research articles Journal of proteome research

Application of Proteomics in the Discovery of Candidate Protein Biomarkers in a Diabetes Autoantibody Standardization Program Sample Subset

Thomas O. Metz,*,† Wei-Jun Qian,† Jon M. Jacobs,† Marina A. Gritsenko,† Ronald J. Moore,† Ashoka D. Polpitiya,† Matthew E. Monroe,† David G. Camp, II,† Patricia W. Mueller,‡ and Richard D. Smith†

Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, and Diabetes and Molecular Risk Assessment Laboratory, United States Centers for Disease Control and Prevention, Atlanta, Georgia

Received September 17, 2007

Novel biomarkers of type 1 diabetes must be identified and validated in initial, exploratory studies before they can be assessed in proficiency evaluations. Currently, untargeted "-omics" approaches are underutilized in profiling studies of clinical samples. This report describes the evaluation of capillary liquid chromatography (LC) coupled with mass spectrometry (MS) in a pilot proteomic analysis of human plasma and serum from a subset of control and type 1 diabetic individuals enrolled in the Diabetes Autoantibody Standardization Program, with the goal of identifying candidate biomarkers of type 1 diabetes. Initial high-resolution capillary LC-MS/MS experiments were performed to augment an existing plasma peptide database, while subsequent LC-FTICR studies identified quantitative differences in the abundance of plasma proteins. Analysis of LC-FTICR proteomic data identified five candidate protein biomarkers of type 1 diabetes. α-2-Glycoprotein 1 (zinc), corticosteroid-binding globulin, and lumican were 2-fold up-regulated in type 1 diabetic samples relative to control samples, whereas clusterin and serotransferrin were 2-fold up-regulated in control samples relative to type 1 diabetic samples. Observed perturbations in the levels of all five proteins are consistent with the metabolic aberrations found in type 1 diabetes. While the discovery of these candidate protein biomarkers of type 1 diabetes is encouraging, follow up studies are required for validation in a larger population of individuals and for determination of laboratory-defined sensitivity and specificity values using blinded samples.

Keywords: diabetes • proteomics • Fourier transform ion cyclotron resonance • label-free quantitation • liquid chromatography • mass spectrometry

Introduction

The best current approach for predicting who may be at risk for developing type 1 diabetes mellitus before onset of clinical symptoms is by measurement of autoantibodies to islet cell antigens.[1–3] In 1999, the United States Centers for Disease Control (CDC) and Prevention's National Diabetes Laboratory joined with the Immunology of Diabetes Society (IDS) in a collaborative agreement to form the Diabetes Autoantibody Standardization Program (DASP), with the common goal of improving the performance of assays for islet cell autoantibodies.[4,5] The DASP conducts periodic evaluations (since 2000) of selected laboratories performing assays for islet cell autoantibodies, and the results of the DASP 2005 proficiency evaluation demonstrated a high degree of concordance between laboratories (51 laboratories in 18 countries) in radiobinding assays for glutamic acid decarboxylase (GAD) and protein tyrosine phosphatase (IA-2) antibodies. However, for the major autoantibodies predictive of type 1 diabetes, the sensitivities adjusted to 95% specificity varied considerably: 81 ± 8% for GAD, 71 ± 5% for IA-2, and 47 ± 19% for insulin (mean ± standard deviation; P.W. Mueller, personal communication, 08/08/2007). These results indicate rather high sensitivity for the GAD autoantibody among laboratories but only moderate and poor sensitivity for IA-2 and insulin autoantibodies, respectively (all laboratories reported high specificity for the three autoantibodies, with mean values generally ≥94%). While the results from the first[5] and 2005 DASP proficiency evaluations are encouraging in terms of progress made in improved accuracy and reproducibility of measurement of the three islet cell autoantibodies among laboratories, surrogate biomarkers predictive of those at high risk for developing type 1 diabetes would benefit the clinical community, particularly if such surrogate biomarkers result in higher sensitivity and specificity.

Before surrogate biomarkers predictive of type 1 diabetes can be used for proficiency evaluations such as the DASP, they must first be identified and validated in exploratory studies. Hypoth-

* Corresponding author. P.O. Box 999/MS K8-98, Richland, WA 99352. Phone: (509) 376-8333. Fax: (509) 376-2303. E-mail: thomas.metz@pnl.gov.
† Pacific Northwest National Laboratory.
‡ United States Centers for Disease Control and Prevention.

10.1021/pr700606w CCC: $40.75
Published on Web 12/20/2007

esis-driven comparative studies may be performed to identify candidate biomarkers but are limited by the current knowledge of metabolic, signaling, and inflammatory pathways in the context of type 1 diabetes. In contrast, "-omics" technologies aim to discover candidate biomarkers by broadly surveying large numbers of species in an untargeted fashion using the selected approach. Proteomics[6–8] involves the structural and functional characterization of large numbers of proteins and generally attempts to determine quantitative differences in concentration (or expression) between comparative samples in a discovery-based (i.e., global) fashion. The strength of proteomics lies in the depth of coverage (depending on the particular technological platform chosen) of the proteome and the large amounts of data that can be generated and evaluated. Potential candidate biomarkers are likely waiting to be discovered in this sea of data, and effective strategies for characterizing, archiving, and contrasting data from multiple sample types under varying conditions (e.g., control versus disease) are needed.

In this work, we have applied the accurate mass and time (AMT) tag strategy[9–11] and label-free quantitation in an initial proteomic study of a DASP sample subset. LC-MS/MS analyses of healthy control and type 1 diabetic plasma and serum samples were initially performed to generate a list of confidently identified peptides, which were combined with peptides identified in two independent studies[12,13] of the human plasma proteome to create as comprehensive a peptide mass and time tag database as possible. Subsequent high-throughput and quantitative LC-MS analyses revealed significant differences in the plasma levels of a number of proteins including α-2-glycoprotein 1 (zinc), clusterin (apolipoprotein J), corticosteroid-binding globulin, lumican, and serotransferrin among healthy control individuals and patients recently diagnosed with type 1 diabetes. In particular, α-2-glycoprotein 1 (zinc), corticosteroid-binding globulin, and lumican showed significantly higher abundances in patients with type 1 diabetes, relative to healthy controls.

Materials and Methods

Reagents. All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless stated otherwise. Ammonium bicarbonate and methanol were purchased from Fisher Scientific (Fair Lawn, NJ), while sequencing grade trypsin was purchased from Promega (Madison, WI). Ammonium formate was obtained from Fluka (St. Louis, MO), and a bicinchoninic acid (BCA) Protein Kit was purchased from Pierce (Rockford, IL). Purified, deionized water, >18 MΩ, (Nanopure Infinity ultrapure water system, Barnstead, Newton, WA) was used to make all aqueous biological and HPLC buffers, with the exception of immunodepletion buffers A and B, which were purchased from Agilent Technologies (Palo Alto, CA). The composition of these buffers was not provided by the manufacturer.

Human Plasma and Serum Samples. The DASP is conducted in accordance with the Human Subjects policies and regulations of the CDC. Similarly, this work was approved by the Institutional Review Board of the Pacific Northwest National Laboratory. Human plasma and serum samples from the DASP (years 2000–2005) were received frozen on dry ice; these samples correspond to healthy control individuals ($n = 10$) and patients recently diagnosed with type 1 diabetes ($n = 10$). The patient samples were collected from donors under the age of 30 within 14 days of starting insulin treatment[5] before insulin antibodies (as opposed to insulin autoantibodies) were induced. Because patient samples were collected at multiple sites worldwide, investigators contributing material to the DASP were not limited to a single blood collection and subsequent plasma or serum preparation protocol.

