

- (51) International Patent Classification:
G01N 33/68 (2006.01) C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2016/035117

(22) International Filing Date:
31 May 2016 (31.05.2016)

(25) Filing Language:
English

(26) Publication Language:
English

(30) Priority Data:
62/168,671 29 May 2015 (29.05.2015) US

(71) Applicants: CEDARS-SINAI MEDICAL CENTER [US/US]; 8700 Beverly Boulevard, Los Angeles, California 90048 (US). THE JOHNS HOPKINS UNIVERSITY [US/US]; 3400 North Charles Street, Baltimore, Maryland 21218 (US).

(72) Inventors: GROTE, Eric; 303 N. Swall Drive, Beverly Hills, California 90211 (US). FU, Qin; 230 S. Hamilton Drive, #108, Beverly Hills, California 90211 (US). VAN EYK, Jennifer; 429 N. Croft Ave., Los Angeles, California 90048 (US).

(74) Agents: VAKHARIA-RAO, Hema et al.; Nixon Peabody LLP, 300 South Grand Avenue, Suite 4100, Los Angeles, California 90071-3151 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

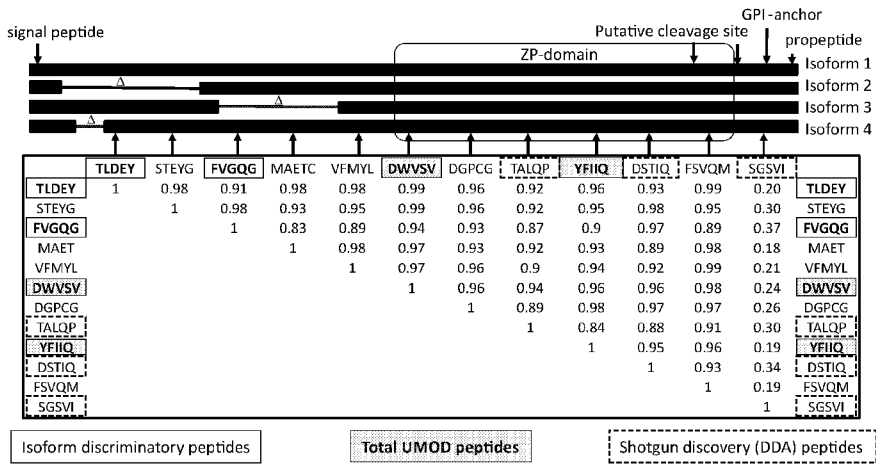
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).
- Published:

— with international search report (Art. 21(3))

— with sequence listing part of description (Rule 5.2(a))

(54) Title: CORRELATED PEPTIDES FOR QUANTITATIVE MASS SPECTROMETRY

FIG. 1B



(57) Abstract: Described herein are methods for identifying signature peptides for quantifying a polypeptide of interest in a sample. The methods include cleaving the polypeptide into peptides; detecting a multiplicity of the peptides with a quantitative analytical instrument; comparing the linearity of signals attributable to pairs of the peptides in a multiplicity of samples; and selecting signature peptides from a group of peptides with more highly correlated signals.

CORRELATED PEPTIDES FOR QUANTITATIVE MASS SPECTROMETRY

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. §119(e) to U.S. provisional patent application No. 62/168,671, filed on May 29, 2015.

GOVERNMENT RIGHTS

[0002] This invention was made with U.S. Government support under Grant No. HHSN268201000032C awarded by the National Institutes of Health. The U.S. Government has certain rights in this invention.

FIELD OF INVENTION

[0003] This invention relates to the identification of correlated signature peptides for quantification.

BACKGROUND

[0004] All publications herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference. The following description includes information that may be useful in understanding the present invention. It is not an admission that any of the information provided herein is prior art or relevant to the presently claimed invention, or that any publication specifically or implicitly referenced is prior art.

[0005] Selected reaction monitoring (SRM), also known as multiple reaction monitoring, is a quantitative mass spectrometry (MS) technique that targets predefined precursor and product ions specific to a particular analyte of interest. Proteins are typically quantified by cleaving them into peptides with a specific protease such as trypsin, measuring the concentration of one or more signature peptides, and then inferring the concentration of the parent protein.

[0006] Uromodulin was selected as an exemplary target to test SRM peptide selection workflows because of its physiological importance, biological complexity and association with disease phenotypes. Uromodulin, also known as UMOD or Tamm-Horsfall Glycoprotein, is the most abundant protein in normal human urine, but its functions remain incompletely understood. Data from genetically modified mice suggests that uromodulin protects against urinary tract infections and calcium oxalate crystals, and participates in the regulation of sodium reuptake to control blood pressure and glomerulocystic kidney disease. In these diseases, abnormal uromodulin processing leads to its accumulation in the ER.

Additionally, common uromodulin variants are associated with chronic kidney disease and hypertension, possibly via effects on salt reabsorption in the kidney. Some disease-associated variants are present at lower concentrations in urine. Exact quantitation of urinary uromodulin as a novel biomarker of susceptibility to CKD and hypertension is therefore of clinical interest and may represent a future readout to monitor blood pressure lowering treatment.

[0007] Uromodulin is well-represented in proteomic MS databases. For example, aside from a 99 amino acid N-terminal region with only one tryptic cleavage site, Peptide Atlas has MS data representing 97% of the mature protein. Nevertheless, MS analysis is complicated by the existence of four major isoforms, a variety of silent, protective, and disease-associated SNPs and mutations, and multiple glycosylation sites and disulfide bonds. In addition, urine is challenging to analyze because its pH is inconsistent between samples and there are widely varying concentrations of uromodulin, serum albumin, total protein, urea, salts, creatinine, and other metabolites.

[0008] SWATH (sequential window acquisition of all theoretical fragment ion spectra) is a new strategy for high throughput, label-free protein quantification. It generates global, quantitative protein maps using data-independent acquisition of collision-induced dissociation (CID) spectra of all precursor ions. As a data-independent acquisition (DIA) method, SWATH-MS has a greater coverage of peptide identification compared to classical discovery approaches.

[0009] Using known fingerprints of target peptides comprising precursor mass, chromatographic retention time and MRM transitions, SWATH protein maps can be interrogated for targeted quantification of proteins of interest based on high resolution MRM-like signatures. SWATH acquires all MRM transitions of all precursors and thus does not require tedious assay development and allows for a more dynamic data interpretation compared to classical MRM experiments. New proteins can be added to the list of targets during the process of data interpretation without the requirement of additional data acquisition.

[0010] How does SWATH work? The mass spectrometer does not select and isolate a specific precursor ion for CID but fragments everything within a mass window such as m/z 25 to acquire a single CID fragment-ion spectrum. To cover the full mass range between m/z 400-1250 the mass spectrometer sequentially acquires one full MS spectrum and about 34

CID-MS/MS spectra with isolation windows of m/z 25 during one cycle of roughly 3.5 seconds. Theoretically fragment ions of all precursor ions detectable throughout the selected mass range and along the chromatographic elution period are recorded. Such complex CID data however, cannot be matched to peptide sequences from databases through the commonly used search engines like Mascot, SEQUEST, ProteinPilot etc. Instead SWATH MS/MS data are searched against spectral libraries which can be generated from previous discovery data of data-dependent acquisitions.

[0011] A variety of methods have been previously used to identify signature peptides for protein quantification. One common approach is to target peptides that were identified in a data-dependent MS screen on related samples, as these peptides are guaranteed to be detectable by MS. A limitation of this approach is that discovery MS and quantitative MS are traditionally performed on different types of MS instruments with different LC systems, ionization, collision cells, and fragmentation patterns. Consequently, the dominant peptides that provide for highly confident protein identification on one instrument do not always yield sufficient MS signals for quantitation on a different instrument. In addition, long peptides (e.g. >10 aa) generally yield more MS/MS fragment ions for confident identification, whereas shorter peptides are more likely to yield a limited number of dominant fragment ions for sensitive SRM quantitation. A related approach is to target peptides found in spectral peptide libraries. Available libraries contain spectra representing many thousands of peptides collected from hundreds of MS runs, thereby facilitating the selection of target peptides and transitions that have been reproducibly observed (see e.g. <http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:start>). However, current MS spectral databases are primarily populated with data from discovery MS instruments and are therefore not directly applicable to SRM assays. SRMatlas, an online resource designed to overcome this limitation, has MS spectra from natural and synthetic peptides that were collected on a triple quadrupole mass spectrometer, the most common instrument for SRM. A pre-publication SRMatlas preview covers 99.9% of the human proteome. A third approach, in silico prediction of proteotypic peptides based solely upon a protein's amino acid sequence, provides an alternative to relying on previously acquired spectra that is especially useful for pioneering work on biological samples that have not been subjected to extensive proteomic analysis.

[0012] Peptide selection for a quantitative MS assay requires more than the mere

identification of detectable peptides. If the goal of the experiment is to quantify the total protein concentration, the selected peptides should not contain genetically encoded variations, and should not be susceptible to in vivo or in vitro post-translational modifications. On the other hand, if the goal is to monitor a specific isoform, SNP or post-translational modification, peptide selection is constrained by the need to target specific peptides that may have relatively weak MS signals and therefore require extensive optimization.

[0013] Here we demonstrate that unpredictable confounding factors can interfere with MS quantitation. Thus, selection of peptides for a robust assay requires experimental data. We present an empirical peptide selection workflow to identify surrogate peptides suitable for determining the concentration of targeted proteins in a complex biological milieu by identifying peptides with highly correlated MS signals.

SUMMARY OF THE INVENTION

[0014] The following embodiments and aspects thereof are described and illustrated in conjunction with systems, compositions and methods which are meant to be exemplary and illustrative, not limiting in scope.

[0015] Various embodiments of the present invention provide a method for identifying signature peptides for quantifying a polypeptide in a sample by selecting peptides with MS signals that are highly correlated with the MS signals of other peptides derived from the same polypeptide. In a preferred embodiment, the MS signal is a peak area. In another preferred embodiment, the MS signal is calculated by dividing the peak area of the peptide by the peak area of an SIL internal standard peptide of the same sequence. In various embodiments, the correlation between the MS signals of a pair of peptides is determined by parametric methods such as the Pearson r correlation or by nonparametric methods such as Kendall rank correlation and Spearman rank correlation. In a preferred embodiment, correlations are measured by determining the coefficient of determination (r^2).

[0016] Various embodiments of the present invention provide a method of identifying signature fragments for quantifying a macromolecule in a sample. The method may comprise: acquiring mass spectrometry (MS) data on multiple candidate fragments of the macromolecule from multiple samples; using the MS data to calculate correlation values for pairwise comparisons between each of the multiple candidate fragments; and identifying the highly correlated fragments among the multiple candidate fragments as the signature fragments for quantifying the macromolecule. In some embodiments, the macromolecule is a

polypeptide. In some embodiments, the macromolecule is a nucleic acid. In some embodiments, the macromolecule is a polysaccharide. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0017] Various embodiments of the present invention provide a method of identifying signature peptides for quantifying a polypeptide in a sample. The method may comprise: acquiring mass spectrometry (MS) data on multiple candidate peptides derived from the polypeptide in multiple samples; using the MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0018] In some embodiments, the MS data is acquired through targeted acquisition methods such as Selective Reaction Monitoring (SRM) and Multiple Reaction Monitoring (MRM). In other embodiments, the MS data is acquired through data-independent acquisition methods such as SWATH. In various embodiments, the MS data is SRM data and/or MRM data. In various embodiments, the MS data is SWATH MS data, Shotgun CID MS data, Original DIA MS Data, MSE MS data, p2CID MS Data, PAcIFIC MS Data, AIF MS Data, XDLA MS Data, or FT-ARM MS Data, or a combination thereof. In various embodiments, the MS data comprises raw MS data obtained from a mass spectrometer and/or processed MS data in which peptides and their fragments (e.g., transitions and MS peaks) are already identified, analyzed and/or quantified.

[0019] Various embodiments of the present invention provide a method of quantifying a polypeptide in a sample. The method may comprise: cleaving the polypeptide to yield one or more signature peptide identified according to a method as described herein; analyzing the sample on a mass spectrometer; detecting MS signals of the signature peptide; and quantifying the polypeptide based on the detected MS signals. In some embodiments, multiple polypeptides in a complex sample are quantified.

[0020] Various embodiments of the present invention provide a kit for quantifying a polypeptide in a sample. The kit comprises an internal standard of a signature peptide identified for the polypeptide according to a method as described herein; and instructions for using the internal standard to quantify the polypeptide in the sample. In some embodiments, the kit targets a single polypeptide. In other embodiments, the kit targets multiple

polypeptides (multiplexing). In various embodiments, the kit further comprises a protease for cleaving the polypeptide to yield the signature peptide. In various embodiments, the kit further comprises an antibody specifically binding to the signature peptide. In certain embodiments, such a kit can be used for SISCAPA. In some embodiments, the kit comprises multiple internal standards. In some embodiments, the kit quantifies multiple polypeptides in a complex sample.

[0021] Various embodiments of the present invention provide a system for identifying signature peptides for quantifying a polypeptide. The system may comprises: a mass spectrometer configured for acquiring mass spectrometry (MS) data on multiple candidate peptides derived from the polypeptide in multiple samples; and a computer configured for using the MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide, wherein the mass spectrometer and the computer are connected via a communication link. In some embodiments, the computer is configured for processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0022] Various embodiments of the present invention provide a non-transitory computer-readable storage medium, wherein the non-transitory computer-readable storage medium is configured for storing a program, wherein the program is configured for execution by a processor of a computer, and wherein the program comprises instructions for using mass spectrometry (MS) data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0023] Various embodiments of the present invention provide a computer. The computer may comprises: a memory configured for storing a program; and a processor configured for executing the program, wherein the program comprises instructions for using mass spectrometry (MS) data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the

highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. Various embodiments of the present invention provide a computer implemented method. The method may comprise: providing a computer as described herein; inputting mass spectrometry (MS) data into the computer; and operating the computer to use the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0024] Various embodiments of the present invention provide a non-transitory computer-readable storage medium, wherein the non-transitory computer-readable storage medium is configured for storing a program, wherein the program is configured for execution by a processor of a computer, and wherein the program comprises instructions for operating a mass spectrometer to acquire mass spectrometry (MS) data, for using the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0025] Various embodiments of the present invention provide a computer. The computer comprises: a memory configured for storing a program; and a processor configured for executing the program, wherein the program comprises instructions for operating a mass spectrometer to acquire mass spectrometry (MS) data, for using the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. Various embodiments of the present invention provide a computer implemented method. The method comprises: providing a computer as described herein; connecting the computer via a communication link to a mass spectrometer; and operating the computer to operate the mass spectrometer to acquire mass spectrometry (MS) data, to use the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the

multiple candidate peptides as the signature peptides for quantifying the polypeptide. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0026] Various embodiments of the present invention provide a non-transitory computer-readable storage medium, wherein the non-transitory computer-readable storage medium is configured for storing a program, wherein the program is configured for execution by a processor of a computer, and wherein the program comprises instructions for processing MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and for quantify the polypeptide based on the signature peptide.

[0027] Various embodiments of the present invention provide a computer, comprising: a memory configured for storing a program; and a processor configured for executing the program, wherein the program comprises instructions for processing MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and for quantify the polypeptide based on the signature peptide. Various embodiments of the present invention provide a computer implemented method, comprising: providing a computer as described herein; inputting MS data into the computer; and operating the computer to process MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and to quantify the polypeptide based on the signature peptide.

[0028] Various embodiments of the present invention provide a non-transitory computer-readable storage medium, wherein the non-transitory computer-readable storage medium is configured for storing a program, wherein the program is configured for execution by a processor of a computer, and wherein the program comprises instructions for operating a mass spectrometer to detect MS signals of a signature peptide for quantifying a polypeptide, and quantifying the polypeptide based on the detected MS signals.

[0029] Various embodiments of the present invention provide a computer. The computer may comprise: a memory configured for storing a program; and a processor configured for executing the program, wherein the program comprises instructions for operating a mass spectrometer to detect MS signals of a signature peptide for quantifying a polypeptide, and quantifying the polypeptide based on the detected MS signals. Various embodiments of the present invention provide a computer implemented method. The method may comprise: providing a computer as described herein; connecting the computer via a communication link to a mass spectrometer; and operating the computer to operate the mass spectrometer to detect MS signals of a signature peptide for quantifying a polypeptide, and to quantify the

polypeptide based on the detected MS signals.

[0030] Various embodiments of the present invention provide a method of producing an antibody. The method comprises: providing a signature peptide identified according to a method as described herein; and immunizing an animal using the signature peptide, thereby producing the antibody. In various embodiments, the method further comprises isolating and/or purifying the antibody from the immunized animal.

[0031] Various embodiments of the present invention provide an antibody specifically binding to a signature peptide identified according to a method as described herein, or an antigen-binding fragment thereof.

[0032] Various embodiments of the present invention provide a method of quantifying a polypeptide in a sample. The method may comprise: contacting the sample with an antibody as described herein or an antigen-binding fragment thereof; detecting the binding between the polypeptide and the antibody or the antigen-binding fragment thereof; and quantifying the polypeptide based on the detected binding.

[0033] Various embodiments of the present invention provide a kit quantifying a polypeptide in a sample. The kit comprises: an antibody specifically binding to a signature peptide identified according to a method as described herein; and instructions for using the antibody to quantify the polypeptide in the sample.

[0034] In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

BRIEF DESCRIPTION OF FIGURES

[0035] Exemplary embodiments are illustrated in referenced figures. It is intended that the embodiments and figures disclosed herein are to be considered illustrative rather than restrictive.

[0036] Figure 1A depicts, in accordance with various embodiments of the present invention, amino acid sequence features of uromodulin 1. Candidate tryptic peptides of 6-21 amino acids include two signature peptides reporting the concentration of total uromodulin (thin outline), two signature peptides that discriminate between uromodulin isoforms (bold outline), three peptides identified by data dependent acquisition that were found to have nonlinear responses (thin dashed outline), and six other peptides included in the correlation matrix (bold dashed outline). Potential posttranslational modifications include N-linked

glycosylation (bold font surround with gray box), disulfide bonds (hollow font), and methionine oxidation (bold font).

[0037] Figure 1B depicts, in accordance with various embodiments of the present invention, a coefficients of determination (r^2) matrix for uromodulin. The schema at top presents structural features of the 4 uromodulin isoforms and identifies the location of 12 candidate peptides, which are identified by their first 5 amino acids. To empirically identify signature peptides that can accurately report the concentration of uromodulin protein, each peptide was individually compared with every other peptide for a total of 72 ($12 \times 12 / 2$) comparisons. For each peptide pair, a plot was constructed using SRM measurements from 9 urine samples. Values for the area under the curve for one peptide were plotted on the x axis and values for the area under the curve for the other peptide were plotted on the y axis. A line was fit to the 9 data points, and a coefficient of determination (r^2) was calculated and entered into the matrix.

[0038] Figures 2A-2C depict, in accordance with various embodiments of the present invention, that absolute quantification of uromodulin is reproducible. Four uromodulin peptides were quantified by SRM in 40 urine samples using SIL internal standards for normalization to a standard curve. For presentation, the samples are arranged according to the concentration of the DWVSV peptide. Absolute concentration ($\mu\text{g/ml}$) and reproducibility (%CV) are compared between (Figure 2A) LC-MS injections ($n=3$) for quantitation of the DWVSV-y7 transition in each digest, (Figure 2B) Trypsin digests ($n=3$), and (Figure 2C) different SRM transitions ($n=2, 3$, or 4) for the same peptide. See Table 2 for a list of transitions for each peptide.

[0039] Figure 3 depicts, in accordance with various embodiments of the present invention, that SRM quantification of the 4 empirically selected uromodulin peptides is internally consistent and correlates with ELISA results. Normalized SRM and ELISA data from 40 urine samples are presented as a correlation matrix.

[0040] Figure 4 depicts, in accordance with various embodiments of the present invention, a proposed workflow for empirical peptide selection.

[0041] Figure 5 depicts, in accordance with various embodiments of the present invention, a sample processing workflow highlighting the order of reagent addition and each step where conditions were optimized.

[0042] Figure 6 depicts, in accordance with various embodiments of the present invention,

that some trypsin-sensitive peptides have low SRM correlations. For each peptide, an average SRM correlation was calculated from the coefficients of variation presented in Figure 1B. Trypsin resistance was defined as the ratio of the SRM signal from a digest with 4 μ l trypsin compared to the signal from a digest with 1 μ l trypsin. Trypsin-sensitive peptides had a low score because digestion was complete with 1 μ l trypsin.

[0043] Figure 7 depicts, in accordance with various embodiments of the present invention, that SRM can distinguish between uromodulin isoforms. Uromodulin purified from urine by Millipore (M) and Prospec Bio (P) was compared with recombinant uromodulin-3 (Abnova). A trypsin digest of each protein was analyzed with an SRM assay targeting 11 uromodulin-derived peptides. To normalize the results for each target peptide, raw SRM area-under-the-curve data was divided by the average signal for those samples with detectable peptide.

[0044] Figure 8 depicts, in accordance with various embodiments of the present invention, variability in methionine oxidation. Native and oxidized forms of four uromodulin peptides were quantified by comparing equivalent transitions from raw SRM (area under the curve) data. The urine specimens included pooled normal urine from a -80°C stock, with and without thawing and storage at -20°C for one month, and seven randomly selected clinical urine specimens.

[0045] Figure 9 depicts, in accordance with various embodiments of the present invention, normalization with SIL internal standards. Pooled urine was spiked with a mixture of SIL peptide standards, digested with trypsin, and then divided into aliquots that were desalted on different wells of an HLB microplate. The desalting conditions were altered by varying the total amount of urine protein applied, the number of times each aliquot was passed through the HLB resin, the volume of elution buffer, the number of times the elution buffer was passed through the HLB resin, and the flow rate during elution. Each eluate was dried, resuspended in MS buffer, and then analyzed with an SRM assay targeting the four empirically selected uromodulin peptides and two peptides from human serum albumin. The resuspension volume was adjusted to compensate for differences in the amounts of input peptides. Upper panel: Raw area-under-the-curve data; Lower panel: normalized data calculated by dividing the signal from native peptides by data from the corresponding SIL peptide standard. To compensate for differences between the SRM response for different peptides, all data was divided by the average signal for the corresponding peptide.

[0046] Figure 10 depicts, in accordance with various embodiments of the present invention,

linearity and range of the SRM assay. Purified uromodulin was digested with trypsin, desalted on HLB resin, and resuspended in MS loading buffer supplemented with a mixture of SIL peptides. Serial dilutions were prepared in supplemented loading buffer and then analyzed by SRM. Data is presented for a representative transition reporting on the y7 fragment of the DWVSV peptide.

[0047] Figure 11 depicts, in accordance with various embodiments of the present invention, a selection of surfactants. Pooled human urine was supplemented with various surfactants and then reduced, alkylated, and digested with trypsin. The resulting peptides were desalted on an HLB plate and analyzed by SRM. Data is presented for a representative transition targeting the y10 fragment of the DSTIQVVENGESSQGR peptide.

[0048] Figures 12A-12B depict, in accordance with various embodiments of the present invention, peptide desalting on HLB resin. Figure 12A: SIL peptides (100 fmol/ μ l) were desalted on C18 or C4 OMIX pipet tips or on WCX or HLB Oasis microplates. Recovery was calculated by comparing SRM peak areas before and after desalting. Figure 12B: Various concentrations of SIL peptides in 50 μ l of trypsin-digested urine were desalted on an HLB plate.

[0049] Figure 13 depicts, in accordance with various embodiments of the present invention, a schematic of general workflow for SWATH-MS acquisition and analysis.

[0050] Figure 14 depicts, in accordance with various embodiments of the present invention, an example of TOF MS parameters for TripleTOF MS instruments.

[0051] Figure 15 depicts, in accordance with various embodiments of the present invention, an example of Switch Criteria parameters for TripleTOF MS instruments.

[0052] Figure 16 depicts, in accordance with various embodiments of the present invention, schematic for importing ion library into PeakView software.

[0053] Figure 17 depicts, in accordance with various embodiments of the present invention, example of typical processing settings for SWATH analysis using PeakView software.

[0054] Figure 18 depicts, in accordance with various embodiments of the present invention, schematic for exporting SWATH results from PeakView software.

