., Kiyomoto, H., Abboud, S. L., and Abboud, H. E. (1997) Expression of transforming growth factor-beta and type IV collagen in early streptozotocin-induced diabetes. *Diabetes* **46,** 473–480

73. Ziyadeh, F. N., Sharma, K., Ericksen, M., and Wolf, G. (1994) Stimulation of collagen gene expression and protein synthesis in murine mesangial cells by high glucose is mediated by autocrine activation of transforming growth factor-beta. *J. Clin. Investig.* **93,** 536–542

74. Kreisberg, J. I., and Ayo, S. H. (1993) The glomerular mesangium in diabetes mellitus. *Kidney Int.* **43,** 109–113

75. Kim, S. W., Hwang, H. J., Kim, H. M., Lee, M. C., Shik Lee, M., Choi, J. W., and Yun, J. W. (2006) Effect of fungal polysaccharides on the modulation of plasma proteins in streptozotocin-induced diabetic rats. *Proteomics* **6,** 5291–5302

76. Groop, L., Mäkipernaa, A., Stenman, S., DeFronzo, R. A., and Teppo, A. M. (1990) Urinary excretion of kappa light chains in patients with diabetes mellitus. *Kidney Int.* **37,** 1120–1125

77. Ronco, P. M., Alyanakian, M. A., Mougenot, B., and Aucouturier, P. (2001) Light chain deposition disease: a model of glomerulosclerosis defined at the molecular level. *J. Am. Soc. Nephrol.* **12,** 1558–1565

78. Ronco, P., Plaisier, E., Mougenot, B., and Aucouturier, P. (2006) Immunoglobulin light (heavy)-chain deposition disease: from molecular medicine to pathophysiology-driven therapy. *Clin. J. Am. Soc. Nephrol.* **1,** 1342–1350