Immunoaffinity Subtraction Using the Multiple Affinity Removal System (MARS). For LC-MS/MS analyses, 20 μL of each individual control and patient sample was pooled to form two distinct samples, pooled control and pooled patient (200 μL each). Six high-abundance plasma proteins—albumin, IgG, antitrypsin, IgA, transferrin, and haptoglobin— that constitute approximately 85% of the total protein mass of human plasma were then removed from each pooled sample using a multiple affinity removal system LC column (MARS; 4.6 mm × 100 mm; Agilent) coupled with an Agilent 1100 series HPLC system. Briefly, each pooled sample was diluted 5-fold with loading buffer (buffer A), and 40 μL of diluted sample was injected for each run. High-abundance proteins were removed per the manufacturer's instructions using the ready-to-use loading/washing and eluting buffers (buffers A and B, respectively, Agilent); the flow-through fractions were subsequently pooled for each corresponding sample type (control or patient). The pooled samples were then concentrated in vacuo, desalted using PD-10 columns (GE Healthcare Bio-Sciences, Piscataway, NJ), and concentrated again in vacuo.

For LC-FTICR analyses, 35 μL each of individual control and patient samples was depleted of the top six high-abundance plasma proteins, as described above.

Enzymatic Digestion of Plasma Proteins. Individual and pooled plasma samples were enzymatically digested with trypsin prior to analysis. Briefly, proteins were sequentially denatured and reduced by incubation with 8 M urea followed by incubation with 10 mM dithiothreitol, both for 1 h at 37 °C. Cysteine residues were then alkylated by reaction with 40 mM iodoacetamide in the dark for 1 h at 37 °C, and the samples were subsequently desalted using PD-10 columns. Trypsin was added at a 1:50 ratio (enzyme to protein) in 50 mM ammonium bicarbonate buffer containing 1 mM calcium chloride and 20% acetonitrile, and the samples were incubated for 3 h at 37 °C. Denaturation, reduction, alkylation, and digestion were all performed in a Thermomixer R (Eppendorf, Hamburg, Germany) at the stated temperatures, with constant shaking at 800 rpm. The digests were cleaned using C-18 SPE tubes (Discovery DSC-18, Supelco, Bellefonte, PA), and bound peptides were eluted with 80% acetonitrile containing 0.1% trifluoroacetic acid. Eluted peptides were then concentrated in vacuo to 50–100 μL.

Strong-Cation Exchange Fractionation of Enzymatic Digests. Prior to LC-MS/MS analyses, strong-cation exchange fractionation (SCX) of enzymatic digests (pooled control and pooled patient samples) was performed as previously described.[14,15] Briefly, peptides were diluted with 900 μL of 10 mM ammonium formate (pH 3.0) in water containing 25% acetonitrile and fractionated by SCX chromatography using a Polysulfethyl A column (2.1 mm × 200 mm; PolyLC, Columbia, MD) that was preceded by a guard column (2.1 mm × 10 mm) of the same material. The separations were performed at a flow rate of 0.2 mL/min using an Agilent 1100 series HPLC system with mobile phases consisting of 10 mM ammonium formate (pH 3.0) in water containing 25% acetonitrile (A) and 500 mM ammonium formate (pH 6.8) in water containing 25% acetonitrile (B). After loading 900 μL of sample onto the column (~900 μg peptides), the gradient was maintained at 100% A for 10 min. Peptides were then separated using a gradient from 0

to 50% B over 40 min, followed by a gradient of 50–100% B over 10 min. The gradient was then held at 100% B for 10 min. A total of 25 fractions each were collected for enzymatic digests of pooled control and pooled patient samples, and all fractions were dried under vacuum.

Reversed-Phase Capillary LC-MS/MS and LC-FTICR Analyses. For LC-MS/MS analyses, dried peptide fractions were reconstituted in 30 μL of 25 mM ammonium bicarbonate, pH 7.8, and analyzed using a custom-built capillary LC system coupled online to a linear ion trap mass spectrometer (LTQ; ThermoElectron, Waltham, MA) by way of an in-house manufactured electrospray ionization interface as previously described.[16] Reversed-phase capillary columns were prepared by slurry packing 5-μm Jupiter C18 bonded particles (Phenomenex, Torrence, CA) into a 150 μm × 65 cm fused silica capillary (Polymicro Technologies, Phoenix, AZ) that utilized a 2-μm stainless steel retaining screen within a stainless steel union (Valco Instruments Co., Houston, TX). The mobile phase consisted of 0.2% acetic acid and 0.05% TFA in water (A) and 0.1% TFA in 90% acetonitrile/10% water (B). Mobile phases were degassed online using a vacuum degasser (Jones Chromatography Inc., Lakewood, CO), and the HPLC system was equilibrated at 5000 psi with 100% mobile phase A for initial starting conditions. After loading 10 μg of peptides onto the column, the mobile phase was held at 100% A for 20 min. Exponential gradient elution was performed by increasing the mobile-phase composition from 0 to 77% B over 100 min, using a stainless steel mixing chamber, followed by column washing at ~100% B for 10 min. The exponential gradient results from continuous delivery of mobile phase B into the mixer under constant pressure[17] and can be approximated using a multistep gradient performed on a conventional LC system. To identify the eluting peptides, the LTQ was operated in a data-dependent MS/MS mode (m/z 400–2000), in which a full MS scan was followed by ten MS/MS scans using a normalized collision energy of 35%. A dynamic exclusion window of 1 min was used to discriminate against previously analyzed ions. The temperature of the heated capillary and the ESI voltage were 200 °C and 2.2 kV, respectively.

Additionally, unfractionated peptide digests from both the control and patient sample pools were analyzed in duplicate.

For LC-FTICR analyses, peptide samples were analyzed in triplicate using chromatographic and electrospray conditions identical to those for LC-MS/MS analyses. The LC system was interfaced to a modified 9.4 T FTICR MS (Bruker Daltonics, Billerica, MA) via an electrodynamic ion funnel[18] coupled to a radio frequency quadrupole for collisional ion focusing and high-efficiency ion accumulation and transport to the cylindrical ICR cell. The temperature of the heated capillary and the ESI voltage were 200 °C and 2.2 kV, respectively.

Development of the AMT Tag Database. The SEQUEST analysis software[19] was used to match the MS/MS fragmentation spectra with sequences from the April 2005 IPI human protein database (version 3.05), containing 49 161 entries; a dynamic mass modification on cysteinyl residues corresponding to alkylation by iodoacetamide (57.0215 Da) was applied during SEQUEST analyses. Peptides identified in both this study and two independent human plasma proteomic studies[12,13] were stored as AMT tags in a Microsoft SQL Server database.

Peptides passing the SEQUEST filter criteria shown in Table 1 were added to the AMT tag database. The database includes up to the top 3 scoring peptide hits per MS/MS spectra if the peptide identifications meet all criteria. In situations where SEQUEST produced more than one peptide identification from a given spectrum, it is possible to identify the true match if accurate mass information is exploited as in the subsequent LC-FTICR analyses. The final plasma AMT tag database contained 19 579 human plasma peptides available for matching to subsequent high-throughput LC-FTICR experiments of similar samples. The peptide elution times from each LC-MS/MS analysis were normalized to a range of 0 to 1 using a predictive peptide LC normalized elution time (NET) model and linear regression, as previously reported.[20] A NET average and standard deviation were assigned to each identified peptide if the same peptide was observed in multiple analyses. Both calculated monoisotopic masses and observed NETs of identified peptides were included in the AMT tag database.

Table 1. SEQUEST Filter Criteria Used to Identify Peptides for AMT Tag Database Population[a]

charge state	cleavage type	Xcorr	DelCn
1+	fully tryptic	≥1.5	≤0.05
1+	partially or nontryptic protein terminal	≥3.0	≤0.05
2+	fully tryptic	≥2.7	≤0.05
2+	partially or nontryptic protein terminal	≥3.7	≤0.05
≥3+	fully tryptic	≥3.3	≤0.05
≥3+	partially or nontryptic protein terminal	≥4.5	≤0.05

[a] Xcorr - cross correlation score between observed and theoretical peptide fragmentation spectra. DelCn - difference between the Xcorr of the top scoring peptide and Xcorr of the nth ranked peptide in a given spectrum.