DESCRIPTION OF THE INVENTION

[0055] All references cited herein are incorporated by reference in their entirety as though fully set forth. Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this

invention belongs. Allen *et al.*, *Remington: The Science and Practice of Pharmacy* 22nd ed., Pharmaceutical Press (September 15, 2012); Hornyak *et al.*, *Introduction to Nanoscience and Nanotechnology*, CRC Press (2008); Singleton and Sainsbury, *Dictionary of Microbiology and Molecular Biology* 3rd ed., revised ed., J. Wiley & Sons (New York, NY 2006); Smith, *March's Advanced Organic Chemistry Reactions, Mechanisms and Structure* 7th ed., J. Wiley & Sons (New York, NY 2013); Singleton, *Dictionary of DNA and Genome Technology* 3rd ed., Wiley-Blackwell (November 28, 2012); and Green and Sambrook, *Molecular Cloning: A Laboratory Manual* 4th ed., Cold Spring Harbor Laboratory Press (Cold Spring Harbor, NY 2012), provide one skilled in the art with a general guide to many of the terms used in the present application.

[0056] For references on mass spectrometry and proteomics, see e.g., Salvatore Sechi, *Quantitative Proteomics by Mass Spectrometry (Methods in Molecular Biology)* 2nd ed. 2016 Edition, Humana Press (New York, NY, 2009); Daniel Martins-de-Souza, *Shotgun Proteomics: Methods and Protocols* 2014 edition, Humana Press (New York, NY, 2014); Jörg Reinders and Albert Sickmann, *Proteomics: Methods and Protocols (Methods in Molecular Biology)* 2009 edition, Humana Press (New York, NY, 2009); and Jörg Reinders, *Proteomics in Systems Biology: Methods and Protocols (Methods in Molecular Biology)* 1st ed. 2016 edition, Humana Press (New York, NY, 2009).

[0057] For references on how to prepare antibodies, see e.g., Greenfield, *Antibodies A Laboratory Manual* 2nd ed., Cold Spring Harbor Press (Cold Spring Harbor NY, 2013); Köhler and Milstein, *Derivation of specific antibody-producing tissue culture and tumor lines by cell fusion*, Eur. J. Immunol. 1976 Jul, 6(7):511-9; Queen and Selick, *Humanized immunoglobulins*, U. S. Patent No. 5,585,089 (1996 Dec); and Riechmann *et al.*, *Reshaping human antibodies for therapy*, Nature 1988 Mar 24, 332(6162):323-7.

[0058] One skilled in the art will recognize many methods and materials similar or equivalent to those described herein, which could be used in the practice of the present invention. Other features and advantages of the invention will become apparent from the following detailed description, taken in conjunction with the accompanying drawings, which illustrate, by way of example, various features of embodiments of the invention. Indeed, the present invention is in no way limited to the methods and materials described. For convenience, certain terms employed herein, in the specification, examples and appended claims are collected here.

[0059] Unless stated otherwise, or implicit from context, the following terms and phrases

include the meanings provided below. Unless explicitly stated otherwise, or apparent from context, the terms and phrases below do not exclude the meaning that the term or phrase has acquired in the art to which it pertains. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. It should be understood that this invention is not limited to the particular methodology, protocols, and reagents, etc., described herein and as such can vary. The definitions and terminology used herein are provided to aid in describing particular embodiments, and are not intended to limit the claimed invention, because the scope of the invention is limited only by the claims.

[0060] As used herein the term “comprising” or “comprises” is used in reference to compositions, methods, and respective component(s) thereof, that are useful to an embodiment, yet open to the inclusion of unspecified elements, whether useful or not. It will be understood by those within the art that, in general, terms used herein are generally intended as “open” terms (e.g., the term “including” should be interpreted as “including but not limited to,” the term “having” should be interpreted as “having at least,” the term “includes” should be interpreted as “includes but is not limited to,” etc.). Although the open-ended term “comprising,” as a synonym of terms such as including, containing, or having, is used herein to describe and claim the invention, the present invention, or embodiments thereof, may alternatively be described using alternative terms such as “consisting of” or “consisting essentially of.”

[0061] Unless stated otherwise, the terms “a” and “an” and “the” and similar references used in the context of describing a particular embodiment of the application (especially in the context of claims) can be construed to cover both the singular and the plural. The recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (for example, “such as”) provided with respect to certain embodiments herein is intended merely to better illuminate the application and does not pose a limitation on the scope of the application otherwise claimed. The abbreviation, “e.g.” is derived from the Latin *exempli gratia*, and is used herein to indicate a non-limiting

example. Thus, the abbreviation “e.g.” is synonymous with the term “for example.” No language in the specification should be construed as indicating any non-claimed element essential to the practice of the application.

[0062] The term “sample” or “biological sample” as used herein denotes a sample taken or isolated from a biological organism, e.g., a tumor sample from a subject. Exemplary biological samples include, but are not limited to, cheek swab; mucus; whole blood, blood, serum; plasma; urine; saliva; semen; lymph; fecal extract; sputum; other body fluid or biofluid; cell sample; tissue sample; tumor sample; and/or tumor biopsy etc. The term also includes a mixture of the above-mentioned samples. The term “sample” also includes untreated or pretreated (or pre-processed) biological samples. In some embodiments, a sample can comprise one or more cells from the subject. In some embodiments, a sample can be a tumor cell sample, e.g. the sample can comprise cancerous cells, cells from a tumor, and/or a tumor biopsy.

[0063] As used herein, a “subject” means a human or animal. Usually the animal is a vertebrate such as a primate, rodent, domestic animal or game animal. Primates include chimpanzees, cynomolgous monkeys, spider monkeys, and macaques, e.g., Rhesus. Rodents include mice, rats, woodchucks, ferrets, rabbits and hamsters. Domestic and game animals include cows, horses, pigs, deer, bison, buffalo, feline species, e.g., domestic cat, and canine species, e.g., dog, fox, wolf. The terms, “patient”, “individual” and “subject” are used interchangeably herein. In an embodiment, the subject is mammal. The mammal can be a human, non-human primate, mouse, rat, dog, cat, horse, or cow, but are not limited to these examples. In addition, the methods described herein can be used to treat domesticated animals and/or pets.

[0064] “Mammal” as used herein refers to any member of the class Mammalia, including, without limitation, humans and nonhuman primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs, and the like. The term does not denote a particular age or sex. Thus, adult and newborn subjects, as well as fetuses, whether male or female, are intended to be included within the scope of this term.

[0065] As used herein, SRM stands for selected reaction monitoring. As used herein, MRM stands for multiple reaction monitoring. As used herein, SWATH stands for sequential

window acquisition of all theoretical fragment ion spectra. As used herein, DIA stands for data-independent analysis. As used herein, MS stands for mass spectrometry. As used herein, ARIC stands for atherosclerosis risk in communities. As used herein, PDAY stands for Pathobiological Determinants of Atherosclerosis in Youth. As used herein, PTM stands for post-translational modifications. As used herein, SIL stands for stable isotope-labeled.

[0066] As used herein, “MS data” can be raw MS data obtained from a mass spectrometer and/or processed MS data in which peptides and their fragments (e.g., transitions and MS peaks) are already identified, analyzed and/or quantified. MS data can be Selective Reaction Monitoring (SRM) data, Multiple Reaction Monitoring (MRM) data, Shotgun CID MS data, Original DIA MS Data, MSE MS data, p2CID MS Data, PAcIFIC MS Data, AIF MS Data, XDLA MS Data, SWATH MS data, or FT-ARM MS Data, or their combinations.

[0067] As used herein, “acquiring MS data” can be accomplished without operating a mass spectrometer (for example, through retrieving results from MS experiments run previously and/or MS databases), or can be accomplished through operating a mass spectrometer to run MS experiments on samples.

[0068] As used herein, a pairwise correlation matrix refers to a matrix in which multiple candidate peptides are placed on a top (or bottom) row and a left (or right) column in the same order, and correlation values for each pair of candidate peptides are placed at their column-row intersections. The multiple candidate peptides can be derived from a single polypeptide or multiple polypeptides (for examples, protein isoforms, variants, or a family of related proteins). In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0069] As used herein, the terms “correlation”, “correlation value” and “correlation coefficient” can be used interchangeably to refer to any statistical measure that indicates the extent to which two or more variables fluctuate together. Non-limiting examples of “correlation value” include parametric methods such as the Pearson correlation coefficient; and nonparametric methods such as Kendall rank correlation coefficient and Spearman rank correlation coefficient. In preferred embodiments of the present invention, the “correlation value” is a coefficient of determination (r^2) value.

[0070] This approach of the present invention, based on SRM and/or SWATH MS, allows for the detection and accurate quantification of specific peptides in complex mixtures.

[0071] Selected Reaction Monitoring or Multiple Reaction Monitoring (SRM/MRM) mass

spectrometry is a technology with the potential for reliable and comprehensive quantification of substances of low abundance in complex samples. SRM is performed on triple quadrupole-like instruments, in which increased selectivity is obtained through collision-induced dissociation. It is a non-scanning mass spectrometry technique, where two mass analyzers are used as static mass filters, to monitor a particular fragment of a selected precursor. The specific pair of mass-over-charge (m/z) values associated to the precursor and fragment ions selected is referred to as a “transition”. The detector acts as a counting device for the ions matching the selected transition thereby returning an intensity distribution over time. MRM is when multiple SRM transitions are measured within the same experiment on the chromatographic time scale by rapidly switching between the different precursor/fragment pairs. Typically, the triple quadrupole instrument cycles through a series of transitions and records the signal of each transition as a function of the elution time. The method allows for additional selectivity by monitoring the chromatographic co-elution of multiple transitions for a given analyte.

[0072] SWATH MS a data independent acquisition (DIA) method which aims to complement traditional mass spectrometry-based proteomics techniques such as shotgun and SRM methods. In essence, it allows a complete and permanent recording of all fragment ions of the detectable peptide precursors present in a biological sample. It thus combines the advantages of shotgun (high throughput) with those of SRM (high reproducibility and consistency).

[0073] In a preferred embodiment, the developed assays can be applied to the quantification of polypeptides(s) in biological sample(s). Any kind of biological samples comprising polypeptides can be the starting point and be analyzed in the above procedure. Indeed any protein/peptide containing sample can be used for and analyzed by the assays produced here (cells, tissues, body fluids, waters, food, terrain, synthetic preparations, etc.). The assays can also be used with peptide mixtures obtain by digestion or with any non-digested sample. Digestion of a polypeptide includes any kind of cleavage strategies, such as, enzymatic, chemical, physical or combinations thereof.

[0074] The deciding factors of which polypeptide will be the one of interest varies. It can be decided by performing a literature search and identifying proteins that are functionally related, are candidate protein biomarkers which can be used in screening for drug discovery, biomarker discovery and/or disease clinical phase trials or are diagnostic markers to screen

for pharmaceutical/medical purposes. The polypeptide of interest may be determined by experimental analysis. The selection of the polypeptides is done at the beginning, and used in the invention to develop assays to specifically monitor quantitatively the set of polypeptides in samples of interest.

[0075] According to a preferred embodiment, the following parameters of the assay are determined: trypsin digestion and peptide clean up, best responding polypeptides, best responding fragments, fragment intensity ratios (increased high and reproducible peak intensities), optimal collision energies, and all the optimal parameters to maximize sensitivity and/or specificity of the assays.

[0076] In another preferred embodiment, quantification of the polypeptides and/or of the corresponding proteins or activity/regulation of the corresponding proteins is desired. A selected peptide is labeled with a stable-isotope and used as an internal standard to achieve absolute quantification of a protein of interest. The addition of a quantified stable-labeled peptide analogue of the tag to the peptide sample in known amount; and subsequently the tag and the peptide of interest is quantified by mass spectrometry and absolute quantification of the endogenous levels of the proteins is obtained.

[0077] According to a preferred embodiment, the analysis and/or comparison is done on protein samples of wild-type or physiological/healthy origin with protein samples of mutant or pathological origin.

[0078] The present invention supports the use of SRM and SWATH as platform and uses a correlation matrix to identify signature polypeptides for quantitative proteomics. The approach is applicable to the analysis of proteins from all organisms, from cells, organs, body fluids, and in the context of in vivo and/or in vitro analyses. Examples of applications of the invention include the development, use and commercialization of quantitative assays for sets of polypeptides of interest. The invention can be beneficial for the pharmaceutical industry (e.g. drug development and assessment), the biotechnology industry (e.g. assay design and development and quality control), and in clinical applications (e.g. identification of biomarkers of disease and quantitative analysis for diagnostic, prognostic and/or therapeutic use). The invention can also be applied to water, drink, food and food ingredient testing, for example, quantifying nutrients, contaminants, toxins, antibiotics, steroids, hormones, pathogens, and allergens in water, drinks, foods and food ingredients.

Methods of the Invention

[0079] Various embodiments of the present invention provide for a method for identifying signature peptides for quantifying a polypeptide of interest in a sample. The methods include cleaving the polypeptide into peptides; detecting a multiplicity of the peptides with a quantitative analytical instrument; comparing the linearity of signals attributable to pairs of the peptides in a multiplicity of samples; and selecting signature peptides from a group of peptides with more highly correlated signals. In some embodiments, the quantitative analytical instrument is a mass spectrometer configured for selected reaction monitoring. In other exemplary embodiments, the mass spectrometer is a Triple-Time Of Flight (Triple-TOF) mass spectrometer configured for SWATH.

[0080] In various embodiments, the samples are biological samples or complex biological samples. In exemplary embodiments, the complex samples include, but are not limited to urine, blood fractions, tissues and/or tissue extracts, cells, body fluids, waters, food, terrain and/or synthetic preparations.

[0081] In some embodiments, coefficients of determination are calculated to quantify the linearity of the signals attributable to pairs of peptides in the multiplicity of samples.

[0082] In various embodiments, the peptides are derived by proteolysis or chemical cleavage of the polypeptide. In an embodiment, a protease is utilized to cleave the polypeptide into peptides. For example, the protease is trypsin. In additional embodiments, other proteases or cleavage agents may be used including but not limited to chymotrypsin, endoproteinase Lys-C, endoproteinase Asp-N, pepsin, thermolysin, papain, proteinase K, subtilisin, clostripain, exopeptidase, carboxypeptidase, cathepsin C, cyanogen bromide, formic acid, hydroxylamine, NTCB, or a combination thereof.

[0083] In various other embodiments, a list of candidate peptides to be targeted for detection on the analytical instrument is generated by modeling protein cleavage. In exemplary embodiments, a list of candidate peptides to be targeted for detection on the analytical instrument is generated by modeling trypsin digestion of the polypeptide. In some embodiments, the list of candidate peptides is narrowed by eliminating peptides that, for example, cannot be detected on the analytical instrument. In some embodiments, a list of candidate peptides is narrowed by eliminating: a peptide that has not been previously detected on a mass spectrometer, a peptide susceptible to a modification that interferes with accurate quantitation, a miscleaved peptide comprising an internal protease recognition site, a

peptide with relatively inaccessible ends evidenced by the presence of miscleaved peptides, a peptide that is not unique to the sequence of the protein of interest, a peptide not present in the mature protein, or a combination thereof.

[0084] In an embodiment, the detection of a peptide is improved by changing the conditions for fragmenting that peptide prior to detecting a multiplicity of the peptides with the mass spectrometer. In exemplary embodiments, the fragmentation condition is the collision energy.

[0085] In some embodiments, the selected signature peptides (i) have higher intensity signals than non-selected peptides in the group of peptides with correlated highly correlated signals, (ii) have signals that can be robustly detected above background noise and contaminants, and/or (iii) can discriminate between forms of the protein of interest and/or a combination thereof.

[0086] In various other embodiments, the method further comprises adding a stable isotope-labeled peptide to the sample prior to mass spectrometry. In some embodiments, the absolute amount of a peptide in the sample is determined by comparing the MS signals of natural and stable isotope-labeled peptides.

[0087] Various other embodiments of the present invention also provide a method for identifying signature fragments for quantifying a macromolecule of interest in a sample. The method includes cleaving the macromolecule into fragments; detecting a multiplicity of the fragments with a quantitative analytical instrument; comparing the linearity of signals attributable to pairs of the fragments in a multiplicity of samples; and selecting signature fragments from a group of fragments with more highly correlated signals.

[0088] Various embodiments of the present invention provide for a method for identifying signature peptides for quantifying a polypeptide of interest comprising: identifying one or more polypeptides of interest; establishing a list of candidate peptides *in silico*; digesting the polypeptide of interest with a protease to obtain a mixture of peptides; analyzing the mixture of peptides on a mass spectrometer to identify transitions with high and reproducible peak intensities; optimizing collision energy for each transition with high and reproducible peak intensities; using the optimized parameters to assay a digested complex sample using mass spectrophotometry; calculating correlation values for pairs of target peptide; determining correlated signature peptides that have high coefficients of determination; and quantitatively assessing the signature peptides in varying experimental situations. In other embodiments, optimization is performed when the signal is marginal and not performed if the signal is

strong. In another embodiment, multiple complex samples are digested so that there are enough points on the graph to compare the signals between a pair of peptides to make a linear fit. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0089] In various other embodiments, the lengths of the lengths of the peptides are within the range of 6 and 21 amino acids.

[0090] In other embodiments, the comprehensive list of candidate peptides is narrowed by eliminating peptides. In other embodiments, conventional criteria are used to eliminate peptides from the comprehensive list of candidate peptides by eliminating peptides that: (i) were never detected by MS on any instrument, (ii) are not unique to the sequence of the protein of interest, (iii) are not located within the mature protein, (iv) contain amino acid residues such as methionine, cysteine, and/or asparagine that are subjected to posttranslational modifications that interfere with accurate quantitation by mass spectrometry, (v) are miscleaved or partially cleaved, (vi) are post-translationally modified in vivo, (vii) and/or a combination thereof.

[0091] In various other embodiments, transitions for each peptide with high and reproducible peak intensities are identified. In other embodiments, the collision energy for each transition is optimized. In other embodiments, mass spectrometry comprises selected reaction monitoring (SRM), also known as multiple reaction monitoring (MRM). In other embodiments, SRM or MRM is performed on a triple quadrupole mass spectrometer. In other embodiments, the peptides uniquely associated with the polypeptide of interest are those with high correlations, strong signals, high signal/noise and/or sequences unique to the protein of interest.

[0092] In various other embodiments, an average is calculated from the coefficients of determination for each peptide in a correlation matrix. Signature peptides are then selected from among those peptides with the highest 30%, 40%, 50%, 60%, 70%, 80% or 90% of averages.

[0093] In various other embodiments, a subset of correlated peptides is selected from among the set of peptides in a correlation matrix. Members of the subset all have coefficients of determination of more than 0.60, 0.65, 0.70, 0.75, 0.76, 0.77, 0.78, 0.79, 0.80, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 or 0.99 for pairwise combinations with all other members of the subset. Signature peptides are then selected from the subset of correlated peptides.

[0094] In various other embodiments, stable isotope-labeled peptide standards for absolute quantification are used. In other embodiments, the peptide labeled with a stable isotope is used as an internal standard to obtain absolute quantification of the polypeptide of interest. In other embodiments, the peptides are quantified and then the amount of the parent protein present is inferred before digesting the sample with trypsin. In other embodiments, MS responses are used to determine an upper limit of quantification (ULOQ) and a lower limit of quantification (LLOQ).

[0095] Various embodiments of the present invention provide a method of identifying signature fragments for quantifying a macromolecule in a sample. The method comprises: acquiring mass spectrometry (MS) data on multiple candidate fragments of the macromolecule from multiple samples; using the MS data to calculate correlation values for pairwise comparisons between each of the multiple candidate fragments; and identifying the highly correlated fragments among the multiple candidate fragments as the signature fragments for quantifying the macromolecule. In some embodiments, the macromolecule is a polysaccharide. In some embodiments, the macromolecule is a nucleic acid such as DNA and RNA. In some embodiments, the macromolecule is a polypeptide or protein. In some embodiments, the macromolecule is a glycopeptide. In some embodiments, the macromolecule is a metabolic intermediate. In various embodiments, the multiple candidate peptides are derived by proteolysis or chemical cleavage of the polypeptide. In various embodiments, the macromolecule is digested with an enzyme or chemical to yield the multiple candidate fragments. In some embodiments, the enzyme is a nuclease. In some embodiments, the enzyme is a protease. In certain embodiments, the protease is trypsin. In various embodiments, the MS data comprises raw MS data obtained from a mass spectrometer and/or processed MS data in which peptides and their fragments (e.g., transitions and MS peaks) are already identified, analyzed and/or quantified. In various embodiments, the MS data is Selective Reaction Monitoring (SRM) data and/or Multiple Reaction Monitoring (MRM) data. In various embodiments, the MS data is Shotgun CID MS data, Original DIA MS Data, MSE MS data, p2CID MS Data, PAcIFIC MS Data, AIF MS Data, XDLA MS Data, SWATH MS data, or FT-ARM MS Data, or a combination thereof. In some embodiments, the method further comprising processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g.,

transitions and MS peaks) before calculating correlation values. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0096] Various embodiments of the present invention provide a method of identifying signature peptides for quantifying a polypeptide in a sample. The method comprises: acquiring mass spectrometry (MS) data on multiple candidate peptides derived from the polypeptide in multiple samples; using the MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In various embodiments, the multiple candidate peptides are derived by proteolysis or chemical cleavage of the polypeptide. In various embodiments, the polypeptide is digested with an enzyme or chemical to yield the multiple candidate fragments. In various embodiments, the MS data comprises raw MS data obtained from a mass spectrometer and/or processed MS data in which peptides and their fragments (e.g., transitions and MS peaks) are already identified, analyzed and/or quantified. In various embodiments, the MS data is Selective Reaction Monitoring (SRM) data and/or Multiple Reaction Monitoring (MRM) data. In various embodiments, the MS data is Shotgun CID MS data, Original DIA MS Data, MSE MS data, p2CID MS Data, PACIFIC MS Data, AIF MS Data, XDLA MS Data, SWATH MS data, or FT-ARM MS Data, or a combination thereof. In some embodiments, the method further comprising processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0097] Various embodiments of the present invention provide a method for identifying signature peptides for quantifying a polypeptide in a sample by selecting peptides with MS signals that are highly correlated with the MS signals of other peptides derived from the same polypeptide. In a preferred embodiment, the MS signal is a peak area. In another preferred embodiment, the MS signal is calculated by dividing the peak area of the peptide by the peak area of an SIL internal standard peptide of the same sequence. In various embodiments, the correlation between the MS signals of a pair of peptides is determined by parametric methods such as the Pearson r correlation or by nonparametric methods such as Kendall rank correlation and Spearman rank correlation. In a preferred embodiment, correlations are

measured by determining the coefficient of determination (r^2).

Data independent acquisition on TripleTOF mass spectrometers (SWATH)

[0098] Data independent acquisition (DIA) is an emerging technology in the field of mass spectrometry based proteomics. Although the concept of DIA has been around for over a decade, recent advancements, in particular an improved speed of acquisition, of mass analyzers has pushed the technique into the spotlight and allowed for high quality DIA data to be routinely acquired by proteomics labs. Described herein are exemplar protocols used for DIA acquisition using the Sciex TripleTOF mass spectrometers and data analysis using the Sciex processing software.

I. GENERAL

[0099] Data Independent Acquisition Mass Spectrometry (DIA-MS) is a long-standing technique (1, 2) that has garnered increased attention recently due to the development of new pipelines for extracting, identifying, and quantifying peptides using a targeted analysis approach (3, 4). SWATH™ couples DIA-MS with direct searching of individual samples against an established, and often a more exhaustive, peptide MS spectral library (3, 5, 6). SWATH™ is, therefore, a two-step process (Figure 13), development of the MS spectral library, most often on a pooled sample representing the breath of the experimental collection, using information dependent acquisition (IDA) (see **Note 1**) and then the subsequent analysis of each individual sample by DIA. Thus, a major advantage of SWATH™ is that it can maximize the peptides observed both within an individual sample and across all of the samples in an experimental set, thereby increasing proteome coverage, experimental efficiency, reducing quantitative variability, and minimizing missing data across an experimental matrix. It is important to note that SWATH™ is an emerging approach and methods for estimating peptide identification confidence and false discovery rates as well as the ideal approach for estimating peptide and protein quantity from transition extracted ion chromatograms are continuing to evolve along with the sensitivity and capabilities of the instrumentation itself. As with any large-scale quantitative screening method, care should be taken to confirm and validate the biological differences and conclusions that are derived from a SWATH™ experiment.