Processing of Quantitative LC-FTICR Data Sets. LC-FTICR data sets were processed using the PRISM Data Analysis system,[21] a series of software tools (e.g., ICR2LS, VIPER;[22] freely available at http://ncrr.pnl.gov/software/) developed in-house. The first step involved deisotoping of the raw MS data to give the monoisotopic mass, charge state, and intensity of the major peaks in each mass spectrum.[11] The data were next examined in a 2-D fashion to identify groups of mass spectral peaks that were observed in sequential spectra using an algorithm that computes a Euclidean distance in n-dimensional space for combinations of peaks.[11] Each group, generally ascribed to one detected species and referred to as a "feature", has a median monoisotopic mass, central NET, and abundance estimate computed by summing the intensities of the MS peaks that comprise the entire LC-FTICR feature.

The peptide identities of detected features in each data set (here a data set is equivalent to a single LC-FTICR analysis) were determined by comparing their measured monoisotopic masses and NETs to the calculated monoisotopic masses and observed NETs of each of the 19 579 peptides in the AMT tag database within initial search tolerances of ± 6 ppm and ± 0.025 NET for monoisotopic mass and elution time, respectively. This peak-matching process gave an initial list of peptide identifications for each individual data set. The final combined list of peptide identifications was further filtered such that the average monoisotopic mass and NET error were less than 2 ppm and 0.02 NET, respectively; in addition, all peptides were required to be observed in at least 10 of 60 LC-FTICR data sets. All quantified peptides were rolled up to nonredundant protein groups using ProteinProphet.[23]

Data Normalization and Relative Quantitation. Normalization of peptide intensities from multiple LC-FTICR analyses was performed using a simple global normalization procedure, as previously described.[24] Briefly, a composite reference data set was initially generated by averaging peptide intensities across

all data sets. A normalization factor was computed for each data set as a median of peptide abundance ratios to the peptide abundances of the reference set. The peptide abundances for each data set were normalized by simply dividing the originally observed abundances by the computed normalization factor. Following normalization, the peptide abundances observed across triplicate data sets for each control and patient sample were averaged to produce average peptide abundances for each species identified in each individual.

To improve data visualization and relative quantitation, normalized peptide abundances were converted to $\log_2$ ratio format by dividing the average peptide abundances by the abundances of the same peptides in the reference data set, followed by transformation to $\log_2$ scale. Peptide abundances were thus expressed as $\log_2$ ratios relative to the common reference data set. Relative peptide abundances were then "rolled up" to the protein level by taking the median of the relative abundances for those peptides mapping to each protein. To identify candidate protein biomarkers, *t* tests were performed at both the peptide and protein levels comparing the relative protein abundances between the control and patient samples using the open-source tool TIGR Multiexperiment Viewer (MEV).[25]

Results

LC-MS/MS Analyses of Pooled Control and Pooled Patient Samples. An initial LC-MS/MS proteomic characterization of pooled control and pooled patient samples from a DASP sample subset was performed to expand an existing plasma AMT tag database through the inclusion of plasma proteins that may be specific to type 1 diabetes. The peptide identifications obtained in the current study were used to augment an existing AMT tag database of plasma peptides identified in two independent human plasma proteomic studies.[12,13] In this study, a total of 437 024 and 433 837 tandem mass spectra were collected for pooled control and pooled patient samples, respectively, which, when combined with peptide identifications from previous studies,[12,13] resulted in an AMT tag database comprising 19 579 unique peptides passing the SEQUEST filter criteria described in Table 1.

The MS/MS peptide false discovery rate (FDR) was estimated to be 16.2% when using these filter criteria in conjunction with a decoy database search.[26] As described in Materials and Methods, the AMT tag database can include up to the top 3 scoring peptide hits per MS/MS spectrum. However, it is possible to discern the true identification based on the strict monoisotopic mass tolerances utilized when matching LC-FTICR features to the AMT tag database. In addition, the AMT tags available for matching to LC-FTICR data sets can be further refined through the application of more stringent SEQUEST filter criteria without the need to repopulate or construct a new database.

Peptide identifications from SEQUEST processing of LC-MS/MS data were further refined in a separate analysis that examined the highest-scoring peptide for each spectrum, based on the more stringent filter criteria shown in Table 2. These peptide identifications were used for semiquantitative data analyses based on spectrum counting to support LC-FTICR data, as discussed below. The MS/MS peptide FDR was estimated to be 3.6% when using these filter criteria in conjunction with a decoy database search.[26]

Table 2. SEQUEST Filter Criteria Used in Semiquantitative Data Analyses Based on Spectrum Counting[a]

charge state	cleavage type	Xcorr	DelCn
1+	fully tryptic	≥1.6	≥0.1
2+	fully tryptic	≥2.4	≥0.1
2+	partially or nontryptic protein terminal	≥4.3	≥0.1
≥3+	fully tryptic	≥3.2	≥0.1
≥3+	partially or nontryptic protein terminal	≥4.7	≥0.1

[a] Xcorr - cross correlation score between observed and theoretical peptide fragmentation spectra. DelCn - difference between the Xcorr of the top scoring peptide and Xcorr of the second highest scoring peptide in a given spectrum.

Quantitative LC-FTICR Analysis of Individual Control and Patient Samples. For the purpose of candidate biomarker discovery, we applied a label-free quantitation strategy based on LC-FTICR analyses of individual samples from control and patient subjects without any fractionation. Peptides, and subsequently proteins, were identified by matching detected LC-FTICR features to the plasma AMT tag database using the AMT tag approach. In this study, 9116 ± 710 (mean ± standard deviation) features from 59 LC-FTICR data sets (one data set for control individual 5 was excluded as an outlier, see below) were matched to 1517 ± 199 AMT tags, representing ∼16% of data identified. The remaining ∼84% of features remains unidentified due to a variety of reasons including undersampling during LC-MS/MS analyses, the presence of nonpeptide sample components, and both enzymatic and nonenzymatic post-translational modification of proteins that was not accounted for during SEQUEST processing of MS/MS data.

A total of 1930 unique peptides were identified during this quantitative study, corresponding to 120 unique proteins (≥2 unique peptides per protein), based on stringent filtering as described in the Materials and Methods. Figure 1 shows the average monoisotopic mass and NET error histograms for all identified peptides, including those outside of the final monoisotopic mass and NET tolerances. On the basis of the monoisotopic mass error histogram, we estimate the false positive error rate at the peptide level as ∼7% using a monoisotopic mass tolerance of 2 ppm. This error rate was calculated by first determining the baseline level of false database matches, which is represented in Figure 2 as the tails of the monoisotopic mass error histogram. The false positive error rate is then calculated as the area within the monoisotopic mass tolerance of the database search corresponding to false positives (red area in Figure 2) divided by the sum of the areas within the monoisotopic mass tolerance search corresponding to both true (blue area in Figure 2) and false positives. The value obtained in the current work is similar to that reported in a recent label-free LC-MS-based proteomics study, which utilized an artificial 11 Da monoisotopic mass-shift approach to estimate the false positive error rate as 8.6%.[27] In the current study, the protein level false positive error rate is estimated as ∼1% using the method of MacCoss and colleagues[28] and based on the requirement of at least two unique peptides per protein identification.