[0100] In a SWATH™ experiment, proteins are digested and either directly infused or, more often, separated by liquid chromatography (LC) prior to analysis on a TripleTOF mass spectrometers (5600 or 6600, Sciex), a Q-Exactive mass spectrometer (Thermo Scientific), or

any instrument with sufficiently high scan speed and a quadrupole mass filter. On the Triple TOF instruments, precursor peptide ion selection is performed by filtering precursors collectively through mass-to-charge windows, typically 4-10 m/z wide, sequentially across the entire m/z range of interest rather than selectively isolating a single precursor mass/charge (m/z) per MS/MS scan as performed in IDA-MS experiments. Due to the typically wider isolation windows used in DIA experiments, two or more co-eluting precursors are often fragmented collectively to produce an MS2 spectrum containing a convoluted mixture of fragment ions from multiple precursor ions.

[0101] One approach used to increase the ability to find and confidently identify peptides from these complex mixed spectra is to associate specific peptides with defined regions within the chromatographic elution profile. In order to accomplish this, retention time (RT) determination and alignments across samples is a key aspect of searching IDA data. Exogenous supplied RT standards(6) or endogenous RT (7) that are composed of peptides consistently observed across large number of samples must be used for RT calibration in order to properly align individual ion chromatograms across the entire sample's elution profile.

[0102] Optimization of m/z window number and dwell time/ion accumulation time per window is performed so that the instrument cycles through the entire desired precursor m/z range (e.g., 400 – 1250 m/z). This is largely instrument and sample specific. For the 6600 triple TOF, you can go up to 2250 m/z but we typically analyze between 400-1250 m/z for tryptic digests. When analyzing middle down or any peptides larger than the average tryptic peptides the full range can be used with the appropriate considerations to SWATH™ windows and cycle times. Ultimately, the key is to allow the instrument to cycle rapidly enough to capture multiple observations across the chromatographic elution profile for a given ion.

[0103] The data are subsequently searched against a sample specific peptide library that allows a set number of transition ion chromatograms to be extracted for a peptide within the window of its predicted RT (determined by its observed or normalized RT from the peptide library). The peak groups are scored according to several factors intended to discriminate a “true” peptide target from non-specific noise, and the distribution of these target scores are modeled against the distribution of scores attributed to decoy peak groups to determine a score cut off resulting in an acceptable false discovery rate. Relative peptide abundance is

then inferred from the aggregate of the area under the curve for each transition extracted ion chromatograms (XICs), and various statistical approaches are used to roll transition intensity XICs into peptide intensity estimates, which can then be used to estimate the overall protein intensity. In this chapter, we present the typical workflow used currently by our group to prepare, acquire, and analyze proteomic data for a DIA-MS experiment of cell or tissue samples. For simplicity and pragmatism we present the workflow as completed using SCIEX TripleTOF® instruments and data analysis platform exclusively, with mention of alternative approaches as appropriate.

1.1 Quality assurance and quality control (QA/QC) considerations

[0104] Robust quality assurance (QA) or quality control (QC) protocols are essential to monitor instrument performance and improve reproducibility and reliability of data. A QC standard run can be analyzed at fixed times such as the beginning and end of an experiment or day to assess variation in a variety of quality control metrics (8). For the TripleTOF instruments, we conduct internal mass calibrations of mass accuracy and sensitivity for both MS1 and MS2 scans every 3-5 runs by monitoring at least 8 peptides from 100 fmols digested beta-galactosidase standard (Sciex) and 7 transition ions from the 729.3652 $[M+2H]^{2+}$ ion (Table 19).

Table 19. Beta-galactosidase peptides used for autocalibration and quality control.

Beta-Galactosidase Peptide sequence	$[M+2H]^{2+}$	transition ions for 729.36	Fragment
YSQQQLMETSHR	503.2368		
RDWENPGVTQLNR	528.9341		
GDFQFNISR	542.2654		
IDPNAWVER	550.2802		
DVSLHLKPTTQISDFHVATR	567.0565		
VDEDQPFPAVPK	671.3379		
DWENPGVTQLNR	714.8469		
APLDNDIGVSEATR	729.3652		
		175.1190	y1
		347.2037	y3
		563.2784	y5
		729.3652	b7
		832.4523	y8
		1061.5222	y10
		1289.6332	y12

[0105] What also needs to be tracked is sample processing to ensure the quality to what is being analyzed, which is not addressed at in this manuscript but is well established in targeted multiple and selective monitoring work flows. To do this one can include a exogenously protein, such as beta galactosidase, is added into the sample prior to digestion. Beta-galactosidase elected peptides can be quantified (if ^{15}N labeled peptides are added after digestion to the sample) or assessed in each sample (for more details see Chen et al., in Salvatore Sechi, *Quantitative Proteomics by Mass Spectrometry (Methods in Molecular Biology) 2nd ed. 2016 Edition*, Humana Press (New York, NY, 2009))

[0106] Internal peptide retention time (RT) standards are an essential component of both peptide library generation and SWATH™ data analysis, and must be 1) detectable across all individual samples and 2) spread evenly across the chromatogram. Retention time of a given peptide from the library is used to set an extraction window for its peak group identification from the SWATH™ data file, and subsequently also used in scoring the confidence of a given peak group assignment to a peptide sequence from the library. If SWATH™ data files and peptide library files are collected absolutely sequentially with nearly identical chromatography, one might bypass the use of RT alignment standards. Much more commonly, differences in sample matrix, chromatographic set-ups, timing of instrument batch acquisitions, and many other factors can contribute to imperfect chromatographic alignment necessitating RT standards to normalize peptide assay library retention time to SWATH™ acquisition file retention time. Used alone or in combination with retention time standards that are spiked into a sample, endogenous reference peptides can also be used for the calibration of retention times across samples (7). These can be unique to a specific library (sample), however, there are common and conserved peptides that may be present in most, if not all, mammalian cells and tissues which can be used as a complement or replacement to synthetic, externally spiked RT reference peptides (7). QC tools are available to assess quality control metrics in a shotgun or targeted proteomic workflow that allows chromatographic performance and systemic error to be monitored (9). Tracking RT standards across sample runs can also server to assess instrument performance.

[0107] As larger numbers of individual samples are analyzed adopting other routine QC such as randomization or blocking of sampled to minimize sample analysis bias and regular collection of quality control samples spaced evenly and strategically throughout acquisition batches can be necessary components of SWATH™ experimental design.

1.2 Spectral Library Building – Data Generation

[0108] The use of a spectral ion library is most often used for the targeted analysis of SWATH™ data, although other methods are being explored and developed (10, 11), and can be primarily cell or tissue and species specific or a broader library assembled from all relevant peptide observations from a given species (5). Spectral ion libraries are most commonly built using traditional shotgun proteomics in information dependent acquisition (IDA) MS mode. In some cases spectral ion libraries previously generated have been made available to the public from various labs (5, 12, 13). Here we describe the creation of new spectral ion libraries from IDA analysis of proteolytic digestions. Additional detailed information regarding the generation of spectral ion libraries, including the management of protein redundancy and isoform specificity, can be found in Schubert et al (5). It is important to consider differences in peptide fragmentation patterns between instruments, and ideally use IDA data acquired on the same instrument from which you perform your SWATH™ acquisition (14).

[0109] Spectral ion libraries can be constructed in a number of ways. The first and most straightforward way to create an ion library is to analyze a proteolytic digestion in IDA mode of a pooled sample created from all of the individual samples that can be subsequently analyzed by DIA or of samples composing the extremes of the phenotype. This can give the most basic ion library comprising the peptides identified in a single IDA run that can then be used against the SWATH™ acquired version of itself and any other SWATH™ acquired sample of the same general proteome. In an attempt to expand the number of ions selected for fragmentation for library generation from a single IDA run of the pooled sample, multiple runs or technical replicates might help increase the proteome coverage provided to the sample library beyond what may be obtained from a single run and thus may help compensate for the error in sampling that is inherent to DIA methods. Alternatively, deeper and more inclusive ion libraries can be constructed post-digestion using off-line peptide fractionation and analysis of these fractions independently in IDA mode. The IDA runs are then combined to create a more complete and inclusive ion library for the given sample proteome and should ultimately increase the power of DIA-base protein identifications by increasing the number of peptides used to quantitate highly abundant proteins while harnessing the sensitivity of MS2-based quantitation necessary of low abundance proteins and peptides. Some methods commonly used for peptide fractionation are basic-reverse phase HPLC (bRP-HPLC) (15),

strong cation exchange (SCX), and strong anion exchange (SAX) (16) (see **Notes 2 and 3**). Our lab typically uses bRP-HPLC or a solid phase extraction SCX (17) method for peptide fractionation prior to MS analysis. For SWATH™ analysis of post-translational modifications it is recommended to employ enrichment strategies (if applicable) either independently or in combination with the peptide fractionation techniques described and as typically performed in shotgun experiments.

[0110] The following exemplar protocol is for library generation using Sciex TripleTOF™ systems with an Eksigent® 415 nano LC and ekspert 400 autosampler, although alternative LC and autosamplers may be used with the TripleTOF systems.

II. MATERIALS

[0111] Proteolytic peptide mixture, most often MS-grade trypsin (Promega)

[0112] 5600 or 6600 TripleTOF system

[0113] Nano-LC and autosampler (e.g. Eksigent® 415 nano LC, ekspert™ 400 autosampler) and ekspert™ cHiPLC (optional)

[0114] Trap and analytical LC columns (Eksigent® P/N 804-00006 and 804-00001)

[0115] Proteolytic peptide mixture, most often MS-grade trypsin (Promega)

[0116] 5600 or 6600 TripleTOF system

[0117] Retention time standards, either commercial peptides that are spiked in right before MS analysis (e.g. Biogynosis cat# KI-3002-2) or endogenous peptides present in all samples can be used (Parker et al, in press) (see **Note 4**).

[0118] **Software Needed** (see **Note 5**)

[0119] Analyst TF 1.7

[0120] PeakView 2.0 or higher

[0121] Variable Window Calculator

[0122] Protein Pilot 4.5 or higher

[0123] SWATH™ microapp

[0124] Microsoft Excel

[0125] MarkerView (optional)

III. METHODS

3.1 IDA analysis of proteolytic digests for spectral ion library building

[0126] 3.1.1 Create an IDA method in Analyst TF 1.7 with 1 survey scan and 20 candidate ion scans per cycle (see **Note 6**). Check the **Rolling Collision Energy** box.

[0127] 3.1.2 For TOF MS (MS1)

[0128] Under the **MS Tab** set the accumulation time to 250 ms and the mass range from 400-1250Da (**Figure 14**, see **Note 7**). Set the method duration to match the length of your LC gradient method.

[0129] Under the **Switch Criteria** tab set the range to match what you selected under the above window, monitor charge states from 2 to 5 which exceed 150 counts, set the mass tolerance to 50 ppm, and set your exclusion criteria (**Figure 15**, see **Note 8**).

[0130] Under the **Include/Exclude** tab put in any masses you want to monitor or exclude in your analysis.

[0131] Under the **IDA Advanced** tab make sure Rolling Collision Energy is checked and make any other necessary changes that would be pertinent to your experiment.

[0132] Default settings do not need to be changed under the **Advanced MS** tab.

[0133] 3.1.3 For Product Ion (MS2)

[0134] Under the **MS Tab** set the accumulation time to 100 ms and the mass range from 100-1800 Da⁷ and check whether you want high resolution or high sensitivity (the high sensitivity function is most commonly selected for proteomics experiments).

[0135] All other tabs should maintain the same parameters as for the TOF MS and do not need to be changed.

[0136] 3.1.4 Load the sample appropriate Gradient, Loading Pump, and auto-sampler methods and save your Acquisition File.

[0137] 3.1.5 Analyze your peptide samples.

3.2 SWATH-MS Data Acquisition

[0138] 3.2.1 Creation of Variable Window SWATH™ methods

[0139] Optimized SWATH™ methods can be constructed for specific samples using the Sciex Variable Window Calculator application. The steps for creating the customized SWATH™ variable windows for a specific sample are listed in the Variable Window Calculator under the **Instructions and Controls** tab. After following these directions select the number of variable windows (see **Note 9**) you want to analyze in your method and the mass range of the SWATH™ analysis. For general proteomics experiments the window overlap is usually left at 1 Da and the collision energy spread (CES) is usually left at 5. The minimum window width should be set no lower than 4 due to the default parameters in the PeakView software. After the Variable Window calculator is finished creating the optimal

windows for your analysis go to the **OUTPUT for Analyst** tab and copy columns A,B, and C into a new Excel file and save as a Text (Tab Delimited) file which can then be loaded into the SWATH™ method within Analyst TF 1.7.

[0140] 3.2.2 Creation of a SWATH™ method in Analyst TF 1.7

[0141] 3.2.2.1 In Analyst TF 1.7 go to the **Build Acquisition Method** tab on the left hand side of the window. Click on TOF MS and select Create SWATH™ Exp button then select the **Manual** tab within this window.

[0142] 3.2.2.2 Under **SWATH™ Analysis Parameters** select the mass range of the analysis (typically 400-1250 Da for tryptic peptides). Under **Fragmentation Conditions** make sure Rolling Collision energy is checked (the CES set in the Variable Window Calculator can overwrite the CES value inputted on this screen). Under **SWATH™ Detection Parameters** select the mass range to monitor for the SWATH™ MS2 spectra (typically 100-1800 Da) and the accumulation time for each window (typically for 100 VW 30 ms is adequate) (see **Note 10**). Lastly, click the **Read SWATH™ Windows from Text File** box and load in your .txt file create in the Variable Window Calculator.

[0143] The accumulation time for the MS1 can be set between 50-150 ms to give a quick survey scan for each cycle (see **Note 11**). Select the appropriate loading pump, gradient, and auto-sampler methods for the file (see **Note 12**). The gradient method chosen should be the same one that was used during the IDA analysis performed to generate the proteome specific spectral library.

3.3 SWATH™ Data Analysis using PeakView 2.1 and SWATH™ microapp 2.0

[0144] 3.3.1 Introduction to SWATH™ data analysis procedure

[0145] As with many methodologies, there are several options for processing SWATH™ data and analyzing results. Here, we present the protocol to process data through the SCIEX proprietary software. In our lab, we also regularly utilize two alternative pipelines, Skyline (18) and OpenSWATH (4). Skyline is a free and open-source tool built in Windows computing environments for analysis of multiple MS data types, including DIA. OpenSWATH™ is a free and open-source built within the openMS data analysis tool space, and operates optimally in a linux computing environment. A summary of the basic information pertaining to using these two alternate data analysis pathways is provided in Table 20.

Table 20. Selected alternative DIA-MS data analysis approaches

Parameters	Skyline ¹	OpenSWATH ²
Input DIA File format	.WIFF	.mzML/.mzXML ³
Peptide Ion Library	Built from DDA search result files (e.g., pep.xml, .group) or imported as a "transition list"	Built using TPP tools and custom Python scripts ⁴
SWATH Workflow	Internal to Skyline	OpenSwathWorkflow.exe
Output File Format	.csv transition report	.tsv transition report
Visualization	Internal to Skyline	TAPIR ⁵
Peak Picking Algorithm	mProphet ⁶ adaptation	pyProphet ⁷
Multi-Run Alignment	-	Feature Alignment ⁸
Quantitative Statistics	Linked External Tool MSstats ⁹	External Tools (eg. MapDIA ¹⁰ , MSstats)

[0146] ¹MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968 (2010).

[0147] ²Röst HL et al OpenSWATH™ enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology* 10;32(3):219-23 (2014)

[0148] ³ Conversion to mzML or mzXML can be done using the tool msconvert, available at: (<http://proteowizard.sourceforge.net/tools/msconvert.html>). Do not select peak picking, files may expand 10x or more from raw file size.

[0149] ⁴Schubert OT et al., Building high-quality assay libraries for targeted analysis of SWATH™ MS data. *Nature Protocols*, 10(3):426-41 (2015). Note: libraries generated using the pipeline described in the Schubert et al paper can be formatted for use in the PeakView microapp, and substituted in the workflow above.

[0150] ⁵<https://github.com/msproteomicstools/msproteomicstools/blob/master/gui/TAPIR.py>

[0151] ⁶<http://www.mprophet.org/>

[0152] ⁷<https://pypi.python.org/pypi/pyprophet>

[0153] ⁸ python script, available to download from <https://github.com/msproteomicstools>, found in folder msproteomicstools/analysis/alignment/feature_alignment.py

[0154] ⁹<http://www.msstats.org/>

[0155] ¹⁰<http://mapdia.sourceforge.net/Main.html>

[0156] In this section, we provide a summary specific to the approach used in our lab for the general implementation of the SCIEX software tools. We recommend referring to the SCIEX software user manuals for additional guidance.

[0157] 3.3.2 Creation of Spectral Ion Library using Protein Pilot Paragon Method

[0158] 3.3.2.1 Prepare the protein reference database that you use for matching DDA spectra to peptide sequences. For instance, FASTA documents for annotated proteomes can be downloaded from the Uniprot website: (<http://www.uniprot.org/proteomes>). Typically, we chose to use the curated, or reference proteomes, for a given organism of interest.

[0159] If external retention time standards were used in the experiment, such as the Biognosys iRT (see **Note 13**) peptides, copy their sequences and append to your FASTA file by opening it in a text editor. FASTA proteome databases should be saved in the appropriate folder within the Protein Pilot software files on your computer as per the software manual instructions.

[0160] 3.3.2.2 In Protein Pilot, select the option for an LC MS search and prepare a database search method appropriate for your experiment, including all of the raw data files you would like to include to build the ion library.

[0161] 3.3.2.3 Once the search is completed open the “FDR report” generated for the search and record the number of proteins identified at 1% Global FDR to be used as input in the following section.

[0162] 3.3.3 Importing Ion Libraries into the SWATH™ microapp and analyzing SWATH™ data

[0163] 3.3.3.1 Open PeakView and using the tabs at the top of the screen, navigate to Quantitation→SWATH™ Processing→Import Ion Library (**Figure 16**).

[0164] 3.3.3.2 Find the .group file produced from the Protein Pilot search and set the number of proteins to import to the 1% Global FDR (see **Note 14**) recorded in the previous section from the FDR report generated by Protein Pilot. Typically peptides shared by more than one protein are not imported. Under **Select sample type**, chose the option appropriate for whether

the samples were unlabeled (typical) or labeled with a chemical tag (i.e. iTRAQ, SILAC, etc...).

[0165] 3.3.3.3 Select all of the SWATH™ files to be analyzed for a given experiment.

[0166] 3.3.3.4 Set your processing settings. For protein quantitation analysis, examples of typical parameter settings are given in **Figure 17** (see **Note 15**):

[0167] 3.3.3.5 After setting your processing settings click “Process” to analyze your SWATH™ data.

[0168] 3.3.3.6 Once completed you can export the data for visualization in MarkerView by clicking **Quantitation→SWATH™ Processing→Export→Areas** or **Export→All** to get a complete list of all parameters for the analysis in Excel format (**Figure 18**).

IV. NOTES

1. The Sciex terminology Information Dependent Acquisition (IDA) is the same as Data Dependent Acquisition (DDA) and this is the terminology used in the Sciex software for shotgun proteomics experiments. Here, we use the IDA acronym to be consistent with the Sciex terminology and software.

2. bRP-HPLC fractionation may be preferred over SCX or SAX fractionation if downstream phospho-peptide enrichment or analysis of other negatively charged peptides is desired. This is due to a more equal distribution of phospho-peptides throughout basic-RP fractions compared to SCX and SAX fractions, in which phospho-peptides are most dense in the early and late fractions, respectively.

3. The SCX method published by Dephoure and Gygi (17) was based on 10 mg of starting material and was used upstream of phosphopeptide enrichment. Our lab has used this method for both phosphoproteomic and general proteomic analysis and we have scaled back the protocol for 1 mg of starting material, in which we have cut the reagents used in the Dephoure & Gygi paper by 1/10th. If using less than 1 mg of starting material scale back the reagents accordingly (13).

4. If large number of samples include beta-galactosidase for sample preparation assessment and N15 labeled peptides to track (see Chen et al., in Salvatore Sechi, *Quantitative Proteomics by Mass Spectrometry (Methods in Molecular Biology) 2nd ed. 2016 Edition*, Humana Press (New York, NY, 2009)).

5. Sciex software can be downloaded at

<http://www.absciex.com/downloads/software-downloads>

6. The number of survey scans desired for the analysis of concatenated or single run samples for library generation is a matter of user discretion but a typical IDA method on a TripleTOF system uses 20 candidate ions.
7. The 5600 TripleTOF system can go up to 1250 m/z and the 6600 TripleTOF can go up to 2250 m/z. However, we find that for tryptic digests there is little additional peptide data obtained above 1250 m/z. The larger mass range on the 6600 system is beneficial when doing large protein modifications such as glycoproteomics or when using alternative proteolytic methods that produce larger peptides (i.e. Lys-C, CNBr).
8. These values are meant to be used as a general guide in setting up an IDA method. Optimization for individual systems and sample types may be required for optimal results. For PTM and low abundant peptide analysis the accumulation times may be adjusted to allow for increased signal in both the MS1 and MS2 scans.
9. The number of variable windows chosen should be considered carefully as the more windows selected the shorter the dwell time has to be for each window. For general purposes 100 VW and a 30 ms dwell time should be sufficient to yield good quantitation of peptides.
10. If accumulation times less than 30 ms are desired it is recommended that they be tested prior to large scale sample analysis to ensure the accumulation time chosen can give adequate signal for quantitation.
11. If using the 5600 TripleTOF system, the minimum accumulation time for the MS1 should be set to 150 ms to ensure the MS1 quality is sufficient to perform the background calibrations during the run. The 6600 TripleTOF system does not use this background calibration so a shorter MS1 accumulation time (50ms) may be used to get a quick survey scan.
12. The LC and auto-sampler methods can vary between labs and the gradient lengths can vary depending on the complexity of the samples. Typically, for complex mixtures a gradient of 5-35%B over 90-120 minutes is suitable and for less complex samples (i.e. immunoprecipitations, purified proteins) shorter gradients between 30 and 60 minutes may be sufficient.
13. iRT FASTA sequence is available at www.biognosys.com, or type the following into your FASTA file:
 - 13.1.1. >Biognosys iRT Kit Fusion

AGGSSEPVTGLADKVEATFGVDESANKYILAGVESNKDAVTPADFSEWSKFLQFG
AQGSPLFKLGGNETQVRTPVISGGPPYYERTPVITGAPYYERGDLDAAASYAPVRTGFI
IDPGGVIRGTFIIDPAAIVR (SEQ ID NO:81)

14. FDR threshold can be set higher or lower depending on the user preference, the higher the FDR is set the more proteins can be incorporated into the library but the confidence of these proteins cannot be as high as if a lower FDR threshold is used.

15. These parameters are meant as a guideline and can be adjusted based on user preferences. Refer to the Sciex PeakView software documentation and the literature regarding optimizing these settings for your particular experiment. Importantly, for PTM analysis, un-check the Exclude Modified Peptides box and increase the number of peptides per protein to a larger value (i.e. 100) to import all peptides identified at the confidence level selected or create a PTM enriched peptide library.