To evaluate the reproducibility of the approach, we analyzed the correlation coefficients of peptide abundances across technical replicates as well as across different patient subjects. Figure 3 shows typical peptide abundance correlation plots between technical replicates (the same digest analyzed in

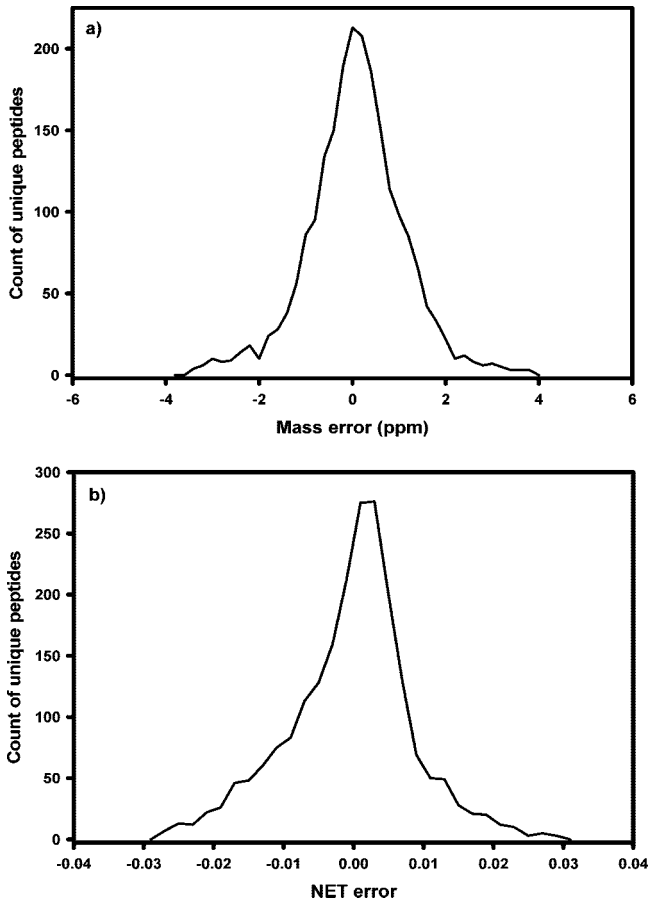

Figure 1. Average monoisotopic mass (a) and NET (b) error histograms for all identified peptides used in the final data analysis.

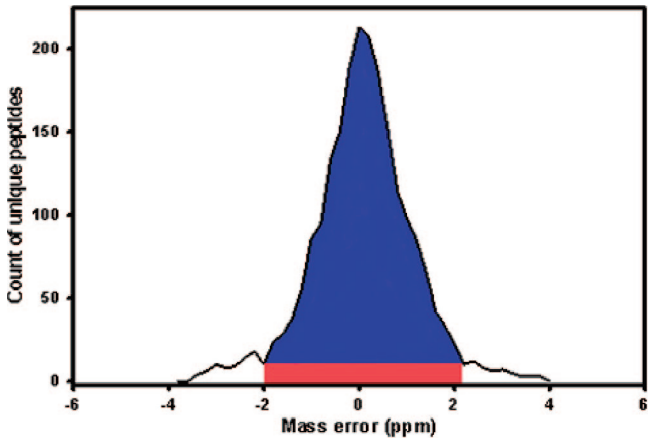

Figure 2. Average monoisotopic mass error histogram for all identified peptides used in the final data analysis, showing area of false positive (FP; red) and true positive (TP; blue) identifications. The false positive error rate was calculated as FP/(FP + TP).

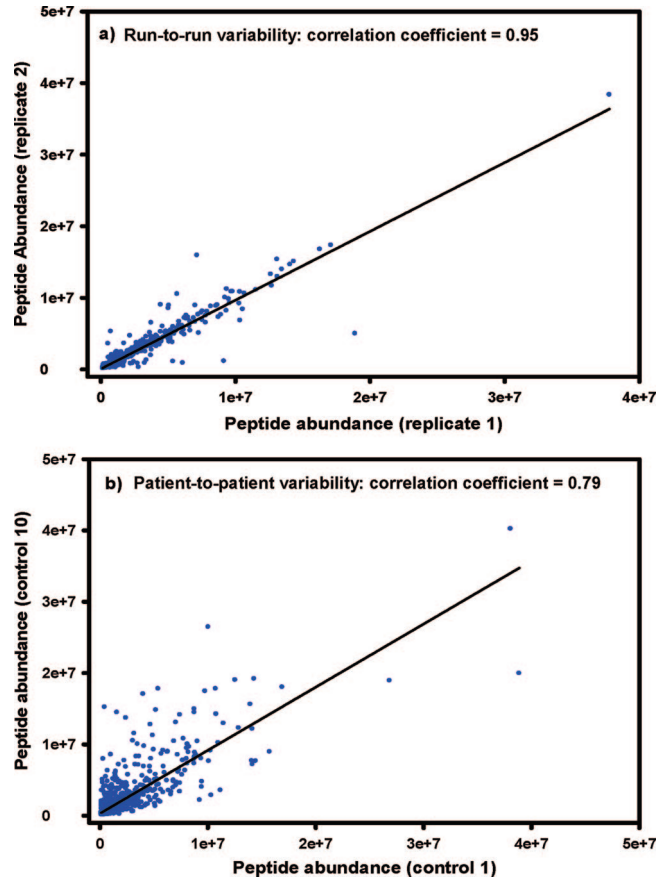

Figure 3. Scatter plots of peptide abundances showing correlation of various replicate analyses. (a) LC-FTICR technical replicates of control sample 1. (b) Biological replicates from two control individuals. The Pearson correlation coefficients for both comparisons are shown.

replicate by LC-FTICR) and between different patient subjects. As shown in Figure 3a, good instrumental reproducibility was observed among technical replicates, with an average Pearson correlation coefficient of 0.95 ± 0.04. However, relatively large subject-to-subject variations were observed as exemplified in Figure 3b. Figure 4 shows the correlation patterns among all 20 subjects, where an average correlation of 0.78 ± 0.08 was observed. The relative consistency of LC-FTICR analyses was also evaluated based on the number of peptides identified in each data set. Figure 5 shows the relative consistency of peptide identifications between technical replicates for each individual and across all 20 individuals. As shown, most technical replicates had standard deviations less than 5% of the total number of unique peptide identifications, except for control individual 5 (C5) for which one LC-FTICR replicate displayed a significant drop in overall intensity. On the basis of this observation, this particular data set was considered as an outlier and was not used in the calculation of average peptide abundances for this individual. The total number of unique peptide identifications across all 20 individuals was also quite consistent given the expected biological variations between samples.

To identify statistically confident candidate protein biomarkers, a global normalization approach was applied across all data sets. Peptide relative abundances in $\log_2$ ratio format were mean-centered, and protein-level relative abundances were estimated as the median of peptide relative abundances for all peptides originating from a given protein. Following the application of a statistical t test at the protein level, 30 out of a total 120 proteins were shown to have significant abundance differences ($p < 0.05$) between control and patient individuals (see Supplemental Data Table 1). These proteins were further examined at the peptide level where we required the average peptide p-value to be <0.1. Figure 6 shows a heat map of relative peptide abundances for 9 proteins that were found to

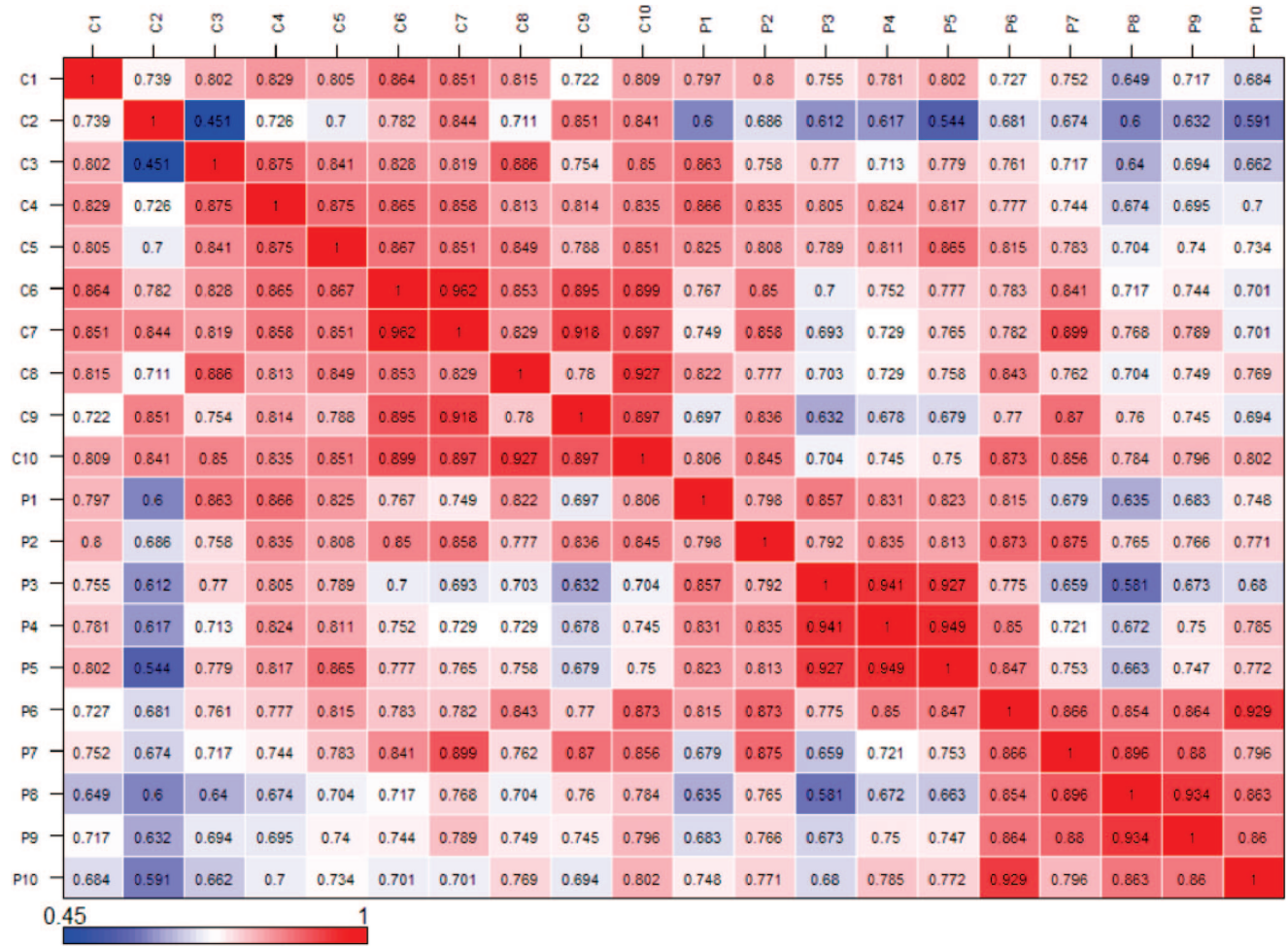

Figure 4. Heat map showing Pearson correlation coefficients of average peptide abundances ($n = 3$*) for individual control and patient samples. Each block in the heat map represents a comparison between two individuals, and the Pearson correlation value is shown. C = control and P = patient. *Note: One technical replicate for control individual 5 displayed a significant drop in overall intensity. On the basis of this observation, this particular data set was considered as an outlier and was not used in the calculation of average peptide abundances for this individual.

be significantly different between control and patient individuals based on both the peptide- and protein-level *t* tests. Of these 9 proteins, only 5 were identified with ≥2 unique peptides in at least 8 of 10 control and patient individuals; therefore, emphasis will be placed on α-2-glycoprotein 1 (zinc), clusterin (apolipoprotein J), corticosteroid-binding globulin, lumican and serotransferrin. Table 3 lists the mean abundances for these proteins in control and patient individuals, as well as the protein and average peptide *p*-values. It is important to note that by using spectrum counting methods, data from the LC-MS/MS analyses of pooled samples showed similar trends, but not large magnitudes of change (except in the case of lumican and serotransferrin), in support of the findings from LC-FTICR analyses of individual samples (see Table 4); however, it is difficult to evaluate the impact of the LC-MS/MS data due to the qualitative nature of the pooled sample comparison. In addition, while pooling individual samples offers the advantage of obtaining a representative profile of the control and patient plasma proteome, it also presents the disadvantage that individuals with higher or lower expression levels of a given protein or proteins may bias the measurement of said protein(s).

Discussion

α-2-Glycoprotein 1 (Zinc). Relative levels of α-2-glycoprotein 1 (zinc) were determined in all individuals using three unique peptides. As shown in Figure 6, this protein was consistently up-regulated in patient samples versus controls; however, control individuals 2 and 3 exhibited relatively higher levels compared to the other eight control individuals. α-2-Glycoprotein 1 (zinc) is a member of the immunoglobulin superfamily[29] and displays lipid mobilization activity.[30,31] The patient samples collected as part of the DASP study correspond to recently diagnosed individuals, several of whom may have been ketotic; therefore, increased levels of α-2-glycoprotein 1 (zinc) in patient relative to control samples may be an indication of a system-wide mobilization of fats for energy production. Indeed, the patient samples were suspected to be hyperlipidemic relative to control samples during sample preparation for proteomic analysis, based on observed opaqueness and increased viscosity (also possibly due to hyperglycemia) of the plasma and serum. In addition, several isoforms of α-2-glycoprotein 1 (zinc) have been shown to exist in human plasma[32] and human serum,[33] leaving open the possibility of

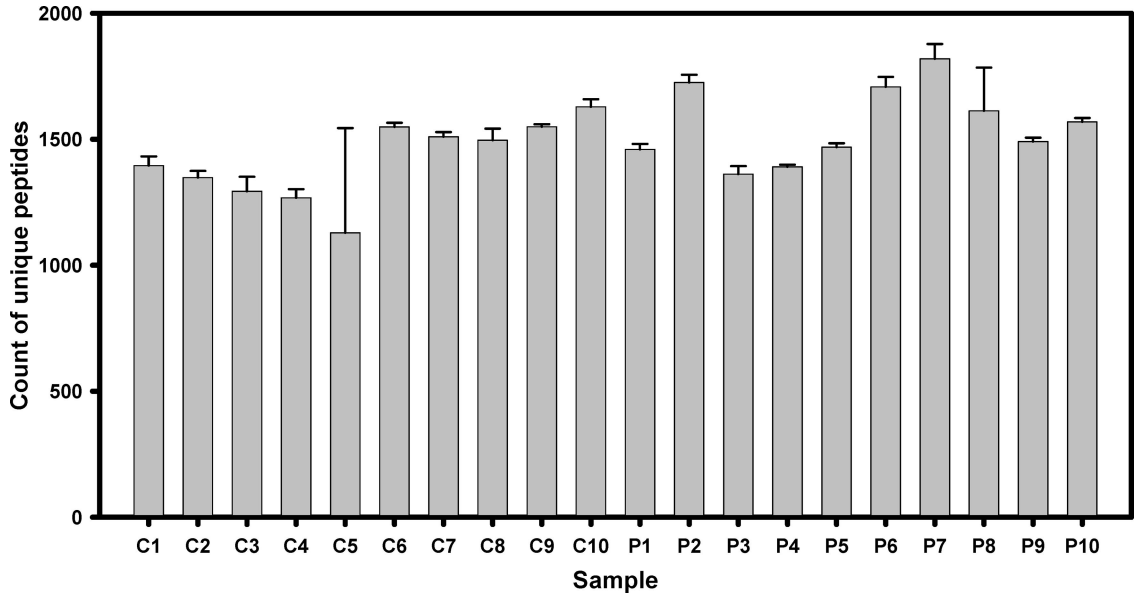

Figure 5. Reproducibility of LC-FTICR analyses in terms of the number of unique peptide identifications per individual. Data are shown as mean ± standard deviation. *C* = control and *P* = patient.