V. REFERENCES

1. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature methods* 1 (1):39-45. doi:10.1038/nmeth705
2. Dong MQ, Venable JD, Au N, Xu T, Park SK, Cociorva D, Johnson JR, Dillin A, Yates JR, 3rd (2007) Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*. *Science* 317 (5838):660-663. doi:10.1126/science.1139952
3. Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, Aebersold R (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & cellular proteomics : MCP* 11 (6):O111 016717. doi:10.1074/mcp.O111.016717
4. Rost HL, Rosenberger G, Navarro P, Gillet L, Miladinovic SM, Schubert OT, Wolski W, Collins BC, Malmstrom J, Malmstrom L, Aebersold R (2014) OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* 32 (3):219-223. doi:10.1038/nbt.2841
5. Schubert OT, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE, Lam H, Amodei D, Mallick P, MacLean B, Aebersold R (2015) Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nature protocols* 10 (3):426-441. doi:10.1038/nprot.2015.015

6. Wang J, Perez-Santiago J, Katz JE, Mallick P, Bandeira N (2010) Peptide identification from mixture tandem mass spectra. *Molecular & cellular proteomics : MCP* 9 (7):1476-1485. doi:10.1074/mcp.M000136-MCP201
7. Parker S, Rost H, Rosenberger G, Collins BC, Maelstrom L, Amodei D, Venkatramen V, Raedschelders K, Van Eyk J, Aebersold R (2015) Identification of a Set of Conserved Eukaryotic Internal Retention Time Standards for Data-Independent Acquisition Mass Spectrometry. *Molecular & cellular proteomics : MCP* Conditionally Accepted
8. Bereman MS (2015) Tools for monitoring system suitability in LC MS/MS centric proteomic experiments. *Proteomics* 15 (5-6):891-902. doi:10.1002/pmic.201400373
9. Bereman MS, Johnson R, Bollinger J, Boss Y, Shulman N, MacLean B, Hoofnagle AN, MacCoss MJ (2014) Implementation of statistical process control for proteomic experiments via LC MS/MS. *J Am Soc Mass Spectrom* 25 (4):581-587. doi:10.1007/s13361-013-0824-5
10. Tsou CC, Avtonomov D, Larsen B, Tucholska M, Choi H, Gingras AC, Nesvizhskii AI (2015) DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature methods* 12 (3):258-264, 257 p following 264. doi:10.1038/nmeth.3255
11. Ting S, Egertson J, MacLean B, Kim S, Payne S, Noble W, MacCoss MJ Pecan: Peptide Identification Directly from Data-Independent Acquisition (DIA) MS/MS Data. In: American Society for Mass Spectrometry, Baltimore, MD, 2014.
12. Toprak UH, Gillet LC, Maiolica A, Navarro P, Leitner A, Aebersold R (2014) Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Molecular & cellular proteomics : MCP* 13 (8):2056-2071. doi:10.1074/mcp.O113.036475
13. Kirk JA, Holewinski RJ, Kooij V, Agnetti G, Tunin RS, Witayavanitkul N, de Tombe PP, Gao WD, Van Eyk J, Kass DA (2014) Cardiac resynchronization sensitizes the sarcomere to calcium by reactivating GSK-3beta. *The Journal of clinical investigation* 124 (1):129-138. doi:10.1172/JCI69253
14. Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, Rinner O (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* 12 (8):1111-1121. doi:10.1002/pmic.201100463

15. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, Moore RJ, Klemke RL, Camp DG, 2nd, Smith RD (2011) Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 11 (10):2019-2026. doi:10.1002/pmic.201000722
16. Han G, Ye M, Zhou H, Jiang X, Feng S, Jiang X, Tian R, Wan D, Zou H, Gu J (2008) Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics* 8 (7):1346-1361. doi:10.1002/pmic.200700884
17. Dephoure N, Gygi SP (2011) A solid phase extraction-based platform for rapid phosphoproteomic analysis. *Methods* 54 (4):379-386. doi:10.1016/j.ymeth.2011.03.008
18. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26 (7):966-968. doi:10.1093/bioinformatics/btq054

[0169] In some embodiments, acquiring MS data does not require operating a mass spectrometer. For examples, MS data can be acquired from MS experiments run previously and/or MS databases. In some embodiments, previously acquired SWATH MS data can be queried with a more comprehensive library to identify additional MS peaks derived from different and macromolecules.

[0170] In various embodiments, acquiring MS data comprises operating a TripleTOF mass spectrometer, a triple quadrupole mass spectrometer, a liquid chromatography-mass spectrometry (LC-MS) system, a gas chromatography-mass spectrometry (GC-MS) system, or a tandem mass spectrometry (MS/MS) system, a dual time-of-flight (TOF-TOF) mass spectrometer, or a combination thereof.

[0171] In various embodiments, acquiring MS data comprises operating a mass spectrometer. Examples of the mass spectrometer include but are not limited to high-resolution instruments such as Triple-TOF, Orbitrap, Fourier transform, and tandem time-of-flight (TOF/TOF) mass spectrometers; and high-sensitivity instruments such as triple quadrupole, ion trap, quadrupole TOF (QTOF), and Q trap mass spectrometers; and their hybrid and/or combination. High-resolution instruments are used to maximize the detection of peptides

with minute mass-to-charge ratio (m/z) differences. Conversely, because targeted proteomics emphasize sensitivity and throughput, high-sensitivity instruments are used. In some embodiments, the mass spectrometer is a TripleTOF mass spectrometer. In some embodiments, the mass spectrometer is a triple quadrupole mass spectrometer.

[0172] In various embodiments, the MS data is collected by a targeted acquisition method. Examples of the targeted acquisition method include but are not limited to Selective Reaction Monitoring (SRM) and/or Multiple Reaction Monitoring (MRM) methods. In various embodiments, acquiring MS data comprises acquiring Selective Reaction Monitoring (SRM) data and/or Multiple Reaction Monitoring (MRM) data.

[0173] In various embodiments, the MS data is collected by a data independent acquisition method. Examples of the independent acquisition (DIA) method including but not limited to Shotgun CID (see *e.g.*, Purvine et al. 2003), Original DIA (see *e.g.*, Venable et al. 2004), MS^E (see *e.g.*, Silva et al. 2005), p2CID (see *e.g.*, Ramos et al. 2006), PAcIFIC (see *e.g.*, Panchaud et al. 2009), AIF (see *e.g.*, Geiger et al. 2010), XDLA (see *e.g.*, Carvalho et al. 2010), SWATH (see *e.g.*, Gillet et al. 2012), and FT-ARM (see *e.g.*, Weisbrod et al. 2012). More information can be found in, for example, Chapman et al. (*Multiplexed and data-independent tandem mass spectrometry for global proteome profiling*, Mass Spectrom Rev. 2014 Nov-Dec;33(6):452-70). In various embodiments, acquiring MS data comprises acquiring Shotgun CID MS data, Original DIA MS Data, MS^E MS data, p2CID MS Data, PAcIFIC MS Data, AIF MS Data, XDLA MS Data, SWATH MS data, or FT-ARM MS Data, or a combination thereof. In certain embodiments, acquiring MS data comprises acquiring MS data comprises acquiring SWATH MS data.

[0174] In various embodiments, the sample is food, water, cheek swab, blood, serum, plasma, urine, saliva, semen, cell sample, tissue sample, or tumor sample, or a combination thereof.

[0175] In various embodiments, the highly correlated peptides form a subset of all queried peptides and have correlation values when compared with other members of the subset that are more than 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.40, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.50, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.60, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.70, 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.80, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88 or 0.89. In various embodiments, the highly correlated peptides form a subset of all queried peptides and have correlation values when compared

with other members of the subset that are more than 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 or 0.99. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0176] In various embodiments, the method further comprises ranking the correlation values of the multiple candidate peptides. In various embodiments, the highly correlated peptides have correlation values ranked in the top 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 among the multiple candidate peptides. In various embodiments, the highly correlated peptides have correlation values ranked in the top 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% among the multiple candidate peptides. In various embodiments, the highly correlated peptides have correlation values ranked in the top 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 30% or 20% among the multiple candidate peptides. In certain embodiments, the highly correlated peptides have correlation values ranked in the top 2, 3, 4, 5, 6, 7, 8, 9, or 10 among the multiple candidate peptides. In certain embodiments, the highly correlated peptides have correlation values ranked in the top 80%, 70%, 60%, 50%, 40%, 30% or 20% among the multiple candidate peptides. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0177] In various embodiments, all of the correlation values of a candidate peptide are considered as indicators for the candidate peptide's correlation level. In various embodiments, a highly correlated peptide has all or half of its correlation values more than 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.40, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.50, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.60, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.70, 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.80, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88 or 0.89. In various embodiments, a highly correlated peptide has all or half of its correlation values more than 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 or 0.99. In various embodiments, a highly correlated peptide has all or half of its correlation values more than 0.990, 0.991, 0.992, 0.993, 0.994, 0.995, 0.996, 0.997, 0.998 or 0.999. In various embodiments, a highly correlated peptide has all or half of its correlation values ranked in the top 2, 3, 4, 5, 6, 7, 8, 9, or 10 among the multiple candidate peptides. In various embodiments, a highly correlated peptide has all or half of its correlation values ranked in the top 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% among the multiple candidate peptides. In various embodiments, a highly correlated peptide has all or half of its

correlation values ranked in the top 80%, 70%, 60%, 50%, 40%, 30% or 20% among the multiple candidate peptides. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0178] In various other embodiments, a subset of correlated peptides is selected from among the set of peptides in a correlation matrix. Members of the subset all have correlation values of more than 0.80, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 or 0.99 for pairwise combinations with all other members of the subset. Signature peptides are then selected from the subset of correlated peptides. In various other embodiments, an average is calculated from the correlation values for each peptide in a correlation matrix. Signature peptides are then selected from among those peptides with the highest 30%, 40%, 50%, 60%, 70%, 80% or 90% of averages. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0179] In various embodiments, the correlation values of a candidate peptide are used to calculate the candidate peptide's mean or media correlation value, which is then considered as an indicator of the candidate peptide's correlation level. In various embodiments, a highly correlated peptide has a mean or median correlation value more than 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.40, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.50, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.60, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.70, 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.80, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88 or 0.89. In various embodiments, a highly correlated peptide has a mean or median correlation value more than 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 or 0.99. In various embodiments, a highly correlated peptide has a mean or median correlation value more than 0.990, 0.991, 0.992, 0.993, 0.994, 0.995, 0.996, 0.997, 0.998 or 0.999. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0180] In various embodiments, the method further comprises ranking the mean or median correlation values of the multiple candidate peptides. In various embodiments, a highly correlated peptide has a mean or median correlation value ranked in the top 2, 3, 4, 5, 6, 7, 8, 9, or 10 among the multiple candidate peptides. In various embodiments, a highly correlated peptide has a mean or median correlation value ranked in the top 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% among the multiple candidate peptides. In various embodiments, the highly correlated peptide has a mean or median correlation values ranked in the top 80%,

70%, 60%, 50%, 40%, 30% or 20% among the multiple candidate peptides. In certain embodiments, the highly correlated peptides have mean or median correlation values ranked in the top 2, 3, 4, 5, 6, 7, 8, 9, or 10 among the multiple candidate peptides. In certain embodiments, the highly correlated peptides have mean or median correlation values ranked in the top 80%, 70%, 60%, 50%, 40%, 30% or 20% among the multiple candidate peptides. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0181] In various embodiments, a method as described herein is an iterative process. For a non-limiting example, an initial set of multiple candidate peptides are subject to a first round of signature peptide identification according to a method as described herein, including but limited to the steps of: (1) using the MS data to calculate correlation values for pairwise comparisons among the complete initial set of multiple candidate peptides; (2) calculating each candidate peptide's mean or median correlation value; (3) ranking the multiple candidate peptides' mean or median correlation values; and (4) retaining those candidate peptides with mean or median correlation values among the top 90%, 80%, 70%, 60%, or 50% as the second set of multiple candidate peptides. Then, the second set of multiple candidate peptides are subject a second round of signature peptide identification, with the above steps (1)-(4) being repeated. This iterative process continues until reaching the final set of highly correlated peptides that are hence identified as the signature peptides for quantifying the polypeptide. In various embodiments, there can be 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more rounds of signature peptide identification. In various embodiments, the final set of highly correlated peptides have mean or median correlation value more than 0.80, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 or 0.99. In various embodiments, the final set of highly correlated peptides have mean or median correlation value more than 0.990, 0.991, 0.992, 0.993, 0.994, 0.995, 0.996, 0.997, 0.998 or 0.999. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0182] In various embodiments, the multiple candidate peptides are obtained from a data-dependent MS screen, data-independent MS data, targeted peptides data, MS spectral database, or proteotypic peptide prediction, or a combination thereof. In some embodiments, the proteotypic peptide prediction is a prediction of protease digestion of the polypeptide. In some embodiments, the proteotypic peptide prediction is a prediction of trypsin digestion of the polypeptide.

[0183] In various embodiments, the method further comprises eliminating peptides that satisfy one or more of the following criteria: (i). not previously detected by MS; (ii). not unique to the polypeptide; (iii). absent from the polypeptide's mature form; (iv.) containing an uncleaved protease recognition site; (v.) susceptible to post-translational modification (PTM), or known to be post-translationally modified in some forms of the protein; (vi.) containing methionine and/or cysteine residues; (vii.) sensitive to endogenous proteases, or miscleaved or incompletely cleaved; (viii.) having m/z values lower than the quantifiable range for the mass spectrometer or sample type (for example, an m/z bottom cutoff value); (ix.) having m/z values higher than the quantifiable range for the mass spectrometer or sample type (for example, an m/z top cutoff value); and (x.) having signal intensities lower than an intensity bottom cutoff value in the acquired MS data (for example, less than 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 16-fold, 17-fold, 18-fold, 19-fold, or 20-fold higher than the background noise in the MS data).

[0184] Examples of PTMs include but are not limited to N-linked glycosylation, O-linked glycosylation, C-mannosylation, GPI anchors (glypiation), phosphorylation on tyrosine, serine or threonine, disulfide bonds, deamidation of asparagine, and methionine oxidation. In various embodiments, one or more of these elimination criteria are applied before acquiring the MS data. In various embodiments, one or more of these elimination criteria are applied before calculating correlation values. In various embodiments, one or more of these elimination criteria are applied after acquiring the MS data. In various embodiments, one or more of these elimination criteria are used after calculating correlation values.

[0185] In various embodiments, the m/z bottom cutoff value is about 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, or 300. In one embodiment, the m/z bottom cutoff value is about 200.

[0186] In various embodiments, the m/z top cutoff value is about 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, or 2500. In various embodiments, the m/z top cutoff value is about 2000.

[0187] In some embodiments, the intensity bottom cutoff value is the background noise' intensity value. In some embodiments, the intensity bottom cutoff values is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 times of the background noise' intensity value.

In one embodiment, the intensity bottom cutoff values 10 times of the background noise' intensity value.

[0188] In various embodiments, the identified signature peptides have high and reproducible signal intensities in the acquired MS data. In some embodiments, the identified signature peptides have peak areas of more than 100, 200, 300, 400, 500, 600, 700, 800, 1000, 1250, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 3000, 3500 or 4000. In one embodiment, the identified signature peptides have signal intensities more than 2000.

[0189] Various cutoff values described herein (e.g., the m/z bottom cutoff value, the m/z top cutoff value, and the intensity bottom cutoff value) can have variations for different samples and instruments. It is contemplated that an ordinarily skilled artisan will recognize characteristics of different samples and instruments and apply appropriate cutoff values with respect to those characteristics.

[0190] The identified signature peptides can be used to build quantitative assays of the polypeptide. Various embodiments of the present invention also provide a method of quantifying a polypeptide in a sample. The method comprises: cleaving the polypeptide to yield a signature peptide identified according to a method as described herein; analyzing the sample on a mass spectrometer; detecting MS signals of the signature peptide; and quantifying the polypeptide based on the detected MS signals. In some embodiments, multiple polypeptides in a complex sample are quantified.

[0191] In various embodiments, the method further comprises spiking the sample with an internal standard of the signature peptide and detecting the internal standard's MS signals in the sample. In some embodiments, the internal standard comprises the signature peptide labeled with a stable isotope. Examples of the stable isotope include but are not limited to ^{15}N (nitrogen-15), ^{13}C (carbon-13), and ^2H (deuterium). In various embodiments, the method further comprises normalizing the signature peptide's MS signals detected in the sample to the internal standard's MS signals detected in the sample.

Internal Standards and Methods of Making Internal Standards

[0192] Stable Isotope Labeled (SIL) peptides, small molecules and lipids, including but not limited to peptides synthesized with $^{13}\text{C}/^{15}\text{N}$ universally-labeled Arg (+10) and Lys (+8).

[0193] Stable Isotope Labeled (SIL) proteins, including but not limited to ^{15}N labeled proteins, ^{15}N - ^{13}C labeled proteins, and ^{15}N - ^{13}C - ^2H -labeled proteins.

[0194] Metabolically labeled proteins. There are multiple methods of this type of in vivo labeling. One exemplar method is Stable Isotope Labeling by Amino acids in Cell culture (SILAC). Cells are cultured in growth medium that contains $^{13}\text{C}_6$ -lysine and/or $^{13}\text{C}_6$ -arginine. Another exemplar method is to feed carnivores with $^{13}\text{C}_6$ -lysine and/or $^{13}\text{C}_6$ -arginine to animals.

[0195] Stable isotopic labeling. Chemical or enzymatic stable isotopic labeling methods are used for samples that are not amenable to metabolic labeling (e.g., clinical samples) and/or when experimental time is limited. Non-limiting examples include adding isotopic atoms or isotope-coded tags to peptides or proteins.

[0196] As one non-limiting example: enzymatic labeling with ^{18}O takes advantage of the proteolytic mechanism of trypsin to incorporate two heavy oxygen atoms from H_2^{18}O at the C-terminus of every newly digested peptide.

[0197] As another non-limiting example: Global Internal Standard Technology (GIST), which uses deuterated (^2H) acylating agents such as N-acetoxysuccinimide (NAS) to label primary amino groups on digested peptides. Acylation of these groups, though, changes the ionic states of peptides and may affect the ionization efficiency of peptides with C-terminal lysines.

[0198] As another non-limiting example: chemical labeling by stable isotope dimethylation. This approach uses formaldehyde in deuterated water to label primary amines with deuterated methyl groups.

[0199] As another non-limiting example: Isotope-Coded Affinity Tags (ICAT). This method originally comprised a sulfhydryl-reactive chemical crosslinking group, linkers with various amounts of heavy (deuterated) isotopes, and a biotin molecule for collection of labelled peptides on a streptavidin matrix.

[0200] As another non-limiting example: isobaric mass tags. A benefit of isobaric mass tags is the multiplex capabilities and thus increased throughput potential of this approach. Commercially available isobaric mass tags (e.g., TMT*, iTRAQ*)

[0201] The Isobaric tags for relative and absolute quantitation (iTRAQ) method is based on the covalent labeling of the N-terminus and side chain amines of peptides from trypsin digested proteins with tags of varying mass. This method offers the simultaneous analysis of 4, 6 or 8 biological samples. While the exact tags used vary depending on manufacturer, the basic components of all isobaric mass tag reagents consist of a mass reporter (tag) that has a

unique number of ^{13}C substitutions, a mass normalizer that has a unique mass that balances the mass of the tag to make all of the tags equal in mass.

[0202] Tandem mass tags (TMT or TMTs) are chemical labels. The tags contain four regions, namely a mass reporter region (M), a cleavable linker region (F), a mass normalization region (N) and a protein reactive group (R). The chemical structures of all the tags are identical but each contains isotopes substituted at various positions, such that the mass reporter and mass normalization regions have different molecular masses in each tag. The combined M-F-N-R regions of the tags have the same total molecular weights and structure so that during chromatographic or electrophoretic separation and in single MS mode, molecules labelled with different tags are indistinguishable. Upon fragmentation in MS/MS mode, sequence information is obtained from fragmentation of the peptide back bone and quantification data are simultaneously obtained from fragmentation of the tags, giving rise to mass reporter ions.

[0203] Isotope-Coded Protein Label (ICPL) isobaric mass tagging has also been adapted for use with protein labeling. ICPL is based on tagging stable isotope derivatives at the free amino groups of intact proteins, the method is applicable to any protein sample, including tissue extracts and body fluids. Some commercially available kits also offer isobaric tags with sulfhydryl-reactivity and anti-TMT antibody for affinity purification of cysteine-tagged peptides prior to LC-MS/MS.

[0204] In various embodiments, the method further comprises generating a standard curve for the polypeptide using external standards. Examples of the external standards include but are not limited to a series of known concentrations of the polypeptide to be quantified. In various embodiments, the method further comprises spiking the external standards with an internal standard of the signature peptide and detecting the internal standard's MS signals in the external standards. In various embodiments, the method further comprises normalizing the signature peptide's MS signals detected in the external standards to the internal standard's MS signals detected in the external standards. In various embodiments, the method further comprises quantifying the polypeptide in a sample based on the detected MS signals in the sample and the generated standard curve. In various embodiments, the same MS protocol or technique is used to analyze the external standards to generate the standard curve and to analyze the sample to quantify the polypeptide.

Systems and Computers of the Invention

[0205] Various embodiments of the present invention provide a system for identifying signature peptides for quantifying a polypeptide. The system comprises: a mass spectrometer configured for acquiring mass spectrometry (MS) data on multiple candidate peptides derived from the polypeptide in multiple samples; and a computer configured for using the MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide, wherein the mass spectrometer and the computer are connected via a communication link. In some embodiments, the computer is also configured for processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0206] Various embodiments of the present invention provide a system for identifying signature peptides for quantifying a polypeptide. The system comprises: a mass spectrometer configured for acquiring mass spectrometry (MS) data on multiple candidate peptides derived from the polypeptide in multiple samples; a first computer configured for processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks); and a second computer configured for using the processed MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide, wherein the mass spectrometer and the computers are connected via a communication link. In some embodiments, the first and second computers are the same computer. In other embodiments, the first and second computers are separate computers. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0207] In various embodiments, the computer comprises: a memory configured for storing a program; and a processor configured for executing the program, wherein the program comprises instructions for using the MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and for identifying the highly correlated

peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In some embodiments, the program further comprises instructions for processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0208] Various embodiments of the present invention provide a non-transitory computer-readable storage medium, wherein the non-transitory computer-readable storage medium is configured for storing a program, wherein the program is configured for execution by a processor of a computer, and wherein the program comprises instructions for using mass spectrometry (MS) data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In various embodiments, the MS data comprises raw MS data obtained from a mass spectrometer and/or processed MS data in which peptides and their fragments (e.g., transitions and MS peaks) are already identified, analyzed and/or quantified. In some embodiments, the program further comprises instructions for processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In some embodiments, the program further comprises instructions for operating a mass spectrometer to acquire MS data. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0209] Various embodiments of the present invention provide a computer. The computer comprises: a memory configured for storing a program; and a processor configured for executing the program, wherein the program comprises instructions for using mass spectrometry (MS) data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In various embodiments, the MS data comprises raw MS data obtained from a mass spectrometer and/or processed MS data in which peptides and their fragments (e.g., transitions and MS peaks) are already identified, analyzed and/or quantified.

In some embodiments, the program further comprises instructions for processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0210] Various embodiments of the present invention provide a computer implemented method. The method comprises: providing a computer as described herein; inputting mass spectrometry (MS) data into the computer; and operating the computer to use the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In some embodiments, the method further comprises operating the computer to process the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0211] Various embodiments of the present invention provide a non-transitory computer-readable storage medium, wherein the non-transitory computer-readable storage medium is configured for storing a program, wherein the program is configured for execution by a processor of a computer, and wherein the program comprises instructions for operating a mass spectrometer to acquire mass spectrometry (MS) data, for using the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In various embodiments, the MS data comprises raw MS data obtained from a mass spectrometer and/or processed MS data in which peptides and their fragments (e.g., transitions and MS peaks) are already identified, analyzed and/or quantified. In some embodiments, the program further comprises instructions for processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In some embodiments, the program further comprises instructions for operating a mass spectrometer to acquire MS data.

In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0212] Various embodiments of the present invention provide a computer. The computer comprises: a memory configured for storing a program; and a processor configured for executing the program, wherein the program comprises instructions for operating a mass spectrometer to acquire mass spectrometry (MS) data, for using the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In various embodiments, the MS data comprises raw MS data obtained from a mass spectrometer and/or processed MS data in which peptides and their fragments (e.g., transitions and MS peaks) are already identified, analyzed and/or quantified. In some embodiments, the program further comprises instructions for processing the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0213] Various embodiments of the present invention provide a computer implemented method. The method comprises: providing a computer as described herein; connecting the computer via a communication link to a mass spectrometer; and operating the computer to operate the mass spectrometer to acquire mass spectrometry (MS) data, to use the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide. In some embodiments, the method further comprises operating the computer to process the MS data to identify, analyze and/or quantify the multiple candidate peptides and fragments thereof (e.g., transitions and MS peaks) before calculating correlation values. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 16. In some embodiments, the correlation values are coefficient of determination (r^2) values.

[0214] Various embodiments of the present invention provide a non-transitory computer-readable storage medium, wherein the non-transitory computer-readable storage medium is configured for storing a program, wherein the program is configured for execution by a processor of a computer, and wherein the program comprises instructions for processing MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and for quantify the polypeptide based on the signature peptide. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0215] Various embodiments of the present invention provide a computer, comprising: a memory configured for storing a program; and a processor configured for executing the program, wherein the program comprises instructions for processing MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and for quantify the polypeptide based on the signature peptide. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0216] Various embodiments of the present invention provide a computer implemented method, comprising: providing a computer as described herein; inputting MS data into the computer; and operating the computer to process MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and to quantify the polypeptide based on the signature peptide. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0217] Various embodiments of the present invention provide a non-transitory computer-readable storage medium, wherein the non-transitory computer-readable storage medium is configured for storing a program, wherein the program is configured for execution by a processor of a computer, and wherein the program comprises instructions for operating a mass spectrometer to detect MS signals of a signature peptide for quantifying a polypeptide, and quantifying the polypeptide based on the detected MS signals. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0218] Various embodiments of the present invention provide a computer. The computer comprises: a memory configured for storing a program; and a processor configured for

executing the program, wherein the program comprises instructions for operating a mass spectrometer to detect MS signals of a signature peptide for quantifying a polypeptide, and quantifying the polypeptide based on the detected MS signals. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0219] Various embodiments of the present invention provide a computer implemented method. The method comprises: providing a computer as described herein; connecting the computer via a communication link to a mass spectrometer; and operating the computer to operate the mass spectrometer to detect MS signals of a signature peptide for quantifying a polypeptide, and to quantify the polypeptide based on the detected MS signals. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0220] In accordance with the present invention, a “communication link,” as used in this disclosure, means a wired and/or wireless medium that conveys data or information between at least two points. The wired or wireless medium may include, for example, a metallic conductor link, a radio frequency (RF) communication link, an Infrared (IR) communication link, an optical communication link, or the like, without limitation. The RF communication link may include, for example, WiFi, WiMAX, IEEE 802.11, DECT, 0G, 1G, 2G, 3G or 4G cellular standards, Bluetooth, and the like.

[0221] Computers and computing devices typically include a variety of media, which can include computer-readable storage media and/or communications media, in which these two terms are used herein differently from one another as follows.

[0222] Computer-readable storage media can be any available storage media that can be accessed by the computer, is typically of a non-transitory nature, and can include both volatile and nonvolatile media, removable and non-removable media. By way of example, and not limitation, computer-readable storage media can be implemented in connection with any method or technology for storage of information such as computer-readable instructions, program modules, structured data, or unstructured data. Computer-readable storage media can include, but are not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disk (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or other tangible and/or non-transitory media which can be used to store desired information.

Computer-readable storage media can be accessed by one or more local or remote computing devices, e.g., via access requests, queries or other data retrieval protocols, for a variety of operations with respect to the information stored by the medium.

[0223] On the other hand, communications media typically embody computer-readable instructions, data structures, program modules or other structured or unstructured data in a data signal that can be transitory such as a modulated data signal, e.g., a carrier wave or other transport mechanism, and includes any information delivery or transport media. The term “modulated data signal” or signals refers to a signal that has one or more of its characteristics set or changed in such a manner as to encode information in one or more signals. By way of example, and not limitation, communication media include wired media, such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media.

[0224] In view of the exemplary systems described above, methodologies that may be implemented in accordance with the described subject matter will be better appreciated with reference to the flowcharts of the various figures. For simplicity of explanation, the methodologies are depicted and described as a series of acts. However, acts in accordance with this disclosure can occur in various orders and/or concurrently, and with other acts not presented and described herein. Furthermore, not all illustrated acts may be required to implement the methodologies in accordance with the disclosed subject matter. In addition, those skilled in the art will understand and appreciate that the methodologies could alternatively be represented as a series of interrelated states via a state diagram or events. Additionally, it should be appreciated that the methodologies disclosed in this specification are capable of being stored on an article of manufacture to facilitate transporting and transferring such methodologies to computer and computing devices. The term article of manufacture, as used herein, is intended to encompass a computer program accessible from any computer-readable device or storage media.

Capturing Reagents, Antibodies and Immunoassays of the Invention

[0225] Various embodiments of the present invention provide a method of producing a capturing reagent. The method comprises: providing a signature peptide identified according to a method as described herein; and producing the capturing reagent specifically binding to the signature peptide. In some embodiments, the capturing reagent is an antibody. In other embodiments, the capturing reagent is an aptamer. In various embodiments, the aptamer is

DNA aptamer, RNA aptamer, XNA aptamer, or peptide aptamer, or a combination thereof. In various embodiments, the method further comprises using the signature peptide to screen an aptamer library; and identifying an aptamer specifically binding to the signature peptide. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17. In various embodiments, the aptamer specifically binds to the polypeptide to which the signature peptide is identified for. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17.

[0226] Various embodiments of the present invention provide a capturing reagent specifically binding to a signature peptide identified according to a method as described herein. In some embodiments, the capturing reagent is an antibody. In other embodiments, the capturing reagent is an aptamer. In various embodiments, the aptamer is DNA aptamer, RNA aptamer, XNA aptamer, or peptide aptamer, or a combination thereof. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0227] As used herein, aptamers refer to oligonucleotide or peptide molecules that bind to a specific target molecule. Aptamers are usually created by selecting them from a large random sequence pool. Aptamers can be classified as: DNA or RNA or XNA aptamers, which comprise (usually short) strands of oligonucleotides; and peptide aptamers, which comprise a short variable peptide domain, attached at both ends to a protein scaffold.

[0228] Various embodiments of the present invention provide a method of producing an antibody. The method comprises: providing a signature peptide identified according to a method as described herein; and immunizing an animal using the signature peptide, thereby producing the antibody. In various embodiments, the method further comprises isolating and/or purifying the antibody from the immunized animal. In various embodiments, the antibody specifically binds to the signature peptide. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17. In various embodiments, the antibody specifically binds to the polypeptide to which the signature peptide is identified for. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17.

[0229] Various embodiments of the present invention provide an antibody specifically binding to a signature peptide identified according to a method as described herein, or an antigen-binding fragment thereof. In various embodiments, the antibody is a polyclonal antibody or a monoclonal antibody. In various embodiments, the antibody can be of any

animal origin. Examples of the animal origin include but are not limited to human, non-human primate, monkey, mouse, rat, guinea pig, dog, cat, rabbit, pig, cow, horse, goat, and donkey. In some embodiments, the antibody is a humanized antibody. In some embodiments, the antibody is a chimeric antibody. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0230] Various embodiments of the present invention provide a method of quantifying a polypeptide in a sample. The method comprises: contacting the sample with an antibody as described herein or an antigen-binding fragment thereof; detecting the binding between the polypeptide and the antibody or the antigen-binding fragment thereof; and quantifying the polypeptide based on the detected binding. In various embodiments, the method further comprises generating a standard curve for the polypeptide using external standards. Examples of the external standards include but are not limited to a series of known concentrations of the polypeptide to be quantified. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0231] In various embodiments, quantifying a polypeptide in a sample comprises contacting the sample with an antibody as described herein and thereby forming antigen-antibody complexes. In the methods and assays of the invention, the quantity of a polypeptide can be determined using an antibody as described herein and detecting immunospecific binding of the antibody to the polypeptide. Examples of quantitative assays based on the antibody include but are not limited to western blot, enzyme-linked immunosorbent assay (ELISA) and radioimmunoassay.

[0232] Various embodiments of the present invention provide a method of quantifying a polypeptide in a sample. The method comprises using an antibody as described herein with SISCAPA (Stable Isotope Standards and Capture by Anti-Peptide Antibodies). SISCAPA applies existing mass spectrometry quantitation methods (e.g., MRM) to the measurement of signature peptides of protein biomarkers. It improves sensitivity by capture of these signature peptides on immobilized anti-peptide antibodies.

[0233] In various embodiments, the method comprise: cleaving the polypeptide in the sample to yield a signature peptide identified according to a method as described herein; spiking the sample with an internal standard of the signature peptide; capturing the signature peptide and internal standard with a capturing reagent specifically binding to the signature peptide;

analyzing the captured signature peptide and internal standard on a mass spectrometer; detecting MS signals of the signature peptide the internal standard; and quantifying the signature peptide based on the detected MS signals. In some embodiments, the capturing reagent is an antibody or an antigen-binding fragment thereof specifically binding to the signature peptide. In other embodiments, the capturing reagent is an aptamer specifically binding to the signature peptide. In some embodiments, capturing the signature peptide and internal standard comprises forming an antigen-antibody complex between the antibody or its fragment and the signature peptide and an antigen-antibody complex between the antibody or its fragment and the internal standard; isolating the antigen-antibody complexes from the sample; and dissociating the signature peptide and the internal standard from the antibody or its fragment. In various embodiments, the antibody or its fragment is attached to a magnetic bead for capturing the signature peptide and internal standard. In some embodiments, capturing the signature peptide and internal standard comprises forming a target-aptamer complex between the aptamer and the signature peptide and a target-aptamer complex between the aptamer and the internal standard; isolating the target-aptamer complexes from the sample; and dissociating the signature peptide and the internal standard from the aptamer. In various embodiments, the aptamer is attached to a magnetic bead for capturing the signature peptide and internal standard. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0234] SISCAPA technology is the smart shortcut to sensitive quantitation of protein biomarkers and targets. SISCAPA assays combine the precision of MRM mass spectrometry with the power of affinity enrichment to deliver a superior alternative to conventional immunoassays for protein quantitation. The SISCAPA workflow is highly automated and exploits familiar LC-MS/MS platforms widely used for drug and metabolite quantitation. SISCAPA provides a range of practical advantages over conventional ligand binding immunoassays. Sensitivity: SISCAPA improves peptide multiple reaction monitoring (MRM) sensitivity by 3-4 orders of magnitude over non-enriched samples. Specificity: SISCAPA combines antibody immunocapture selectivity with the near-absolute structural specificity of MRM mass spectrometry. Standardization: SISCAPA employs true internal standards (stable isotope labeled synthetic peptides) within each assay for reliable quantitation. Multiplexing: SISCAPA assays can be combined in mix-and-match panels without cross-reactions common

in sandwich immunoassays. Throughput: SISCAPA delivers highly purified peptide analytes, free of matrix components, for decreased LC times and higher throughput. Development: SISCAPA assay development is faster, less expensive and more straightforward than sandwich immunoassay development. More information on SISCAPA can be found in US 9274124 and Anderson, N. L. et al. (Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA), *Journal of Proteome Research* 3: 235-44 (2004)), which are incorporated herein by reference in their entirety as though fully set forth.

[0235] As a non-limiting example, serum or plasma samples to be analyzed by Siscapa MRM are first subjected to proteolytic digestion, yielding a complex mixture of peptides from which one or more signature peptides are selected as targets. Digestion is accomplished by unfolding the proteins in a chaotropic solvent and then adding an enzyme such as trypsin which specifically cleaves the sample proteins at lysine and arginine residues. A synthetic stable isotope labeled version of a target signature peptide is added in known amount to serve as an internal standard for quantitation. The target signature peptide and its corresponding internal standard are then captured by sequence specific anti-peptide antibodies (e.g., an anti-signature peptide antibody as described herein) attached to magnetic beads. A low-abundance target signature peptide can be captured from a large massive digest, extending detection sensitivity by orders of magnitude compared to unfractionated digests. The magnetic beads, with their peptide cargo can then be easily removed from the digest, washed extensively, and then finally placed in an acidic eluent solution in which the peptides disassociate from the antibodies. This specific capture process enriches the target signature peptide and corresponding internal standard by more than 100,000 fold while retaining the quantitative ratio between them. This ratio can then be measured precisely in a mass spectrometer providing a quantitation of the bio marker protein in the original sample. By providing an almost pure sample of the desired target signature peptide for analysis, detection sensitivity is maximized while shortening LC-MS cycle time for higher throughput.

[0236] Antibodies, both polyclonal and monoclonal, can be produced by a skilled artisan either by themselves using well known methods or they can be manufactured by service providers who specialize making antibodies based on known protein sequences. In the present invention, the signature peptide sequences are identified and thus production of antibodies against them is a matter of routine.

[0237] For example, production of monoclonal antibodies can be performed using the traditional hybridoma method by first immunizing mice with an antigen which may be an isolated peptide of choice or fragment thereof (for example, a signature peptide as described herein) and making hybridoma cell lines that each produce a specific monoclonal antibody. The antibodies secreted by the different clones are then assayed for their ability to bind to the antigen using, e.g., ELISA or Antigen Microarray Assay, or immuno-dot blot techniques. The antibodies that are most specific for the detection of the signature peptide can be selected using routine methods and using the antigen used for immunization and other antigens as controls. The antibody that most specifically detects the desired antigen and no other antigens are selected for the processes, assays and methods described herein. The best clones can then be grown indefinitely in a suitable cell culture medium. They can also be injected into mice (in the peritoneal cavity, surrounding the gut) where they produce an antibody-rich ascites fluid from which the antibodies can be isolated and purified. The antibodies can be purified using techniques that are well known to one of ordinary skill in the art.

[0238] Any suitable immunoassay method may be utilized, including those which are commercially available, to determine the level of a polypeptide assayed according to the invention. Extensive discussion of the known immunoassay techniques is not required here since these are known to those of skill in the art. Typical suitable immunoassay techniques include sandwich enzyme-linked immunoassays (ELISA), radioimmunoassays (RIA), competitive binding assays, homogeneous assays, heterogeneous assays, etc.

[0239] For example, in the assays of the invention, “sandwich-type” assay formats can be used. An alternative technique is the “competitive-type” assay. In a competitive assay, the labeled probe is generally conjugated with a molecule that is identical to, or an analog of, the analyte. Thus, the labeled probe competes with the analyte of interest for the available receptive material. Competitive assays are typically used for detection of analytes such as haptens, each hapten being monovalent and capable of binding only one antibody molecule.

[0240] The antibodies can be labeled. In some embodiments, the detection antibody is labeled by covalently linking to an enzyme, label with a fluorescent compound or metal, label with a chemiluminescent compound. For example, the detection antibody can be labeled with catalase and the conversion uses a colorimetric substrate composition comprises potassium iodide, hydrogen peroxide and sodium thiosulphate; the enzyme can be alcohol dehydrogenase and the conversion uses a colorimetric substrate composition comprises an

alcohol, a pH indicator and a pH buffer, wherein the pH indicator is neutral red and the pH buffer is glycine-sodium hydroxide; the enzyme can also be hypoxanthine oxidase and the conversion uses a colorimetric substrate composition comprises xanthine, a tetrazolium salt and 4,5-dihydroxy-1,3-benzene disulphonic acid. In one embodiment, the detection antibody is labeled by covalently linking to an enzyme, label with a fluorescent compound or metal, or label with a chemiluminescent compound.

[0241] Direct and indirect labels can be used in immunoassays. A direct label can be defined as an entity, which in its natural state, is visible either to the naked eye or with the aid of an optical filter and/or applied stimulation, e.g., ultraviolet light, to promote fluorescence. Examples of colored labels which can be used include metallic sol particles, gold sol particles, dye sol particles, dyed latex particles or dyes encapsulated in liposomes. Other direct labels include radionuclides and fluorescent or luminescent moieties. Indirect labels such as enzymes can also be used according to the invention. Various enzymes are known for use as labels such as, for example, alkaline phosphatase, horseradish peroxidase, lysozyme, glucose-6-phosphate dehydrogenase, lactate dehydrogenase and urease.

[0242] The antibody can be attached to a surface. Examples of useful surfaces on which the antibody can be attached for the purposes of detecting the desired antigen include nitrocellulose, PVDF, polystyrene, and nylon.

[0243] In some embodiments of the processes, assays and methods described herein, detecting the binding of an antibody to a polypeptide includes contacting the sample with an antibody as described herein that specifically binds a signature peptide, forming an antigen-antibody complex between the antibody and the polypeptide present in the sample, washing the sample to remove the unbound antibody, adding a detection antibody that is labeled and is reactive to the antibody bound to the polypeptide in the sample, washing to remove the unbound labeled detection antibody and converting the label to a detectable signal, wherein the detectable signal is indicative of the quantity of the polypeptide in the sample. In some embodiments, the effector component is a detectable moiety selected from the group consisting of a fluorescent label, a radioactive compound, an enzyme, a substrate, an epitope tag, electron-dense reagent, biotin, digonigenin, hapten and a combination thereof. In some embodiments, the detection antibody is labeled by covalently linking to an enzyme, labeled with a fluorescent compound or metal, labeled with a chemiluminescent compound. The quantity of the polypeptide may be obtained by assaying a light scattering intensity resulting

from the formation of an antigen-antibody complex formed by a reaction of the polypeptide in the sample with the antibody, wherein the light scattering intensity of at least 10% above a control light scattering intensity indicates the likelihood of chemotherapy resistance.

Kits of the Invention

[0244] Various embodiments of the present invention provide a kit for quantifying a polypeptide in a sample. The kit comprises an internal standard of a signature peptide identified for the polypeptide according to a method as described herein; and instructions for using the internal standard to quantify the polypeptide in the sample. In various embodiments, the kit further comprises a protease for cleaving the polypeptide to yield the signature peptide. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17. In some embodiments, the kit comprises multiple internal standards. In some embodiments, the kit quantifies multiple polypeptides in a complex sample.

[0245] In accordance with the present invention, “a” should be construed to cover both the singular and the plural. In some embodiments, the kit targets a single polypeptide. In various embodiments, the kit includes one or more signature peptides for the single polypeptide.

[0246] In other embodiments, the kit targets multiple polypeptides (multiplexing). In some embodiments, the multiple polypeptides are related by their functions or pathways. When the kit targets multiple polypeptides, the kit includes multiple internal standards of multiple signature peptides for multiple polypeptides.

[0247] As a non-limiting example, for Uromodulin, a kit includes an internal standard for quantifying a UMOD signature peptide. In other examples, the kit would have signature peptides representing multiple target polypeptides or proteins, and the concentration of each signature peptide would be either identical, or balanced to approximate the concentration of the target polypeptides or proteins.

[0248] In various embodiments, the kit can be used for MRM assays for greater sensitivity. In some embodiments, the signature peptides is identified by SRM, and/or MRM, and/or SWATH.

[0249] In various embodiments, the kit further comprises an antibody specifically binding to the signature peptide. In certain embodiments, such a kit can be used for SISCAPA.

[0250] Various embodiments of the present invention provide a kit quantifying a polypeptide in a sample. The kit comprises: a protease for cleaving the polypeptide to yield a signature

peptide identified according to a method as described herein; an internal standard of the signature peptide; and instructions for using the protease and the internal standard to quantify the polypeptide in the sample. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17. In some embodiments, multiple polypeptides in a complex sample are quantified.

[0251] In various embodiments, the internal standard comprises the signature peptide labeled with a stable isotope. Examples of the stable isotope include but are not limited to ^{15}N (nitrogen-15), ^{13}C (carbon-13), and ^2H (deuterium). In various embodiments, the kit further comprises external standards. Examples of the external standards include but are not limited to a series of known concentrations of the polypeptide to be quantified. In various embodiments, the external standards can be used to generate a standard curve for quantifying the polypeptide in the sample.

[0252] Various embodiments of the present invention provide a kit quantifying a polypeptide in a sample. The kit comprises: an antibody specifically binding to a signature peptide identified according to a method as described herein; and instructions for using the antibody to quantify the polypeptide in the sample. Examples of quantitative assays based on the antibody include but are not limited to western blot, enzyme-linked immunosorbent assay (ELISA), radioimmunoassay and SISCAPA. In various embodiments, the kit further comprises external standards. Examples of the external standards include but are not limited to a series of known concentrations of the polypeptide to be quantified. In various embodiments, the external standards can be used to generate a standard curve for quantifying the polypeptide in the sample. In certain embodiments, the polypeptide is uromodulin, serum albumin or any one listed in Table 17. In certain embodiments, the signature peptide is any one listed in Tables 2, 13, or 17.

[0253] Various other embodiments of the present invention also provide for a kit for quantifying proteins of interest. The kit comprises stable isotope-labeled peptides and/or polypeptides matching the sequence of peptides with highly correlated signals; reagents to prepare a sample for mass spectrometry; and instructions for using said kit.

[0254] In some embodiments, the kit further comprises orthologous proteins from species other than the species to which the sample belongs as a control for digestion. For example, non-human protein and peptides (e.g., β -galactosidase and its corresponding SIL peptides)

can be included in the kit as a digestion control. In various embodiments, the SIL peptides are a pre-defined mixture appropriate for quantitation, approximating the concentration of peptide in a digested biological sample. In other words, the SIL peptides are provided at concentrations ranges that encompass target protein's levels generally detected in samples.

[0255] In various embodiments, the instructions describe target peptide and fragment masses for the signature peptide and internal standard (e.g., SIL peptides). In some embodiments, the instructions describe methods for achieving complete digestion, etc.

[0256] The present invention is also directed to a kit to quantify signature polypeptides in a sample. The kit is useful for practicing the inventive method of accurately quantifying correlated polypeptides. The kit is an assemblage of materials or components, including at least one of the inventive compositions. Thus, in some embodiments the kit contains a composition including the signature polypeptide, as described above.

[0257] The exact nature of the components configured in the inventive kit depends on its intended purpose. For example, some embodiments are configured for assaying different types of samples, such as but not limited to cells, tissues, body fluids, waters, food, terrain and/or synthetic preparations.

[0258] Instructions for use may be included in the kit. "Instructions for use" typically include a tangible expression describing the technique to be employed in using the components of the kit to effect a desired outcome, such as to identify and quantify polypeptides. Optionally, the kit also contains other useful components, such as, diluents, buffers, pharmaceutically acceptable carriers, syringes, catheters, applicators, pipetting or measuring tools, bandaging materials or other useful paraphernalia as will be readily recognized by those of skill in the art.

[0259] The materials or components assembled in the kit can be provided to the practitioner stored in any convenient and suitable ways that preserve their operability and utility. For example the components can be in dissolved, dehydrated, or lyophilized form; they can be provided at room, refrigerated or frozen temperatures. The components are typically contained in suitable packaging material(s). As employed herein, the phrase "packaging material" refers to one or more physical structures used to house the contents of the kit, such as inventive compositions and the like. The packaging material is constructed by well-known methods, preferably to provide a sterile, contaminant-free environment. The packaging materials employed in the kit are those customarily utilized in proteomics. As used herein,

the term “package” refers to a suitable solid matrix or material such as glass, plastic, paper, foil, and the like, capable of holding the individual kit components. Thus, for example, a package can be a glass vial used to contain suitable quantities of an inventive composition containing the signature peptides. The packaging material generally has an external label which indicates the contents and/or purpose of the kit and/or its components.

[0260] Many variations and alternative elements have been disclosed in embodiments of the present invention. Still further variations and alternate elements will be apparent to one of skill in the art. Among these variations, without limitation, are the selection of constituent modules for the inventive methods, compositions, kits, and systems, and the various conditions, diseases, and disorders that may be diagnosed, prognosed or treated therewith. Various embodiments of the invention can specifically include or exclude any of these variations or elements.

[0261] In some embodiments, the numbers expressing quantities of ingredients, properties such as concentration, reaction conditions, and so forth, used to describe and claim certain embodiments of the invention are to be understood as being modified in some instances by the term “about.” As one non-limiting example, one of ordinary skill in the art would generally consider a value difference (increase or decrease) no more than 5% to be in the meaning of the term “about.” Accordingly, in some embodiments, the numerical parameters set forth in the written description and attached claims are approximations that can vary depending upon the desired properties sought to be obtained by a particular embodiment. In some embodiments, the numerical parameters should be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. Notwithstanding that the numerical ranges and parameters setting forth the broad scope of some embodiments of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as practicable. The numerical values presented in some embodiments of the invention may contain certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

[0262] Groupings of alternative elements or embodiments of the invention disclosed herein are not to be construed as limitations. Each group member can be referred to and claimed individually or in any combination with other members of the group or other elements found herein. One or more members of a group can be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the

specification is herein deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

EXAMPLES

[0263] The following examples are provided to better illustrate the claimed invention and are not to be interpreted as limiting the scope of the invention. To the extent that specific materials are mentioned, it is merely for purposes of illustration and is not intended to limit the invention. One skilled in the art may develop equivalent means or reactants without the exercise of inventive capacity and without departing from the scope of the invention.