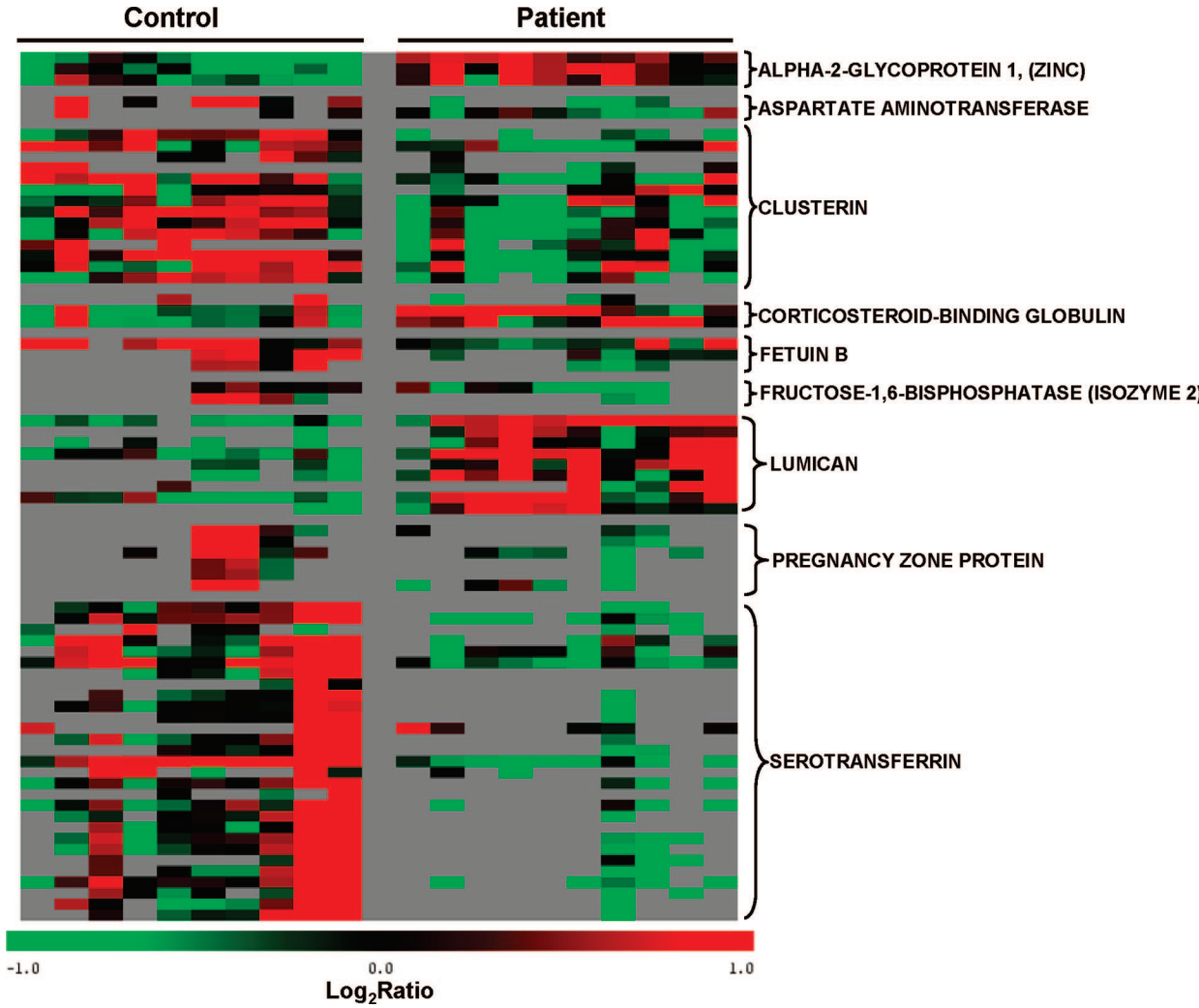

Figure 6. Heat map of relative protein abundances displayed as their constituent relative peptide abundances. Peptide relative abundances in $\log_2$ ratio format were mean-centered, and protein level relative abundances were estimated as the median of peptide relative abundances for all peptides originating from a given protein. Only those proteins passing a *t* test on both the protein ($p < 0.05$) and peptide ($p < 0.1$) levels are shown. Each row represents a unique peptide.

Table 3. Proteins Identified in LC-FTICR Data Showing Statistically Significant Abundance Differences Based on Both a Peptide and Protein Level *t* Test

protein description	protein IPI number	gene	mean protein abundance (control)	mean protein abundance (patient)	protein *p* value	average peptide *p* value
α-2-glycoprotein 1 (zinc)	IPI00166729.2	AZGP1	−0.79	0.43	0.00081	0.0022
clusterin	IPI00291262.3	CLU	0.43	−0.62	0.00030	0.077
corticosteroid-binding globulin	IPI00027482.1	SERPINA6	−0.41	0.47	0.012	0.084
lumican	IPI00020986.2	LUM	−0.68	0.55	0.00012	0.015
serotransferrin	IPI00022463.1	TF	0.30	−0.73	0.002	0.055

Table 4. Peptide Spectrum Counts Obtained from Capillary LC-MS/MS Data for the Five Candidate Protein Biomarkers of Type 1 Diabetes

protein description	protein IPI number	gene	peptide obsd control	peptide obsd patient
α-2-glycoprotein 1 (zinc)	IPI00166729	AZGP1	40	44
clusterin (apolipoprotein J)	IPI00291262	CLU	195	220
corticosteroid-binding protein	IPI00027482	SERPINA6	40	58
lumican	IPI00020986	LUM	54	108
serotransferrin	IPI00022463	TF	126	43

genetic differences in the expression or isoforms of this protein in patients relative to controls.

Clusterin. Fourteen unique peptides were used to determine relative levels of clusterin in patient and control samples, although detection of peptides was variable among individuals. Clusterin, or apolipoprotein J, demonstrates a nearly ubiquitous tissue distribution and is produced in two immunologically distinct forms by alternative splicing.[34] A glycosylated form is secreted and displays cytoprotective effects, whereas a nonglycosylated, cytoplasmic/nuclear form displays apoptotic properties.[35,36] In this study, 7 control individuals exhibited relatively high levels of clusterin, whereas 5 patients exhibited relatively low levels of this protein. It is possible that low levels of clusterin in patient samples indicate poor cytoprotective capability, which may facilitate autoimmune destruction of pancreatic beta cells leading to the development of type 1 diabetes mellitus.

Corticosteroid-Binding Globulin. Relative levels of corticosteroid-binding globulin were determined in control and patient individuals using up to three unique peptides (at least two unique peptides mapping to this protein were detected in all individuals). This protein was consistently up-regulated in patient samples versus controls; however, control individuals 2 and 9 exhibited relatively higher levels compared to the other eight control individuals. Insulin and insulin-like growth factor I have been reported to inhibit secretion of corticosteroid-binding globulin by a human hepatoma cell line,[37] and plasma corticosteroid-binding globulin was recently shown to correlate positively with fasting glucose and hemoglobin A1c but negatively with insulin response to i.v. and oral glucose administration in obese individuals with glucose intolerance.[38] In the current study, the observation of relatively high levels of corticosteroid-binding globulin in patients relative to controls likely reflects the absence or lower circulating levels of insulin in the plasma and serum of patients.