Example 1 An empirical approach to signature peptide choice for selected reaction monitoring: quantification of uromodulin in urine

[0264] There are many proposed avenues for a seamless transition between biomarker discovery data and selected reaction monitoring (SRM) assays for biomarker validation. Unfortunately, studies with the abundant urinary protein uromodulin showed that these methods do not converge on a consistent set of surrogate peptides for targeted MS. As an alternative, we present an empirical peptide selection workflow for robust protein quantitation.

[0265] The relative SRM signal intensity of 12 uromodulin-derived peptides was compared between tryptic digests of 9 urine specimens. Pairwise coefficients of variation between the 12 peptides ranged from 0.19 to 0.99. A correlation matrix was utilized to identify peptides that reproducibly track the amount of uromodulin protein. Four peptides with robust and highly-correlated SRM signals were selected. Absolute quantitation was performed using stable-isotope labeled versions of these peptides as internal standards and a standard curve prepared from a tryptic digest of purified uromodulin.

[0266] Absolute quantification of uromodulin in 40 clinical urine specimens yielded inter-peptide correlations of ≥ 0.984 and correlations of ≥ 0.912 with ELISA data. The SRM assays were linear over >3 orders of magnitude and had typical inter-digest CV's of $<10\%$, inter-injection CV's of $<7\%$, and inter-transition CV's of $<7\%$.

[0267] Comparing the apparent abundance of a plurality of peptides derived from the same target protein makes it possible to select signature peptides that are unaffected by the unpredictable confounding factors that are inevitably present in biological samples.

Urine samples

[0268] Pooled normal human urine and 10 urine samples from healthy males were purchased from Bioreclamation, Inc. Clinical urine specimens were obtained from 42 participants of the Atherosclerosis Risk in Communities (ARIC) study, detailed description of sample selection and characteristics was published (see e.g., The atherosclerosis risk in communities (ARIC) study: Design and objectives. The aric investigators. American journal of epidemiology 1989;129:687-702; and Kottgen A, Hwang SJ, Larson MG, Van Eyk JE, Fu Q, Benjamin EJ, et al. Uromodulin levels associate with a common UMOD variant and risk for incident ckd. J Am Soc Nephrol 2010;21:337-44).

Urine sample preparation

[0269] The sample preparation process is illustrated in Figure 5. To prepare urine for MS analysis, specimens stored at -80°C were thawed, gently mixed, and then centrifuged for 10 minutes at 10,000 x g at room temperature. 5 µl of urine was supplemented with 3 µl of NH₄HCO₃ (1M), 5 µl water, 2 µl RapiGest (1%), 2 µl of SIL peptides (1000 fmole/µl), and 0.16 µl of β-galactosidase (0.5 µg/µl), which was used as a quality control probe to monitor the consistency of sample processing and analysis. Proteins were reduced with 1 µl TCEP (100 mM) for 30 minutes at 60°C, alkylated in the dark with 1 µl iodoacetamide (50 mM) for 30 minutes at 37°C, and then incubated with 0.8 µl trypsin (0.125 µg/µl, Promega Gold) in a 37°C shaker for 6 hours. Digested peptides were purified on an HLB microplate and resuspended in MS loading buffer.

Mass spectrometry

[0270] SRM assays were performed on an LC/MS system comprising a high flow HPLC (Shimadzu Prominence) with an XBridge BEH 30 C18 reverse-phase column (Waters) linked to a triple quadrupole mass spectrometer (Q-Trap 6500 or Q-Trap 5500, Sciex) with a TurboV ion source (Sciex). A detailed description of SRM LC-MS/MS methods and parameters is provided herein. The SRM data was processed using Multiquant (Sciex).

[0271] Data-dependent MS experiments for discovery were performed on an Orbitrap Elite MS (Thermo Scientific, USA) coupled to an Easy-nLC 1000 chromatography system (Thermo Scientific, USA), and a TripleTOF® 5600 MS (Sciex) coupled to an Eksport nanoLC 415 chromatography system as described herein. Data was processed through SORCERER™ (Sage-N-Research Inc.), ProteinPilot™ (Sciex), or PASS (Integrated Analysis Inc.) software.

Quantitation of uromodulin

[0272] The absolute concentration of uromodulin was determined using stable isotope-labeled (SIL) peptides as internal standards and purified uromodulin (EMD Milipore) as an external standard, as described herein.

Peptide selection methods

[0273] For data-dependent LC MS/MS, a tryptic digest of purified uromodulin was analyzed on an Orbitrap MS, in both higher-energy collisional dissociation (HCD) and collision induced dissociation (CID) fragmentation modes, and on a Triple-TOF MS. Proteome Discoverer was used to search MS spectra files and rank peptides. Peptides are commonly ranked by intensity and spectral counting. These methods can give different results, so both were compared. The database methods involved searching human proteome databases from National Institute of Standards and Technology (NIST), PeptideAtlas, and SRMAtlas for uromodulin peptides. Predictions were obtained through the PeptideAtlas interface.

Optimization of urine sample preparation

[0274] Figure 5 presents an overview of the sample preparation workflow highlighting each parameter that was optimized to standardize the trypsin digestion and peptide cleanup procedures.

[0275] (a) Surfactants. Three different surfactants (0.1% RapiGest, 1% sodium deoxycholate (SDC) and 0.01% sodium dodecyl sulfate (SDS) were tested (Figure 11). All of the surfactants increased the SRM signal of the DSTIQ uromodulin peptide when compared with a no surfactant control. RapiGest provided the highest and most consistent response. Surfactants may help to disassemble large UMOD aggregates, thereby increasing the accessibility of trypsin cleavage sites, and may stabilize peptides after digestion. RapiGest has an additional advantage in that it degrades at low pH, so it doesn't interfere with MS like other detergents. In comparison with urea, which is generally used to denature proteins prior to trypsin digestion, surfactants do not modify proteins covalently and are added at a much lower concentration.

[0276] (b) Digestion time. The signals for two uromodulin peptides selected from data-dependent MS discovery data reached a plateau after 4-6 hours. Reduced signals detected after 16 hours in trypsin suggest that these uromodulin-derived peptides are either unstable or susceptible to cleavage by an endogenous protease. In the optimized procedure, urine was supplemented with RapiGest (0.01%) and digested with trypsin for 6 hours.

[0277] (c) Excess trypsin to overcome inhibitors in urine. To optimize trypsin digestion

conditions and insure that incomplete proteolysis did not compromise protein quantitation, pooled urine and a mixture of purified uromodulin and serum albumin were digested with varying amounts of trypsin and then analyzed with an SRM assay targeting 12 uromodulin peptides. In general, more trypsin was required to release peptides from the native uromodulin in urine than from the pure protein mix, even though there was twice as much uromodulin protein in the pure samples. This difference suggests that urine contains a trypsin inhibitor. The amount of this unidentified inhibitor could vary between urine specimens in an uncontrolled manner. For quantitative analysis, urine was digested with a three-fold excess over the amount of trypsin required for complete digestion of the most trypsin-resistant sites.

[0278] (d) Inconsistent results with peptides from protease-sensitive unfolded domains. The amount of trypsin required for complete release of different uromodulin peptides varied by more than 10-fold (Figure 6). As expected, the most trypsin-resistant peptides were derived from folded domains of the uromodulin protein (see Figure 1B). Notably, the three uromodulin peptides with the most disparately variable SRM signals were completely released by a low concentration of trypsin (Figure 6). These peptides may arise from unfolded regions of the protein that are sensitive to natural proteases in urine, which could have different activity in different individuals.

[0279] (e) Selecting HLB as the SPE resin for peptide desalting. The yield of uromodulin peptides after desalting on various SPE resins was evaluated using SIL peptides. HLB resin had the highest yield of the DSTIQVVENGESSQGR and SGSVIDQSR peptides (Figure 12A). Recovery of the SIL peptides from HLB resin was consistent for peptide concentrations ranging from 6.25 to 100 fmol/ μ l in 50 μ l urine (Figure 12B). Desalting on these SPE resins was performed following the manufacturers' suggested protocols. C4 and C18 OMIX Tips (Agilent) fit on a standard pipette. Liquid is passed through the resin by pipetting in and out. Tips were conditioned twice with 10 μ l 50% acetonitrile, 0.1% trifluoroacetic acid (TFA) and equilibrated twice with 10 μ l 0.1% TFA. SIL peptides were acidified with 0.1% TFA, loaded five times on the C4 or C18 resin, washed with 0.1% TFA, eluted with 75% acetonitrile, 0.5% formic acid, dried in a speed-vac, and then dissolved in MS loading buffer. For weak cation exchange (WCX), a 96 well microplate (Waters) was wetted with 200 μ l methanol, equilibrated with 200 μ l water, loaded with SIL peptides in 4% H₃PO₄, washed three times with 200 μ l of 25 mM KH₂PO₄/K₂HPO₄ (pH7), and washed again with 200 μ l methanol. Peptides were eluted with 50 μ l 2% formic acid in methanol,

dried in a speed vac, and resuspended in MS loading buffer. The HLB resin was wetted with 200 μ l methanol and then equilibrated three times with 200 μ l of 0.1% formic acid. SIL peptides in 200 μ l of 4% H₃PO₄ were loaded on the microplate, washed three times with 200 μ l 0.1% formic acid, and then slowly eluted with 200 μ l of 80% acetonitrile, 0.1% formic acid. The eluates were dried in a speed-vacuum and then dissolved in MS loading buffer.

[0280] (f) Normalization to SIL peptide internal standards. Theoretically, SIL peptides should behave identically to native peptides with the same sequence. Thus, any losses of native peptides during sample processing due to peptide instability, insolubility, or low yield after SPE should be accompanied by loss of an equal fraction of the SIL peptide. The utility of SIL peptides as internal standards was tested in an experiment where the desalting conditions were intentionally varied using techniques expected to affect peptide recovery (Figure 9). SIL peptides were added to a large batch of pooled urine, which was digested with trypsin and then divided into aliquots. Each aliquot was separately desalted under different conditions and then analyzed with an SRM assay tracking 6 peptides. As expected, the absolute and relative amounts of native peptides recovered varied tremendously (upper panel). However, a more consistent ratio was observed after normalization to the SIL peptide internal standards (lower panel). These results demonstrate that normalization is highly effective, and highlight the importance of consistent desalting procedures, which were employed in all other experiments.

[0281] (g) Spiked β -galactosidase as a probe for quality control. For quality control, urine samples were spiked with 0.08 μ g β -galactosidase protein and 2 pmol β -galactosidase SIL peptides prior to reduction, alkylation, trypsin digestion, and desalting. The consistency of sample processing was judged by comparing the ratio between digested natural peptide and SIL internal standard peptide for 3 tryptic peptides from β -galactosidase: WVGYGQDSR, IDPNAWVER, and GDFQFNIS. The %CVs for these three peptides were 16.9%, 4.9%, and 3.4%, respectively, in the experiment where uromodulin was quantified in 42 urine samples.

MS methods to identify detectable uromodulin peptides

[0282] The data-dependent acquisition MS experiment for initial peptide selection was performed on an Orbitrap XL mass spectrometer (ThermoFisher) with an on-line nano-HPLC system (1200 Series, Agilent Technologies). Peptides were separated on a reverse-phase analytical column packed with 10 cm of C18 beads (Biobasic C18 PicoFrit column, New Objective, Woburn, MA). A linear AB gradient comprising 5–60% B for 25 min was used

where solvent A was 0.1% formic acid and solvent B was 90% acetonitrile in 0.1% formic acid, followed by 100% B for 2 min. The flow rate was 300 nL/min. The instrument was operated in a data-dependent mode in which a full scan was followed by MS/MS scans of the five most intensive ions, which were automatically selected for collision-induced dissociation (CID). Data analysis was performed on a Sorcerer server using Sequest.

[0283] To compare peptides identifications, the same digested and desalted peptide mixture was run in duplicate on Orbitrap, Triple-TOF, and Triple-Quadrupole instruments. Specifically, the sample was analyzed using an Orbitrap Elite mass spectrometer (Thermo Scientific, USA) online coupled to an Easy-nLC 1000 system (Thermo Scientific, USA). The injection volume was 10 μ L of the sample, representing 0.2 μ g of peptides. After injection the samples were preconcentrated with 0.1% TFA on a trap column (Acclaim PepMap 100, 300 μ m $\times$ 5 mm, C18, 5 μ m, 100 $\text{\AA}$; maxiam pressure 800bar). Subsequently, the peptides were transferred to the analytical column (Acclaim PepMap RSLC, 75 μ m $\times$ 15 cm, nano Viper, C18, 2 μ m, 100 $\text{\AA}$) and separated by a 2% to 30% gradient over 70 mins (solvent A: 0.1% FA in water, solvent B: 0.1% FA in acetonitrile; flow rate 350 nL/min; column oven temperature 45°C). The MS was operated in a data-dependent mode. Full scan MS spectra were acquired at a resolution of 60,000 in the Orbitrap analyzer, followed by tandem mass spectra of the 20 most abundant peaks in the linear ion trap after peptide fragmentation by collision-induced dissociation (CID) or high-energy collision dissociation (HCD). For 5600 Triple-TOF, source conditions were as follows: Spray voltage was set to 2.3 kV, source gas was set to 15, curtain gas was set to 20, interface heater temperature was set to 160, and declustering potential was set to 100. Rolling collision energy was used for MS2 experiments and the 20 most abundant ions were selected for fragmentation. Peptides were loaded onto an Eksigent ekspertTM 415 nanoLC equipped with ekspertTM cHiPLC and ekspertTM nanoLC 400 autosampler. Samples were separated using a nano cHiPLC 200 μ m $\times$ 15 cm ChromXP C18-CL 3 μ m 120 $\text{\AA}$ column using a flow rate of 1000 nL/min and a linear gradient of 5-35% solvent B (0.1% formic acid in acetonitrile) for 123 min, 35-95% B for 3 minutes, holding at 95% for 10 minutes, then re-equilibration at 5% B for 15 minutes.

[0284] LC MS/MS data-dependent acquisition spectral data were searched on Mascot against a Human database and the results were imported into Proteome Discoverer, which allowed peptides to be ranked according to their intensity or spectral count. SEQUEST searches were conducted using the SORCERER platform by Sage-N. The human proteome database from

NIST was also imported into Proteome Discoverer. The SRM Atlas and PeptideAtlas online resources were queried for uromodulin. The consensus prediction amalgamates the results from five predictive algorithms, including STEPP (see e.g., Webb-Robertson et al., A support vector machine model for the prediction of proteotypic peptides for accurate mass and time proteomics, *Bioinformatics*. 2010 Jul 1;26(13):1677-83.)

MS methods for targeting uromodulin

[0285] SRM assays were performed on an LC/MS system with a reverse-phase column (XBridge BEH 30 C18 column, 2.1mmx100mm, 3.5 μ m, Waters, Milford, MA) plumbed into an HPLC (Shimadzu Prominence) linked to a triple quadrupole mass spectrometer (Q-Trap 6500 or Q-Trap 5500, Sciex) with a TurboV ion source (Sciex). Peptides (5 μ l) were injected in triplicate at a rate of 0.2 ml/min. The chromatography buffers were 0.1% formic acid (buffer A) and 95% acetonitrile in 0.1% formic acid (buffer B). The % buffer A increased from 18 to 27% over 7 minutes.

[0286] Uromodulin peptides and transitions were identified using Skyline software (see e.g., MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010;26:966-8), and then imported into Analyst 2.1 software (Sciex). An initial set of six transitions for each peptide was identified from the NIST spectral library. The best two to five of these were selected based upon signal intensity on a triple quadrupole MS. Synthetic stable-isotope peptides were obtained once the final peptides were selected. The collision energy and collision cell exit potential were then optimized using the Autotune function in Analyst with a continuous infusion of synthetic peptides.

[0287] Transitions (Table 7) were initially selected based upon high signal intensity. Two transitions were subsequently eliminated because they had overlapping interferences and/or misshapen peaks. Of note, some of the remaining transitions report on b2 or a2 fragment ions with short sequences and fragment $m/z < \text{parent } m/z$, making them prone to interference. However, we used these transitions because of their high signal intensity. To validate the fragment $m/z < \text{parent } m/z$ transitions, we showed that (1) they co-elute with fragment $m/z > \text{parent } m/z$ transitions from the same peptide, (2) have symmetrical peaks, (3) no spurious noise was observed, even in urine samples with low uromodulin concentrations and (4) the correlation between the measured amounts of different transitions from the same peptide in 9 urine samples, including with transitions having fragment $m/z > \text{parent } m/z$, was nearly

perfect ($r^2 > 0.995$).

Absolute quantification of uromodulin

[0288] The concentration of uromodulin was determined by comparison to a standard curve prepared from purified uromodulin through the use of stable isotope-labeled (SIL) internal standard peptides. The SIL peptides had a C-terminal [^{15}N]-Lys or [^{15}N]-Arg and were synthesized and HPLC-purified by New England Peptide. A mixture of ^{15}N peptides (3 nmoles each) from uromodulin (4 peptides), and β -galactosidase (3 peptides) was prepared in 20% acetonitrile, 0.1% formic acid and then divided into 100 pmoles aliquots. Each aliquot was dried in a speed-vacuum and then stored at -80°C until use. Peptides were re-suspended as a 10X stock (2 pmole/ μl) in 50 μl of MS loading buffer (20% acetonitrile, 0.1% formic acid, and 15 $\mu\text{g/ml}$ glucagon). Glucagon was included as a carrier to stabilize low concentration peptides.

[0289] Standard curves were prepared from human uromodulin purified from pooled urine (EMD Millipore, marketed as Human Tamm-Horsfall Glycoprotein) and recombinant β -galactosidase (Sigma). The concentrations of these proteins were determined by the manufacturers. 100 pmoles each of protein were dissolved in 150 mM NH_4HCO_3 with 0.1% RapiGest (Waters). The proteins were reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP, Pierce) for 30 minutes at 60°C , alkylated in the dark with 5 mM iodoacetamide for 30 minutes at 37°C , and then incubated overnight with 1.5 μg Trypsin (Promega Gold) in a final volume of 50 μl in a shaker block at 37°C .

[0290] Digested peptides were desalted on an HLB microplate in a vacuum manifold (Waters). The HLB resin was wetted with 700 μl methanol and then equilibrated three times with 700 μl of 0.1% formic acid. The peptide solution was diluted to 300 μl in 0.1% formic acid, further acidified with 300 μl of 4% H_3PO_4 , loaded on the microplate, and then slowly aspirated through the HLB resin. The resin was washed three times with 0.1% formic acid and then slowly eluted with 400 μl of 80% acetonitrile, 0.1% formic acid. The eluates were dried in a speed-vacuum, dissolved at 1 pmole/ μl in MS loading buffer supplemented with 1X SIL peptide standards, and then serially diluted 1: $\sqrt{10}$ in MS buffer with 1X SIL peptide standards.

Reproducibility and recovery

[0291] Reproducibility and recovery of the SRM assay were established in a different laboratory with different lots of sample preparation reagents on a different MS instrument by

a different operator. These experiments tracked the same MS transitions using the same mixture of SIL internal standard peptides, the same LC method, and the same standard curve. The volume of urine digested for each sample was increased from 5 μ l to 20 μ l.

[0292] The reproducibility test compared pooled normal human urine with a pool of diseased urine created by mixing urine specimens with high uromodulin from the ARIC study. On five separate days, five samples from each pool were digested with trypsin, desalted, and analyzed on a Q-Trap 6500 MS. Inter-assay CV's were calculated by comparing pools that were run on five different days (Table 8, top). Intra-assay CV's were calculated by comparing five pools run on the same day (Table 8, middle). Total CV's were calculated from the sum of squares of the mean inter- and intra-assay CVs (Table 8, bottom) (see *e.g.*, Grant RP, Hoofnagle AN. From lost in translation to paradise found: Enabling protein biomarker method transfer by mass spectrometry. Clin Chem 2014;60:941-4). Total CVs were <20%, satisfying the best practice acceptance criterion (see *e.g.*, Lee JW, Devanarayan V, Barrett YC, Weiner R, Allinson J, Fountain S, et al. Fit-for-purpose method development and validation for successful biomarker measurement. Pharmaceutical research 2006;23:312-28).

[0293] SRM results with the four uromodulin peptides showed that diseased urine pool had a 2.5 – 3.0-fold higher uromodulin concentration than healthy urine (Table 9, top). Linearity and recovery were determined using mixtures having healthy to diseased ratios of 1:3, 1:1, and 3:1 (Table 9, bottom). For each mixture, an expected concentration for each peptide was calculated assuming a linear response. Observed and expected concentrations were then compared to calculate the percent recovery. The mean percent recovery of was 104%, with a standard deviation of 6%.

Commonly used signature peptide selection methods yield divergent results

[0294] The first major step in developing an SRM assay is to choose signature peptides for the quantitative analysis. In uromodulin-1, there are 27 predicted tryptic peptides with lengths in the useful range of between 6 and 21 amino acids (Figure 1A). From these, potential signature peptides were identified by data-dependent acquisition, database, and predictive methods. Remarkably, these methods yielded almost completely different results. No clear patterns emerge when comparing the top 10 uromodulin peptides selected using 12 different, but not entirely independent, peptide selection methods (Table 1). Urine matrix and the choice of algorithm for searching discovery data also had a profound influence on peptide ranking (Table 3). There was, however, modest overlap in the ranking of transitions based on

fragment ion intensity (Table 4). These results demonstrate that current peptide selection methods do not converge upon a consistent set of recommended peptides and transitions for quantitative analysis.

[0295] An exemplar sequence of uromodulin is shown as SEQ ID NO:82 below:

```

1      mgqpsltwml mvvvaswfit taatdtsear wcsechsnat ctedeavttc tcqegftgdg
61     ltcvdldeca ipgahnncsan sscvntpgsf scvcpegfrl spglgctdvd ecaepglshc
121    halatcvnvv gsylcvcpag yrgdgwhcec spgscgpgld cvpegdalvc adpcqahrtl
181    deywrsteyg egyacdtdlr gwyrfvqggg armaetcvpv lrcntaapmw lngthpssde
241    givsrkacah wsgheclwda svqvkcacagg yyvynltapp echlayctdp ssvegtceec
301    sidedcksnn grwhcqckqd fnitdislle hrlecgandm kvslgkcklk slgfdkvfmy
361    lsdsrscgfn drdnrdwvsv vtpardgpcg tvltrnetha tysntlylad eiiirdlnik
421    infacsypld mkvslktalq pmvsalnirv ggtgmftvrm alfqtpsytd pyqgssvtls
481    teaflvygtm ldggdlsrfa llmtncyatp ssnatdplky fiiqdrpht rdstiqvven
541    gessqgrfsv qmfrfagnyd lvylhcevy l cdtmnekckp tcsgrtrfrsg svidqsrvl n
601    lgpitrkgvq atvsrafssl gllkvwlp11 lsatl1t1tfq

```

Table 1: Comparison of SRM peptide selection methods^a

	Discovery						Database				Prediction Method	
	Triple TOF		Orbi HCD		Orbi CID		NIST library		PeptideAtlas	SRM Atlas	PeptideAtlas Consensus	STEPP
	Intensity	Count	Intensity	Count	Intensity	Count	Intensity	Count				
de ng	VGGTG	TALQP	VGGTG	VLNLG	SGSVI	DGPCG	DSTIQ	TALQP	FAGNY	VFMYL	DSTIQ	QDFNI
	SLGFD	INFAC	STEYG	SGSVI	INFAC	VGGTG	ACAHW	DSTIQ	DSTIQ	TALQP	TALQP	DSTIQ
#2	FSVQM	DGPCG	INFAC	VGGTG	FSVQM	STEYG	INFAC	INFAC	STEYG	STEYG	QDFNI	NETHA
#3	LECGA	YFIHQ	YFIHQ	DGPCG	YFIHQ	VLNLG	DWVSV	STEYG	VFMYL	VGGTG	NETHA	VWLPL
#4	DWVSV	VGGTG	FSVQM	TALQP	TALQP	DWVSV	STEYG	VFMYL	INFAC	QDFNI	STEYG	VLNLG
#5	GVQAT	SGSVI	DSTIQ	FSVQM	VFMYL	DSTIQ	VFMYL	DWVSV	FSVQM	DWVSV	DWVSV	TALQP
#6	VLNLG	VLNLG	VFMYL	STEYG	VLNLG	YFIHQ	FSVQM	YFIHQ	ACAHW	FALLM	VGGTG	AFSSL
#7	TALQP	DWVSV	TALQP	VFMYL	VGGTG	TALQP	VLNLG	VGGTG	VGGTG	FSVQM	AFSSL	DWVSV
#8	VFMYL	VFMYL	VLNLG	DWVSV	MAETC	VFMYL	DGPCG	TLDEY	DWVSV	VLNLG	VWLPL	SLGFD
#9	FVGQG	LECGA	DGPCG	INFAC	DWVSV	INFAC	MAETC	VLNLG	FVGQG	NETHA	VLNLG	VGGTG
#10												

a. Peptides are identified by the sequence of their first 5 amino acid residues. See Table 5 for the full sequence and amino acid numbers of each peptide.

Table 3: Effects of urine matrix and the search algorithm on peptide ranking of Orbitrap results after CID fragmentation and data-dependent acquisition

Rank	Intensity			Count		
	Pure UMOD*		Urine	Pure UMOD*		Urine
	Mascot	SEQUEST		Mascot	SEQUEST	
1	SGSVI	VLNLG	VLNLG	DGPCG	MAETC	VGGTG
2	INFAC	VGGTG	MAETC	VGGTG	STEYG	DGPCG
3	FSVQM	DGPCG	GDGWH	STEYG	DSTIQ	MAETC
4	YFIIQ	TALQP	SGSVI	VLNLG	VGGTG	DSTIQ
5	TALQP	MAETC	TALQP	DWVSV	TALQP	STEYG
6	VFMYL	KGVQA	VGGTG	DSTIQ	DWVSV	KGVQA
7	VLNLG	SGSVI	DWVSV	YFIIQ	VLNLG	TALQP
8	VGGTG	KACAH	INFAC	TALQP	INFAC	INFAC
9	MAETC	INFAC	LECGA	VFMYL	VFMYL	VLNLG
10	DWVSV	ACAHW	VFMYL	INFAC	SGSVI	DWVSV

*The same MS data (.RAW) file was searched with both Mascot and SEQUEST.

Table 4: Fragmentation comparison

	Triple Quad	Triple TOF	Orbi HCD	Orbi CID	NIST Library	Peptide Atlas	SRM Atlas	Fragmentation rank
TLDEYWR	y5	y5			y5	y5	y2	1
	y4	a2			a2	b2	y5	2
	y3	y4			b2	y4	b2	3
	(-H2O)	b2			y4	y3	y3	4
	b3	y3			y3	y2	y3	5
	y6	y6			y2	b4	y4	6
FVGQGGAR	y6	y6			y6	y6		1
	y7	a2			a2	b2		2
	y4	a1			a1	a2		3
	(-H2O)	b2			b2	b6		4
	y5	y7			b6	y4		5
		y4			y4	b4		6
DWVSVVTPAR	b2	y7	y7	y5	y5	y5	b2	1
	y7	b2	a2	y7	y4	y7	y7	2
	y5	a2	b2	y4	y7	b6	y5	3
	a2	y8	y8	b2	b6	b7	y3	4
	y8	y5	y5	b6	y3	y8	y4	5
	b3	y4	a3	y8	b7	b5	y6	6
YFIQDR	a2	y5	a1	y5	y5	y5	b2	1
	y5	a2	y5	y4	y4	y4	y2	2
	b2	y4	a2	b2	a2	b2	y5	3
	y4	b2	y4	a2	b2	a2	y3	4
	y3	a1	y6	b3	y3	y3	y4	5
	y6	y6	y1	y3	b3	b3	y3	6

An empirical workflow for SRM peptide selection

[0296] In order to identify the best signature peptides for quantifying uromodulin in urine, the first step was to eliminate peptides that were never detected by MS on any instrument, were not unique to uromodulin, or were located within a C-terminal region thought to be absent from the mature protein. Several peptides with methionine or cysteine residues, which are susceptible to in vivo and in vitro modifications affecting their m/z ratio were also eliminated. This process narrowed the original set of 27 theoretical peptides down to 12 candidates for further testing (Table 5).

Table 5: Summary of the peptide selection process

ide	Theoretical UMOD tryptic peptides	Round I-selecting 12 peptides (Why excluded from round I)	Round II-selecting final 4 peptides (why exculde from round II)	Final selection Isoform specificity
Y	27 tryptic peptides (6-21 a.a.)			Isoforms 1, 3, and 4
STEYG	R.TLDEYWR.S [178, 184] R.STEYGEYACDTDLR.G [185, 199]		Low signal and Cys	
FVGQG	R.FVGQGGAR.M [204, 211]			Isoforms 1, 2, and 4
MAETC	R.MAETCVPVLR.C [212, 221]			
ACAHW	K.ACAHWSGHCCCLWDASVQVK.A [246, 264]	3 Cys, conserved	low signal and MetOx	
WHCQC	R.WHCQCCK.Q [312, 317]	Low m/z, 2 cys		
FN	K.QDFNITDISLLEHR.L [318, 331]	No Spectra		
GA	R.LECGANDMK.V [332, 340]	MetOx		
FD	K.SLGFDK.V [350, 355]	Low m/z, not unique (published)		
VYL	K.VFMYLSDSR.C [356, 364]		low signal and MetOx	
FN	R.CSGFNDR.D [365, 371]	no spectra		
VS	R.DWVSVVTPAR.D [375, 384]			Universal peptide
DGPCG	R.DGPCGTVLTR.N [385, 394]			
NETHA	R.NETHATYSNTLYLADEIIIR.D [395, 414]	No Spectra	Cys, glycosylated tryptic site	
INFAC	K.INFACSYPLDMK.V [420, 431]	MetOx		
TALQP	K.TALQPMVVSALNIR.V [436, 448]		low signal and MetOx	
VGGTG	R.VGGTGMFTVR.M [449, 458]	MetOx		
1	R.FALLMTNCTYATPSSNATDPLK.Y [498, 518]	MetOx, high m/z		
	K.YFIIQDR.C [519, 525]			Universal peptide

2	R.DSTIQVVENGESSQGR.F [531, 546]		Low correlation	
IM	R.FSVQMFR.F [547, 553]		low signal and MetOx	
C	K.CKPTCSGTR.F [577, 585]	Post CT cleavage	not in the processed form	
I	R.SGSVIDQSR.V [588, 596]		low correlation and not in the processed form	
G	R.VNLGPITR.K [597, 605]	Post CT cleavage	not in the processed form	
AT	K.GVQATVSR.A [607, 614]	Post CT cleavage	not in the processed form	
AFSSL	R.AFSSLGLLK.V [615, 623]	Post CT cleavage	not in the processed form	
VWLPL	K.VWLPLLLSATLTLTFQ.- [624, 639]	No spectra, Post CT cleavage	not in the processed form	

[0297] A tryptic digest of purified uromodulin was used to identify a set of transitions for each peptide that had high and reproducible peak intensities on a triple quadrupole mass spectrometer. The digest was then repeatedly injected to optimize the collision energy for each transition. The resulting parameters were used to investigate the performance of the 12 candidates in urine matrices. After establishing robust procedures for trypsin digestion and peptide cleanup, each peptide was evaluated in a set of tryptic digests of urine specimens obtained from healthy individuals. For this initial analysis, raw area-under-the-peak measurements were compared without normalization.

[0298] The measured amounts of the uromodulin signature peptides used for quantifying uromodulin protein should be linearly related to the amount of input protein and to the amount of other well-behaved signature peptides. To identify peptides with this property, coefficients of determination (r^2) were calculated for pairwise comparisons between each of the 12 candidate peptides across 9 urine samples (Figure 1B). As expected, r^2 values for pairs of transitions from the same peptide were always >0.998 , indicating that any variations from true linearity were due to the effects of differences between individual urine samples on the overall detectability of specific peptides. In contrast, low correlations were observed between several pairs of peptides, indicating that at least one peptide in each of these pairs was not accurately reporting the protein concentration. The identity of the peptides with poor correlations could not have been predicted from SRM chromatograms, as all of the peptides had symmetrical and unambiguously quantifiable peaks with no indication of interference in all urine samples.

[0299] Notably, the peptides with the lowest correlations were highly accessible to trypsin digestion, suggesting that these peptides may be derived from regions of the protein that are sensitive to endogenous proteases that vary between individuals (Figure 6). Also, the poorly correlated SGSVIDQSR peptide, although routinely detected in urine and purified uromodulin, is thought to be located within a C-terminal propeptide associated with the GPI anchor and may be absent from the mature protein.

[0300] From the r^2 data, we selected a set of four signature peptides that were all highly correlated with each other, having r^2 values of at least 0.9. Two of these peptides, DWVSVVTPAR (DWVSV) and YFIIQDR (YFIIQ), were present in all uromodulin isoforms. The other two, TLDEYWR (TLDEY) and FVGQGGAR (FVGQG), can discriminate between isoforms (Figures 1A-1B and Figure 7). In making our selections, we

also considered the total SRM signal intensity of each peptide, background noise, LC retention time, and peak shape (Table 6). Additionally, four Met-containing peptides included in the empirical test had acceptable raw pairwise correlations, but were excluded because the extent of Met oxidation was highly variable (Figure 8).

Table 6: SRM response for 12 individual peptides

	Total from top 3 transitions		Best transition for each peptide		Selection
	Peptide sequence	SRM signal	Peptide fragment	SRM signal	
Uromodulin	TLDEYWR	222228	TLDEYWR y5	184155	Yes
	DGPC[+57]GTVLTR	210975	DGPC[+57]GTVLTR y8+2	132294	No
	YFIQDR	202797	YFIQDR y5	139368	Yes
	FVGQGGAR	189641	FVGQGGAR y6	168847	Yes
	DWVSVVTPAR	111931	DWVSVVTPAR y7	51717	Yes
	SGSVIDQSR	58652	SGSVIDQSR y5	33819	No/yes for comparison
	FSVQMFR	55306	FSVQMFR y4	22738	No
	STEYGEYAC[+57]DTDLR	47853	STEYGEYAC[+57]DTDLR y9	14321	No
	VFMYLDSR	46293	VFMYLDSR y7	24338	No
	MAETC[+57]VPVLR	28391	MAETC[+57]VPVLR y4	14637	No
	DSTIQVWENGESSQGR	13240	DSTIQVWENGESSQGR y5	4497	No/yes for comparison
	TALQPMV/SALNIR	5146	TALQPMV/SALNIR y9	3629	No/yes for comparison

Building a quantitative SRM assay

[0301] For absolute quantitation, SIL peptide versions of the empirically-selected uromodulin signature peptides were spiked into each trypsin digest and used to normalize the data. Our expectation was that the SIL peptides would behave similarly to natural peptides with the same sequence, such that any loss of natural peptides during sample processing would be accompanied by an equivalent loss of SIL peptides. Normalization was found to be remarkably effective in a test where peptide cleanup procedures were deliberately manipulated to alter peptide recovery (Figure 9). The SIL peptides were also used to further optimize the MS parameters (Table 7).

Table 7: SRM parameters

Protein	Q1 Mass (Da)	Q3 Mass (Da)	Acq. Time (ms)	Transition	DP (volts)	EP (volts)	CE (volts)	CXP (volts)
Uromodulin	491.7	768.3	30	TLDEYWR y5	67	10	21	15
	496.7	778.3	20	TLDEYWR [^] y5	51	10	21	4
	491.7	653.3	100	TLDEYWR y4	67	10	24	15
	496.7	663.3	60	TLDEYWR [^] y4	51	10	27	4
	491.7	524.3	100	TLDEYWR y3	67	10	26	15
	496.7	534.3	60	TLDEYWR [^] y3	51	10	27	4
	396.2	545.3	30	FVGQGGAR y6	41	10	19	4
	401.2	555.3	20	FVGQGGAR [^] y6	41	10	19	4
	396.2	644.4	70	FVGQGGAR y7	41	10	21	16
	401.2	654.4	60	FVGQGGAR [^] y7	41	10	21	16
	565.3	302.1	60	DWVSVTPAR b2	46	10	27	8
	570.3	302.1	30	DWVSVTPAR [^] b2	46	10	27	8
	565.3	729.4	60	DWVSVTPAR y7	72	10	26	15
	570.3	739.4	40	DWVSVTPAR [^] y7	46	10	25	4
	565.3	274.1	60	DWVSVTPAR a2	46	10	37	8
	570.3	274.1	50	DWVSVTPAR [^] a2	46	10	37	8
	565.3	828.5	80	DWVSVTPAR y8	72	10	26	15
	570.3	838.5	50	DWVSVTPAR [^] y8	46	10	25	18
	477.8	644.3	40	YFIQDR y5	66	10	22	15
	482.8	654.4	20	YFIQDR [^] y5	56	10	21	4
	477.8	311.1	40	YFIQDR b2	56	10	21	10
	482.8	311.0	20	YFIQDR [^] b2	56	10	21	10
	477.8	531.2	40	YFIQDR y4	66	10	22	15
	482.8	541.3	30	YFIQDR [^] y4	56	10	21	4
	858.5	1072.5	20	DSTIQVVENGESSQGR [^] y10	80	10	40	28
	858.5	973.5	20	DSTIQVVENGESSQGR [^] y9	80	10	43	25
	479.4	515.2	20	SGSVIDQSR [^] y4	80	10	25	14
	479.4	628.2	20	SGSVIDQSR [^] y5	80	10	25	18
	712.4	414.4	20	TALQPMVSALNIR [^] b4	76	10	29	32
	712.4	1010.4	20	TALQPMVSALNIR [^] y9	76	10	35	26
Galactosidase	534.3	286.1	15	WVGYGQDSR b2	51	10	23	8
	539.3	286.1	15	WVGYGQDSR [^] b2	51	10	23	8
	534.3	262.0	15	WVGYGQDSR-y2	51	10	37	8
	539.3	272.0	15	WVGYGQDSR [^] y2	51	10	37	8
	534.3	562.1	15	WVGYGQDSR-y5	51	10	27	6
	539.3	572.1	15	WVGYGQDSR [^] y5	51	10	27	6
	534.3	782.1	15	WVGYGQDSR-y7	51	10	25	6
	539.3	792.1	15	WVGYGQDSR [^] y7	51	10	25	6
	550.3	774.2	15	IDPNAWVER y6	61	10	33	8
	555.4	784.2	15	IDPNAWVER [^] y6	61	10	33	8
	550.3	871.2	15	IDPNAWVER y7	61	10	25	18
	555.4	881.2	15	IDPNAWVER [^] y7	61	10	25	18
	550.3	436.1	15	IDPNAWVER y7+2	61	10	23	8
	555.4	441.1	15	IDPNAWVER [^] y7+2	61	10	23	8
	542.3	262.1	15	GDFQFNISR y2	61	10	21	8
	547.3	272.1	15	GDFQFNISR [^] y2	61	10	21	8
	542.3	636.0	15	GDFQFNISR y5	61	10	25	12
	547.3	646.0	15	GDFQFNISR [^] y5	61	10	25	12
	542.3	764.2	15	GDFQFNISR y6	61	10	25	18
	547.3	774.2	15	GDFQFNISR [^] y6	61	10	25	18

[^] ¹⁵N-labeled amino acid residue at the C-terminus of a SIL peptide

[0302] On a standard curve constructed from a serial dilution of purified uromodulin, the SRM response for 12 abundant transitions representing the 4 signature uromodulin peptides was linear over at least 3 orders of magnitude, with a linearity of ≥ 0.998 (Figure 10). The lower limits of quantitation (LLOQ) ranged between 0.4-14.1 $\mu\text{g/ml}$ (Table 2). The upper limit of quantification for all transitions was greater than 446.4 $\mu\text{g/ml}$, the highest concentration tested. At 446.4 $\mu\text{g/ml}$ uromodulin, recoveries were nearly 100%, and CVs were $<5\%$.

Table 2: LLOQ and ULOQ of Selected Uromodulin Peptides

Signature Peptide			LLOQ ($\mu\text{g/ml}$)			ULOQ		
Peptide	Fragment	Linearity	$\mu\text{g/ml}$	% Recovery ^d	%CV	$\mu\text{g/ml}$	% recovery	% CV
DWVSVVTPAR	y7	1.000	4.5	94.2	6.0	446.4	99.5	1.3
	y8	1.000	1.4	105.0	8.4	446.4	96.2	1.2
	b2	0.998	1.4	97.9	8.0	446.4	100.8	2.0
	a2	1.000	1.4	86.0	6.2	446.4	99.4	1.7
FVGQGGAR	y6	0.999	14.1	109.1	14.4	446.4	100.0	4.6
	y7	1.000	4.5	102.0	8.6	446.4	100.0	3.6
TLDEYWR	y5	0.999	1.4	87.8	14.9	446.4	97.1	2.1
	y4	1.000	1.4	87.1	12.9	446.4	100.7	1.9
	y3	1.000	4.5	94.0	5.6	446.4	99.2	1.9
YFIQDR	y5	1.000	1.4	82.3	11.4	446.4	98.4	2.5
	y4	0.999	0.5	97.9	17.8	446.4	97.6	2.7
	b2	0.999	1.4	84.1	12.8	446.4	96.8	2.0

- Linearity was determined across an 8 point 1: $\sqrt{10}$ dilution series of purified uromodulin
- LLOQ, determined from the standard curve, is defined as the lowest concentration of calibrate at which recovery is $100\% \pm 20\%$ and $\text{CV} < 20\%$.
- ULOQ is defined as the highest concentration of the standard at which recover is $100\% \pm 20\%$ and $\text{CV} < 20\%$.
- Recovery was calculated by back-fitting data to the standard curve. For each data point, the concentration calculated using the linear equation of best fit was compared with the known amount of input protein.

Reproducibility and recovery

[0303] Uromodulin was quantified in pools of healthy and diseased serum to establish the reproducibility and recovery of the final method. For reproducibility, five aliquots of each pooled sample were processed on each of five different days. The inter-assay, intra-assay, and total CV's ranged from 1% - 13%, 1% - 11%, and 5%-13%, respectively (Table 8). For recovery, healthy and diseased serum was mixed at ratios of 1:3, 1:1, and 3:1. Recoveries ranged from 83% to 118%, with a mean and standard deviation of $104\% \pm 6\%$ (Table 9).

Table 8: Reproducibility of the SRM method: Inter-Assay, Intra-Assay, and Total CV's
Inter-Assay CV's^a

Sample type	Sample	DWWSWTPAR	FVGQGGAR	TLDEYWR	YFIIQDR
Healthy pool	1	4%	7%	6%	4%
	2	4%	8%	10%	3%
	3	3%	13%	3%	5%
	4	4%	7%	6%	4%
	5	1%	11%	8%	2%
	mean	3%	9%	7%	4%

Sample type	Sample	DWWSWTPAR	FVGQGGAR	TLDEYWR	YFIIQDR
ARIC pool	1	4%	6%	5%	7%
	2	5%	2%	7%	5%
	3	6%	3%	7%	4%
	4	5%	8%	6%	6%
	5	4%	6%	8%	6%
	mean	5%	5%	7%	6%

Intra-Assay CV's^b

Sample type	Day	DWWSWTPAR	FVGQGGAR	TLDEYWR	YFIIQDR
Healthy pool	1	4%	11%	5%	3%
	2	3%	10%	9%	1%
	3	3%	9%	7%	2%
	4	3%	4%	4%	7%
	5	5%	9%	7%	5%
	mean	4%	9%	6%	4%

Sample type	Day	DWWSWTPAR	FVGQGGAR	TLDEYWR	YFIIQDR
ARIC pool	1	3%	3%	4%	4%
	2	3%	4%	4%	5%
	3	1%	4%	6%	2%
	4	3%	3%	5%	4%
	5	4%	6%	4%	2%
	mean	3%	4%	5%	3%

Total CV's^c

Sample	DWWSWTPAR	FVGQGGAR	TLDEYWR	YFIIQDR
Healthy pool	5%	13%	9%	5%
ARIC pool	7%	8%	8%	7%

- Inter-assay CV's were established for each sample of each pooled samples across 5 days. Experiments were repeated with 5 individual pooled healthy or 5 pooled diseased (2),
- Intra-assay CV's were established from each day 5 healthy pooled or 5 healthy diseased pooled, experiments were repeated 5 days for each pool (2).
- $CV\ total = (\text{mean } CV^2_{intra} + \text{mean } CV^2_{inter})^{1/2}$ (2).

Table 9: Recovery

	DWVSVVTPAR		
Sample	Observed ^a (μg/ml)	Calculated ^b	Recovery ^c
15 μl healthy : 5 μl ARIC	10.0±0.6	9.8	102%±6%
10 μl healthy : 10 μl ARIC	12.7±0.8	12.5	101%±7%
5 μl healthy : 15 μl ARIC	14.9±1.5	15.3	98%±10%

		FVGQGGAR	
Sample	Observed (μg/ml)	Calculated	Recovery
15 μl healthy : 5 μl ARIC	27.4±3.3	27.3	101%±12%
10 μl healthy : 10 μl ARIC	36.9±3.7	34.5	101%±11%
5 μl healthy : 15 μl ARIC	45.9±2.8	41.7	98%±7%

		TLDEYWR	
Sample	Observed (μg/ml)	Calculated	Recovery
15 μl healthy : 5 μl ARIC	13.2±1.0	12.3	108%±8%
10 μl healthy : 10 μl ARIC	16.6±1.2	15.7	106%±8%
5 μl healthy : 15 μl ARIC	20.9±0.9	19.1	110%±5%

		YFIIQDR	
Sample	Observed (μg/ml)	Calculated	Recovery
15 μl healthy : 5 μl ARIC	10.8±1.0	10.8	101%±2%
10 μl healthy : 10 μl ARIC	14.4±0.5	14.4	100%±3%
5 μl healthy : 15 μl ARIC	17.8±0.5	17.9	100%±3%

- The observed concentrations were calculated using peptides as internal standards and purified uromodulin as an external standard, for each admixture sample, mean observed concentration is obtained from 4 replicate.
- Calculated concentration from each pool each determined by 25 samples (5 samples for 5 days) analyzed in Table 8
- Recovery= 100 x (observed/calculated)

The quantitative SRM Assay yields reproducible results comparable to an ELISA

[0304] The quantitative SRM assay was evaluated by measuring the uromodulin concentration in 42 urine specimens that had been previously analyzed using an ELISA assay (see *e.g.*, Kottgen A, Hwang SJ, Larson MG, Van Eyk JE, Fu Q, Benjamin EJ, et al. Uromodulin levels associate with a common UMOD variant and risk for incident ckd. J Am Soc Nephrol 2010;21:337-44). The absolute concentration for each peptide was calculated with reference to a standard curve prepared from data collected in the same sequence of MS runs. Three independent digests were prepared for each urine sample, and the SRM assay was run three times on each digest. Two urine specimens were eliminated from further analysis: one had a uromodulin concentration below the LLOQ, and the other was enriched for uromodulin isoforms 1 and 4 over isoforms 2 and 3, as shown by relatively high amounts of

the TLDEY and FVGQG peptides.

[0305] The results for the remaining 40 samples, acquired from a total of 360 MS runs, were internally consistent (Figures 2A-2C). Coefficients of variation (CV) comparing the three digests for each sample were typically <10%, and CV's comparing the three injections for each digest were typically <7%. CV's comparing peptide concentrations measured using different transitions were typically <10%, with a trend towards higher CV's for low concentration peptides.

[0306] Notably, the UMOD concentration determined by SRM was greater than that determined by ELISA. This discrepancy could be due to i) inconsistency in the documented concentration of the standards used for SRM and ELISA, and/or ii) reduced antibody binding to endogenous uromodulin due to interference from unknown matrix components or structural modifications (e.g. post-translational modifications, proteolysis) lying within one of the uromodulin epitopes. In addition, the calculated concentration of the isoform-discriminatory FVGQG peptide was consistently higher than that of the other peptides, suggesting that the purified uromodulin calibrator had a different ratio of isoforms than the clinical samples or lacked an interfering contaminant common to all urine specimens. Alternatively, the FVGQG peptide could have a different decay rate than the other peptides.