Lumican. Up to nine unique peptides were used to determine relative levels of lumican in patient and control samples, although detection of peptides was variable among individuals. This protein was consistently and strongly up-regulated in patient samples versus controls, with the exception of patients 1 (no up-regulation), 7, and 8 (moderate up-regulation). In addition, the LC-MS/MS data for this protein (Table 4), while semiquantitative, strongly support these results. Lumican is a member of the small leucine-rich proteoglycan family[39] and is a component of the extracellular matrix (ECM) and binding partner of TGF-β, a key mediator of fibrotic diseases. This protein was reported to be expressed in the cornea, dermis, cartilage, heart, lung, and kidney of developing mouse embryos,[40] thus it is difficult to speculate on the specific tissue source of plasma/serum lumican in patient samples. However, lumican has been shown to be up-regulated in the tubulointerstitium and glomeruli of diabetic patients with nephropathy.[39] In addition, lumican (as well as other proteoglycans) has been reported to be up-regulated in kidneys of rats experiencing unilateral ureteral obstruction.[41] These studies suggest that increased thickening of the ECM and possibly glomerulosclerosis may be accompanied by up-regulation of select proteoglycans, including lumican. Because the samples involved in the DASP study are from recently diagnosed patients, it is difficult to estimate the degree to which plasma hyperglycemia affects the morphology and function of the kidney. However, it is possible that increased plasma lumican in patient samples reflects an acute response of the kidney to high levels of plasma glucose.

Serotransferrin. Relative levels of serotransferrin were determined in control and patient individuals using up to 29 unique peptides, although detection of these peptides was highly variable among individuals. This protein was strongly up-regulated in four control individuals and moderately up-regulated in six control individuals relative to patients. In addition, the LC-MS/MS data for this protein (Table 4), while semiquantitative, strongly support these results. Increased urinary transferrin excretion rates have been reported in type 1 diabetic patients with normal urinary albumin excretion rates,[42–44] suggesting that transferrin might be released preferentially from the diabetic kidney leading to lower levels in plasma. Indeed, van Campenhout and colleagues found lower levels of transferrin and total iron-binding capacity in serum of type 1 diabetic subjects relative to controls.[45]

Conclusions

The first application of global, untargeted proteomics in the study of a DASP sample subset has resulted in the identification of five candidate protein biomarkers of type 1 diabetes. However, it cannot be determined from the data obtained in this work whether the five proteins discussed above are predictive or diagnostic of type 1 diabetes, and because this study utilized a relatively low number of biological replicates ($n = 10$) for each condition, higher numbers of control and type 1 diabetic samples are required to validate these results in subsequent studies. Once validated, the above proteins can be further explored in targeted proteomic studies utilizing

isotopically labeled peptide internal standards for absolute quantitation, which will enable the determination of laboratory-defined sensitivity and specificity with blinded samples.

Acknowledgment. The authors would like to thank the DASP Standardization Committee and the members of the diabetes research community who have contributed DASP samples. This work was supported by NIH grant DK070146, while portions of the research were supported by the NIH National Center for Research Resources (RR18522). The DASP is funded at CDC by PL105-33, 106-310, 106-554, and 107-360 administered by the National Institutes of Health. Work was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility located at Pacific Northwest National Laboratory (PNNL) and sponsored by the U.S. Department of Energy (DOE) Office of Biological and Environmental Research. PNNL is a multiprogram national laboratory operated by Battelle Memorial Institute for the DOE under contract DE-AC06-76RLO-1830.

Supporting Information Available: Supplemental tables at the peptide and protein levels. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Bingley, P. J.; Bonifacio, E.; Williams, A. J.; Genovese, S.; Bottazzo, G. F.; Gale, E. A. Prediction of IDDM in the general population: strageties based on combinations of autoantibody markers. *Diabetes* **1997**, *46* (11), 1701–1710.

(2) Seissler, J.; Hatziagelaki, E.; Scherbaum, W. A. Modern concepts for the prediction of type 1 diabetes. *Exp. Clin. Endocrinol. Diabetes* **2001**, *109* (Supplement 2), S304–S316.

(3) LaGasse, J. M.; Brantley, M. S.; Leech, N. J.; Rowe, R. E.; Monks, S.; Palmer, J. P.; Nepom, G. T.; McCulloch, D. K.; Hagopian, W. A. Study, Washington State Diabetes Prediction. Successful prospective prediction of type 1 diabetes in schoolchildren through multiple defined autoantibodies: an 8-year follow-up of the Washington State Diabetes Prediction Study. *Diabetes Care* **2002**, *25* (3), 505–511.

(4) Mueller, P. W.; Bingley, P. J.; Bonifacio, E.; Steinberg, K. K.; Sampson, E. J. Predicting type 1 diabetes using autoantibodies: the latest results from the diabetes autoantibody standardization program. *Diabetes Technol. Ther.* **2002**, *4* (3), 397–400.

(5) Bingley, P. J.; Bonifacio, E.; Mueller, P. W. Diabetes Antibody Standardization Program: first assay proficiency evaluation. *Diabetes* **2003**, *52* (5), 1128–1136.

(6) Jacobs, J. M.; Adkins, J. N.; Qian, W. J.; Liu, T.; Shen, Y.; Camp, D. G., II; Smith, R. D. Utilizing human blood plasma for proteomic biomarker discovery. *J. Proteome Res.* **2005**, *4* (4), 1073–1085.

(7) Sadygov, R. G.; Cociorva, D.; Yates, J. R., III. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat. Methods* **2004**, *1* (3), 195–202.

(8) Monti, M.; Orru, S.; Pagnozzi, D.; Pucci, P. Interaction proteomics. *Biosci. Rep.* **2005**, *25* (1–2), 45–56.

(9) Pasa-Tolic, L.; Masselon, C.; Barry, R. C.; Shen, Y.; Smith, R. D. Proteomic analyses using an accurate mass and time tag strategy. *BioTechniques* **2004**, *37* (4), 621–636.

(10) Adkins, J. N.; Monroe, M. E.; Auberry, K. J.; Shen, Y.; Jacobs, J. M.; Camp, D. G., II; Vitzthum, F.; Rodland, K. D.; Zangar, R. C.; Smith, R. D.; Pounds, J. G. A proteomic study of the HUPO Plasma Proteome Project's pilot samples using an accurate mass and time tag strategy. *Proteomics* **2005**, *5* (13), 3454–3466.

(11) Zimmer, J. S.; Monroe, M. E.; Qian, W. J.; Smith, R. D. Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom. Rev.* **2006**, *25* (3), 450–482.

(12) Liu, T.; Qian, W. J.; Mottaz, H. M.; Gritsenko, M. A.; Norbeck, A. D.; Moore, R. J.; Purvine, S. O.; Camp, D. G., II; Smith, R. D. Evaluation of muli-protein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol. Cell. Proteomics* **2006**, *5* (11), 2167–2174.

(13) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Xiao, W.; Moldawer, L. L.; Kaushal, A.; Monroe, M. E.; Varnum, S. M.; Moore, R. J.; Purvine, S. O.; Maier, R. V.; Davis, R. W.; Tompkins, R. G.; Camp, D. G., II; Smith, R. D. High dynamic range characterization of the trauma patient plasma proteome. *Mol. Cell. Proteomics* **2006**, *5* (10), 1899–1913.

(14) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Camp, D. G., II; Monroe, M. E.; Moore, R. J.; Smith, R. D. Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* **2005**, *4* (6), 2070–2080.

(15) Wang, H.; Qian, W. J.; Chin, M. H.; Petyuk, V. A.; Barry, R. C.; Liu, T.; Gritsenko, M. A.; Mottaz, H. M.; Moore, R. J.; Camp, D. G., II; Khan, A. H.; Smith, D. J.; Smith, R. D. Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. *J. Proteome Res.* **2006**, *5* (2), 361–369.

(16) Shen, Y.; Zhao, R.; Belov, M. E.; Conrads, T. P.; Anderson, G. A.; Tang, K.; Pasa-Tolic, L.; Veenstra, T. D.; Lipton, M. S.; Udseth, H. R.; Smith, R. D. Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem.* **2001**, *73* (8), 1766–1775.

(17) Ding, J.; Sorensen, C. M.; Zhang, Q.; Jiang, H.; Jaitly, N.; Livesay, E. A.; Shen, Y.; Smith, R. D.; Metz, T. O. Capillary LC coupled with high-mass measurement accuracy mass spectrometry for metabolic profiling. *Anal. Chem.* **2007**, *79*, 6081–6093.

(18) Shaffer, S. A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. An ion funnel interface for improved ion focusing and sensitivity using electrospray ionization mass spectrometry. *Anal. Chem.* **1998**, *70* (19), 4111–4119.

(19) Eng, J. K.; McCormack, A. L.; Yates, J. R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(20) Petritis, K.; Kangas, L. J.; Ferguson, P. L.; Anderson, G. A.; Pasa-Tolic, L.; Lipton, M. S.; Auberry, K. J.; Strittmatter, E.; Shen, Y.; Zhao, R.; Smith, R. D. Use of artificial neural networks for the prediction of peptide liquid chromatography elution times in proteome analyses. *Anal. Chem.* **2003**, *75* (5), 1039–1048.

(21) Kiebel, G. R.; Auberry, K. J.; Jaitly, N.; Clark, D. A.; Monroe, M. E.; Peterson, E. S.; Tolic, N.; Anderson, G. A.; Smith, R. D. PRISM: a data management system for high-throughput proteomics. *Proteomics* **2006**, *6* (6), 1783–1790.

(22) Monroe, M. E.; Tolic, N.; Jaitly, N.; Shaw, J. L.; Adkins, J. N.; Smith, R. D. VIPER: an advanced software package to support high-throughput LC-MS peptide identifications. *Bioinformatics* **2007**, *23* (15), 2021–2023.

(23) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75* (17), 4646–4658.

(24) Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W. J.; Webb-Robertson, B. J.; Smith, R. D.; Lipton, M. S. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* **2006**, *5* (2), 277–286.

(25) Saeed, A. I.; Sharov, V.; White, J.; Li, J.; Liang, W.; Bhagabati, N.; Braisted, J.; Klapa, M.; Currier, T.; Thiagarajan, M.; Sturn, A.; Snuffin, M.; Rezantsev, A.; Popov, D.; Ryltsov, A.; Kostukovich, E.; Borisovsky, I.; Liu, Z.; Vinsavich, A.; Trush, V.; Quackenbush, J. TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **2003**, *34* (2), 374–378.

(26) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4* (3), 207–214.

(27) Petyuk, V. A.; Qian, W. J.; Chin, M. H.; Wang, H.; Livesay, E. A.; Monroe, M. E.; Adkins, J. N.; Jaitly, N.; Anderson, D. J.; Camp, D. G., II; Smith, D. J.; Smith, R. D. Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry. *Genome Res.* **2007**, *17* (3), 328–336.

(28) MacCoss, M. J.; Wu, C. C.; Yates, J. R., III. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **2002**, *74*, 5593–5599.

(29) Uria, J. A.; Fueyo, A.; Balbin, M.; Velasco, G.; Pendas, A. M.; Lopez-Otin, C. Alternative splicing gives rise to two novel long isoforms of Zn-alpha 2-glycoprotein, a member of the immunoglobulin superfamily. *Gene* **1996**, *169* (2), 233–236.

(30) Todorov, P. T.; McDevitt, T. M.; Meyer, D. J.; Ueyama, H.; Ohkubo, I.; Tisdale, M. J. Purification and characterization of a tumor lipid-mobilizing factor. *Cancer Res.* **1998**, *58* (11), 2353–2358.

(31) Hirai, K.; Hussey, H. J.; Barber, M. D.; Price, S. A.; Tisdale, M. J. Biological evaluation of a lipid-mobilizing factor isolated from the urine of cancer patients. *Cancer Res.* **1998**, *58* (11), 2359–2365.

(32) Kamboh, M. I.; Ferrell, R. E. Genetic studies of low-abundance human plasma proteins. I. Microheterogeneity of zinc-alpha 2-glycoprotein in biological fluids. *Biochem. Genet.* **1986**, *24* (11–12), 849–857.
(33) Jirka, M.; Blanicky, P. Polymorphism of human serum Zn-alpha 2-glycoprotein and its behaviour during ontogenesis using quantitative crossed starch gel immunoelectrophoresis. *Clin. Chim. Acta* **1980**, *103* (1), 61–66.
(34) Leskov, K. S.; Klokov, D. Y.; Li, J.; Kinsella, T. J.; Boothman, D. A. Synthesis and functional analyses of nuclear clusterin, a cell death protein. *J. Biol. Chem.* **2003**, *278* (13), 11590–11600.
(35) Trougakos, I. P.; So, A.; Jansen, B.; Gleave, M. E.; Gonos, E. S. Silencing expression of the clusterin/apolipoprotein j gene in human cancer cells using small interfering RNA induces spontaneous apoptosis, reduced growth ability, and cell sensitization to genotoxic and oxidative stress. *Cancer Res.* **2004**, *64* (5), 1834–1842.
(36) Gleave, M. E.; Miyake, H. Use of antisense oligonucleotides targeting the cytoprotective gene, clusterin, to enhance androgen- and chemo-sensitivity in prostate cancer. *World J. Urol.* **2005**, *23* (1), 38–46.
(37) Crave, J. C.; Lejeune, H.; Brebant, C.; Baret, C.; Pugeat, M. Differential effects of insulin and insulin-like growth factor I on the production of plasma steroid-binding globulins by human hepatoblastoma-derived (Hep G2) cells. *J. Clin. Endocrinol. Metab.* **1995**, *80*, 1283–1289.
(38) Fernandez-Real, J. M.; Grasa, M.; Casamitjana, R.; Pugeat, M.; Barret, C.; Ricart, W. Plasma total and glycosylated corticosteroid-binding globulin levels are associated with insulin secretion. *J. Clin. Endocrinol. Metab.* **1999**, *84* (9), 3192–3196.
(39) Schaefer, L.; Raslik, I.; Grone, H. J.; Schonherr, E.; Macakova, K.; Ugorcakova, J.; Budny, S.; Schaefer, R. M.; Kresse, H. Small proteoglycans in human diabetic nephropathy: discrepancy between glomerular expression and protein accumulation of decorin, biglycan, lumican, and fibromodulin. *FASEB J.* **2001**, *15* (3), 559–561.
(40) Ying, S.; Shiraishi, A.; Kao, C. W.; Converse, R. L.; Funderburgh, J. L.; Swiergiel, J.; Roth, M. R.; Conrad, G. W.; Kao, W. W. Characterization and expression of the mouse lumican gene. *J. Biol. Chem.* **1997**, *272* (48), 30306–30313.
(41) Silverstein, D. M.; Travis, B. R.; Thornhill, B. A.; Schurr, J. S.; Kolls, J. K.; Leung, J. C.; Chevalier, R. L. Altered expression of immune modulator and structural genes in neonatal unilateral ureteral obstruction. *Kidney Int.* **2003**, *64* (1), 25–35.
(42) Martin, P.; Walton, C.; Chapman, C.; Bodansky, H. J.; Stickland, M. H. Increased urinary excretion of transferrin in children with type 1 diabetes mellitus. *Diabetic Med.* **1990**, *7* (1), 35–40.
(43) O'Donnell, M. J.; Martin, P.; Florkowski, C. M.; Toop, M. J.; Chapman, C.; Holder, R.; Barnett, A. H. Urinary transferrin excretion in type 1 (insulin-dependent) diabetes mellitus. *Diabetic Med.* **1991**, *8* (7), 657–61.
(44) Worthley, D. L.; Harvey, N. T.; Hill, N. L.; Walsh, R. L.; Edwards, J. B.; Roberts, A. P. Urinary transferrin and albumin concentrations in patients with type 1 diabetes and normal controls: the search for the first protein lost. *Clin. Biochem.* **2001**, *34* (1), 83–5.
(45) Van Campenhout, A.; Van Campenhout, C.; Lagrou, A. R.; Moorkens, G.; De Block, C.; Manuel-y-Keenoy, B. Iron-binding antioxidant capacity is impaired in diabetes mellitus. *Free Radical Biol. Med.* **2006**, *40* (10), 1749–55.

PR700606W