[0307] There was a strong correlation (≥ 0.98) between the calculated concentrations of the 4 uromodulin signature peptides (Figure 3). These results represents a significant improvement over the >0.90 correlations for these peptides observed during the peptide selection phase. This improvement was achieved by normalizing to the SIL internal standards, thereby controlling for variations in peptide recovery. In contrast to the superior results for the empirically selected signature peptides, normalized data for 3 peptides that had been previously selected from shotgun proteomics data correlated poorly with each other (r^2 0.28 – 0.70) and with the 4 empirically selected peptides (r^2 0.38 – 0.74). Significantly, there was also a high correlation between the SRM data for the four empirically selected peptides and results from an ELISA assay that had been performed 2 years earlier on the same samples (Figure 3). These results demonstrate that choosing signature peptides based on experimental results generates more reliable SRM data.

[0308] The accuracy of protein quantitation by SRM, SWATH, and other MS techniques is completely dependent upon the selection of appropriate surrogate peptides to represent the protein of interest. Empirically testing a plurality of candidate peptides to identify those with

correlated MS signals makes it possible to select peptides that will generate robust data in the real world. Reliance on other popular methods can lead to confounding results because unpredictable factors can interfere with accurate quantitation.

Using a correlation matrix to identify proteotypic peptides

[0309] In principle, when a protein is completely digested into peptides, the derivative peptides should be present in equimolar amounts. Thus, if one complex biological sample has twice as much of a protein of interest as another, it should, after proteolysis, have twice as much of every derivative peptide. Consequently, in a set of unknown biological samples, the measured amounts of two peptides derived from the same protein should have a linear relationship regardless of the amount of protein in each sample. If the relationship deviates from linearity for any reason, at least one of the peptides is not suitable for determining the concentration of the parent protein.

[0310] We propose an efficient workflow to select representative peptides for absolute MS quantitation of a target protein (Figure 4). The process begins by identifying the set of all potential peptides from an amino acid sequence that are within a detectable m/z range. If the goal of the experiment is to monitor a specific PTM, proteolytic cleavage, isoform, or mutation, peptides representing the desired feature must be retained. Otherwise, the initial set can be trimmed by eliminating peptides that are not present in all forms of the protein to be quantified. Peptides subject to oxidation and other in vitro artifacts should also be eliminated, if possible.

[0311] Preliminary SRM assays are designed to target as many peptides as practically possible and then tested in biological samples representative of the milieu that will be used for quantitative assays. If the peptide is readily detected, these preliminary assays don't have to be fully optimized for MS performance or absolute quantitation, and they can be developed using purified protein, enriched protein or native biological samples. The goal is to quickly measure the relative amounts of each peptide in the full range of appropriate biological samples. A coefficient of correlation (r^2) is calculated for each pair of peptides and then arranged in a matrix, making it possible to identify a subset of well-behaved peptides that all have relatively high correlation scores with each other. The final signature peptides can then be selected based on practical criteria including signal strength and LC elution time.

[0312] There are many potential reasons for the measured amount of a peptide to vary from expectation. Differences in the chemical composition, pH, or ionic strength of the biological

matrix can influence proteolysis, peptide stability, aggregation, or ionization in an MS instrument. Oxidation and other artifactual chemical modifications can change the mass of a peptide and thereby interfere with MS detection. Peptide mass can also be affected by unknown PTMs or polymorphisms. In addition, background noise could arise from unknown components in the biological matrix. By following the proposed workflow, peptides with poor correlations can be readily identified using a correlation matrix and then expeditiously eliminated without actually determining precisely why they are unsuitable for quantitation.

Limitations of previous peptide selection methods

[0313] The most important concept arising from this work is that one cannot take shortcuts in peptide selection and expect to be rewarded with a robust assay. A variety of common peptide selection methods were tested and gave wildly inconsistent results. Notably, 14 different uromodulin peptides were ranked among the top three by one or more methods (Table 1; see also Table 3), but none of these “top 3” peptides were included in the empirically derived SRM assay (Table 7). The most commonly recommended peptide, DSTIQVVENGESSQGR, with 6 different endorsements, had a low SRM signal and a relatively low correlation with other uromodulin peptides. Five other top 3 peptides, including two recommended by SRM Atlas, contained methionine residues, which can have a high degree of variability in the percentage of oxidation. Additionally, two top 3 peptides predicted by purely computational methods were not detected on any MS instruments.

Comparing SRM and ELISA assays

[0314] All four uromodulin peptides in our final assay yielded quantitative SRM results comparable to those obtained with an ELISA (Figure 3). The correlation between different peptides measured by SRM was somewhat higher than the correlation with the ELISA data. This difference may arise because the same tryptic digests were used for all peptides in the SRM assay, whereas the ELISA was performed 2 years earlier (see *e.g.*, Kottgen A, Hwang SJ, Larson MG, Van Eyk JE, Fu Q, Benjamin EJ, et al. Uromodulin levels associate with a common umod UMOD variant and risk for incident ckd. J Am Soc Nephrol 2010;21:337-44).

[0315] SRM assays have several advantages over ELISAs. Most importantly, ELISAs are completely dependent upon antibodies. It takes a long time to produce antibodies with sufficient affinity and specificity, and their corresponding epitopes may be suboptimal for quantitation due to incomplete accessibility, interferences, or variation between protein forms. These concerns are magnified by the fact that epitopes are not even disclosed for the

commercially available ELISA assays targeting uromodulin. Furthermore, SRM assays are more flexible than ELISAs, as they can target multiple peptides including ones that discriminate between isoforms and post-translational modifications.

[0316] In conclusion, the empirical peptide selection workflow described in this paper is useful to identify signature peptides for quantitative MS assays that are demonstrably free from unpredictable artifacts that could interfere with accurate and reproducible quantitation.

Example 2 Peptide selection from SWATH data

[0317] Human aorta tissue was from the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study, an investigation of atherosclerotic lesions (Pathobiological Determinants of Atherosclerosis in Youth (PDAY) Research Group, Natural history of aortic and coronary atherosclerotic lesions in youth. Findings from the PDAY Study, *Arterioscler Thromb.* 1993 Sep;13(9):1291-8). Proteins from 15 aortas were extracted by grinding with a mortar and pestle in 8M urea, 2M Thiourea, 4% CHAPS and 1% DTT. Samples were diluted to 0.8M urea with 100 mM NH_4HCO_3 buffer at pH 8.0 and digested overnight with trypsin. After digestion the samples were desalted by solid phase extraction on a 30 mg Oasis® HLB plate.

MS Data Acquisition

[0318] Chromatography: Peptides from 4 μg aortic protein were separated on a NanoLC™ 415 System (SCIEX) operating in trap-elute mode at microflow rates. A 0.3x150 cm ChromXP™ column (SCIEX) was used with a short gradient (3-35% solvent B in 60 min, B: 100% ACN, 0.1 formic acid in water) at 5 $\mu\text{L}/\text{min}$ (total run time 75 min).

[0319] Mass Spectrometry: The MS analysis was performed on a TripleTOF® 6600 system (SCIEX) using a DuoSpray Source with a 25 μm I.D. hybrid electrodes (SCIEX). Variable window SWATH® Acquisition methods were built using Analyst® TF Software 1.7. 100 Q1 window across the mass range (400-1250) isolation for improved data quality through increased specificity. Variable sized Q1 windows optimized based on precursor density further increased specificity while ensuring broad mass range coverage.

[0320] Data-Independent Acquisition data analysis: Spectral library generation from data-dependent acquisition MS: Profile-mode .wiff files from shotgun data acquisition were converted to mzML format using the AB Sciex Data Converter (in proteinpilot mode) and then re-converted to mzXML format using ProteoWizard v.3.0.6002 (Kessner et al, 2008) for peaklist generation. The MS2 spectra were queried against the reviewed canonical Swiss-Prot

Human complete proteome appended with iRT protein sequence and shuffled sequence decoys (Elias & Gygi, 2007). All data were searched using the X!Tandem Native v.2013.06.15.1, X!Tandem Kscore v.2013.06.15.1 (Craig & Beavis, 2004) and Comet v.2014.02 rev.2 (Eng et al, 2012). The search parameters included the following criteria: static modifications of Carbamidomethyl (C) and variable modifications of Oxidation (M), Phosphorylation (STY). The parent mass tolerance was set to be 50 p.p.m, and mono-isotopic fragment mass tolerance was 100 p.p.m (which was further filtered to be < 0.05 Da for building spectral library); tryptic peptides with up to two missed cleavages were allowed. The identified peptides were processed and analyzed through Trans-Proteomic Pipeline v.4.8 (Keller et al, 2005) and was validated using the PeptideProphet (Keller et al, 2002) scoring. The PeptideProphet results were statistically refined using iProphet (Shteynberg et al, 2011). All the peptides were filtered at a false discovery rate (FDR) of 1% with a peptide probability cutoff ≥ 0.99 . The raw spectral libraries were generated from all valid peptide spectrum matches and then refined into non-redundant consensus libraries (Collins et al, 2013) using SpectraST v.4.0 (Lam et al, 2007). For each peptide, the retention time was mapped into the iRT space (Escher et al, 2012) with reference to a linear calibration constructed for each shotgun run as previously described (Collins et al, 2013). The MS assays, constructed from the Top six most intense transitions (from ion series: b and y and charge states: 1,2) with Q1 range from 400 to 1,200 m/z excluding the precursor SWATH window, were used for targeted data analysis of SWATH maps.

[0321] Targeted data analysis for SWATH-MS: SWATH-MS.wiff files from the data-independent acquisition were first converted to profile mzML using ProteoWizard v.3.0.6002 (Kessner et al, 2008). The whole process of SWATH-targeted data analysis was carried out using OpenSWATH v.2.0.0 (Rost et al, 2014) running on an internal computing cluster. OpenSWATH utilizes a target-decoy scoring system (PyProphet v.0.13.3) such as mProphet to estimate the identification of FDR. The best scoring classifier that was built from the sample of most protein identifications was utilized in this study. Based on our final spectral library, OpenSWATH firstly identified the peak groups from all individual SWATH maps at a global peptide FDR of 1% and aligned them between SWATH maps based on the clustering behaviors of retention time in each run with a non-linear alignment algorithm (Weisser et al, 2013). For this analysis, the MS runs were realigned to each other using LOcally WEighted Scatterplot Smoothing method and the peak group clustering was performed using

“LocalMST” method. Specifically, only those peptide peak groups that deviate within 3 standard deviations from the retention time were reported and considered for alignment with the max FDR quality of 5% (quality cutoff to still consider a feature for alignment). Next, to obtain a high-quality quantitative data at the protein level, we discarded those proteins whose peptides were shared between multiple different proteins (non-proteotypic peptides) (Mallick et al, 2007). Quantitative peptide and protein level summary outputs were then used for all downstream biological analysis.

Selection of highly-correlated signature peptides

[0322] Transition selection. Prism software was used to calculate coefficients of determination between all possible pairs of the six transitions for each peptide. A correlation matrix was constructed, and the mean correlation for each peptide was calculated. Correlations were generally $r^2 > 0.85$. Any transition with a mean correlation 10% below the average mean for all transitions of the peptide was discarded. If any transitions were discarded, a revised correlation matrix was constructed and the mean correlations were recalculated.

[0323] Transitions were also ranked by mean peak area. The transition with the highest mean peak area was selected as the signature transition for the peptide if its mean correlation was within 5% of the highest mean correlation. If not, the transition having the highest peak area and also having a mean correlation within 5% of the highest mean correlation was selected.

[0324] Correlation matrix analysis. A separate correlation matrix was created for each protein of interest. All quantifiable peptides derived from the protein were represented by the peak area from a single signature transition. Prism software was used to calculate coefficients of determination between all possible peptide pairs. The correlation data was transferred to a Microsoft Excel spreadsheet, and an average correlation was determined for each peptide.

[0325] Peptide selection for Serum Albumin. Serum albumin was selected as an exemplary protein to investigate the versatility of the peptide selection methodology because it is well studied in quantitative SRM assays. The PDAY SWATH dataset contains quantified peaks from 63 serum albumin peptides. Table 10 presents a truncated version of a 63x63 matrix of pairwise correlations between these peptides. Columns 5-9 show pairwise correlations for 5 exemplary peptides. Column 10 shows the average of pairwise correlations between the peptide shown in column 2 and the other 62 peptides.

Table 10

Peak Area	Peptide sequence ^a	z	Frag Ion	QTALV	LVNEV	VFDEF	LVAAS	DDNPN	Ave r ²
218306	QTALVELVK	2	y5		0.958	0.941	0.933	0.916	0.937
136844	SHC(CAM)IAEVENDEM(Ox)PA DLPSLAADFVESK	3	b3	0.979	0.912	0.945	0.879	0.866	0.916
739992	LVNEVTEFAK	2	y8	0.958		0.905	0.906	0.934	0.926
441926	RPC(CAM)FSALEVDETYVPK	3	b6	0.950	0.903	0.932	0.888	0.860	0.907
114067	SHC(CAM)IAEVENDEMPADLP SLAADFVESK	3	y11	0.957	0.904	0.936	0.823	0.851	0.894
366357	VFDEFKPLVEEPQNLIK	3	y6	0.941	0.905		0.859	0.876	0.895
98707	QNC(CAM)ELFEQLGEYK	2	y4	0.935	0.942	0.899	0.906	0.899	0.916
238541	AVMDDFAAFVEK	2	y9	0.951	0.878	0.922	0.862	0.834	0.890
56772	RMPC(CAM)AEDYLSVVLNQL C(CAM)VLHEK	4	b7	0.946	0.879	0.868	0.921	0.880	0.899
21997	KQTALVELVK	2	y8	0.912	0.933	0.855	0.847	0.923	0.894
28179	VHTEC(CAM)C(CAM)HGDLLE C(CAM)ADDR	4	y4	0.913	0.944	0.910	0.850	0.944	0.912
419165	LC(CAM)TVATLR	2	y6	0.890	0.900	0.940	0.860	0.855	0.889
282837	LVRPEVDVMC(CAM)TAFHDNE ETFLKK	4	b7	0.949	0.882	0.869	0.813	0.806	0.864
7661	EC(CAM)C(CAM)EKPLLEK	3	y5	0.943	0.946	0.838	0.848	0.898	0.895
53763	LVAASQAALGL	2	b8	0.933	0.906	0.859		0.927	0.906
116552	DDNPNLPR	2	y5	0.916	0.934	0.876	0.927		0.913
501275	FQNALLVR	2	y6	0.918	0.950	0.842	0.850	0.869	0.886
19645	VPQVSTPTLVEVSR	2	y8	0.920	0.971	0.841	0.865	0.897	0.899
18838	RHPYFYAPELLFFAK	3	b5	0.918	0.886	0.831	0.796	0.783	0.843
243203	AEFAEVSK	2	y6	0.883	0.875	0.902	0.857	0.884	0.880
69408	SLHTLFGDK	2	y7	0.866	0.918	0.911	0.809	0.858	0.873
14853	LVRPEVDVM(Ox)C(CAM)T(p)A FHDNEETFLKK	5	b7	0.903	0.837	0.888	0.859	0.868	0.871
7550	RMPC(CAM)AEDY(p)LSVVLNQL LC(CAM)VLHEK	4	y3	0.880	0.923	0.811	0.897	0.930	0.888
368829	KVPQVSTPTLVEVSR	3	y4	0.894	0.888	0.923	0.739	0.777	0.844
2716	KYLYEIAR	2	y6	0.885	0.896	0.902	0.832	0.865	0.876
131057	TYETTLEK	2	y6	0.889	0.921	0.821	0.877	0.953	0.892
227882	RHPDYSVVLRLR	3	y5	0.904	0.877	0.820	0.780	0.716	0.819
15226	HPDYSVVLRLR	3	y4	0.877	0.849	0.886	0.928	0.842	0.876
4327	KLVAASQAALGL	2	b9	0.897	0.831	0.894	0.744	0.726	0.818
191879	LVRPEVDVMC(CAM)TAFHDNE ETFLK	4	b7	0.862	0.793	0.914	0.806	0.737	0.822
65813	FKDLGEENFK	3	y4	0.857	0.874	0.945	0.808	0.841	0.865
206640	YLYEIAR	2	y5	0.878	0.830	0.901	0.752	0.725	0.817
38305	VHTEC(CAM)C(CAM)HGDLLE C(CAM)ADDRADLAK	5	b9	0.840	0.776	0.919	0.844	0.850	0.846
4685	NEC(CAM)FLQHKDDNPNLPR	4	y3	0.855	0.779	0.853	0.804	0.806	0.820
21573	AAFTEC(CAM)C(CAM)QAADK	2	y7	0.821	0.866	0.807	0.816	0.891	0.840

17678	ETYGEMADC(CAM)C(CAM)AK	2	y7	0.826	0.773	0.897	0.773	0.790	0.812
7105	QEPERNEC(CAM)FLQHKDDNP NLPR	5	y4	0.858	0.798	0.858	0.878	0.892	0.857
22522	YIC(CAM)ENQDSISSK	2	y10	0.823	0.872	0.797	0.823	0.939	0.851
1836	ADDKETC(CAM)FAEEGK	3	y5	0.827	0.891	0.779	0.813	0.953	0.853
2995	NEC(CAM)FLQHK	2	y6	0.821	0.903	0.780	0.833	0.831	0.834
35989	C(CAM)C(CAM)TESLVNR	2	y7	0.801	0.797	0.753	0.819	0.808	0.795
30991	LAKT(p)Y(p)ET(p)TLEKC(CAM) C(CAM)AAADPHEC(CAM)YAK	4	y7	0.786	0.737	0.871	0.759	0.788	0.788
41379	M(Ox)PC(CAM)AEDYLSVVLNQ LC(CAM)VLHEK	3	y3	0.822	0.782	0.703	0.769	0.648	0.745
41376	MPC(CAM)AEDYLSVVLNQL C(CAM)VLHEK	3	y3	0.821	0.780	0.702	0.768	0.646	0.743
5762	HPYFYAPELLFFAK	3	b4	0.780	0.758	0.772	0.691	0.614	0.723
3066	SLHTLFGDKLC(CAM)TVATLR	4	y4	0.772	0.769	0.639	0.879	0.819	0.776
21003	LKEC(CAM)C(CAM)EKPLLEK	3	y5	0.740	0.830	0.650	0.745	0.882	0.769
3262	ETYGEM(Ox)ADC(CAM) C(CAM)AK	2	y6	0.753	0.749	0.851	0.714	0.795	0.772
37325	ALVLIIFAQYLQQC(CAM)PFED HVK	3	y7	0.747	0.637	0.745	0.563	0.464	0.631
6779	ADDKETC(CAM)FAEEGKK	4	y6	0.729	0.671	0.749	0.775	0.712	0.727
35973	EFNAETFTFHADIC(CAM)TLSEK	3	y9	0.720	0.631	0.743	0.595	0.532	0.644
54535	DVFLGMFLYEYAR	2	y9	0.715	0.580	0.737	0.557	0.459	0.609
4731	LDELRLDEGK	2	b6	0.657	0.627	0.688	0.630	0.652	0.651
7212	C(CAM)C(CAM)AAADPHE C(CAM)YAK	3	y7	0.677	0.600	0.685	0.663	0.618	0.649
2732	RM(Ox)PC(CAM)AEDYLSVVLN QLC(CAM)VLHEK	4	b7	0.589	0.651	0.420	0.605	0.564	0.566
4755	LVRPEVDVM(Ox)C(CAM)TAFH DNEETFLK	4	b7	0.516	0.594	0.414	0.505	0.436	0.493
6271	S(p)HC(CAM)IAEVENDEM(Ox) PADLPSLAADFVESK	4	y7	0.536	0.459	0.577	0.262	0.247	0.416
5672	AVM(Ox)DDFAAFVEK	2	y4	0.463	0.562	0.314	0.518	0.456	0.463
1422	NYAEAKDVFLGMFLYEYAR	3	y5	0.454	0.495	0.398	0.320	0.482	0.430
4000	LVRPEVDVM(Ox)C(CAM)TAFH DNEETFLKK	5	b7	0.426	0.489	0.280	0.464	0.342	0.400
5581	TC(CAM)VADESAENC(CAM)DK	2	y10	0.389	0.391	0.451	0.288	0.395	0.383
4571	DVFLGM(Ox)FLYEYAR	2	b3	0.372	0.352	0.284	0.299	0.123	0.286
3171	EFNAETFTFHADIC(CAM)TLSEK ER	4	y5	0.322	0.182	0.175	0.447	0.386	0.302
Average coefficient of determination (r^2)				0.799	0.784	0.771	0.751	0.748	

^a Abbreviations: z, charge; (CAM), Carbamidomethylated; (Ox), Oxidized; (P), Phosphorylated

[0326] Peptides containing methionine residues, missed cleavages, and/or phosphorylations were excluded, resulting in a 26x26 matrix of pairwise correlations. The peptides in this

matrix were sorted again by the average of their correlations. Table 11 presents a truncated version of this matrix.

Table 11.

Average Peak Area	Peptide sequence ^a	z	Frag Ion	QTALV	LVNEV	VFDEF	DDNPN	FQNAL	LVAAS	SLHTL	AEFAE	Ave r ²
218306	QTALVELVK	2	y5		0.958	0.941	0.916	0.918	0.933	0.866	0.883	0.848
739992	LVNEVTEFAK	2	y8	0.958		0.905	0.934	0.950	0.906	0.918	0.875	0.844
366357	VFDEFKPLVEEPQNLIK	3	y6	0.941	0.905		0.876	0.842	0.859	0.911	0.902	0.829
98707	QNC(CAM)ELFEQLGEYK	2	y4	0.935	0.942	0.899	0.899	0.920	0.906	0.895	0.794	0.822
28179	VHTEC(CAM)C(CAM)HGDLLC(CAM)ADDR	4	y4	0.913	0.944	0.910	0.944	0.905	0.850	0.896	0.905	0.821
441926	RPC(CAM)FSALEVDETYVPK	3	b6	0.950	0.903	0.932	0.860	0.896	0.888	0.883	0.927	0.817
419165	LC(CAM)TVATLR	2	y6	0.890	0.900	0.940	0.855	0.868	0.860	0.946	0.886	0.811
116552	DDNPNLPR	2	y5	0.916	0.934	0.876		0.869	0.927	0.858	0.884	0.809
501275	FQNALLVR	2	y6	0.918	0.950	0.842	0.869		0.850	0.865	0.821	0.806
19645	VPQVSTPTLVEVSR	2	y8	0.920	0.971	0.841	0.897	0.960	0.865	0.855	0.843	0.803
53763	LVAASQAALGL	2	b8	0.933	0.906	0.859	0.927	0.850		0.809	0.857	0.800
7661	EC(CAM)C(CAM)EKPLLEK	3	y5	0.943	0.946	0.838	0.898	0.935	0.848	0.810	0.823	0.800
69408	SLHTLFGDK	2	y7	0.866	0.918	0.911	0.858	0.865	0.809		0.853	0.794
243203	AEFAEVSK	2	y6	0.883	0.875	0.902	0.884	0.821	0.857	0.853		0.790
131057	TYETTLK	2	y6	0.889	0.921	0.821	0.953	0.927	0.877	0.805	0.830	0.786
15226	HPDYSVLLLR	3	y4	0.877	0.849	0.886	0.842	0.800	0.928	0.864	0.776	0.770
21573	AAFTEC(CAM)C(CAM)QAADK	2	y7	0.821	0.866	0.807	0.891	0.833	0.816	0.811	0.900	0.760
22522	YIC(CAM)ENQDSISSK	2	y10	0.823	0.872	0.797	0.939	0.838	0.823	0.818	0.829	0.755
206640	YLYEIR	2	y5	0.878	0.830	0.901	0.725	0.817	0.752	0.789	0.835	0.743
2995	NEC(CAM)FLQHK	2	y6	0.821	0.903	0.780	0.831	0.828	0.833	0.825	0.669	0.737
35989	C(CAM)C(CAM)TESLVNR	2	y7	0.801	0.797	0.753	0.808	0.668	0.819	0.729	0.726	0.703
5762	HPYFYAPELLFFAK	3	b4	0.780	0.758	0.772	0.614	0.820	0.691	0.769	0.586	0.676
35973	EFNAETFTFHADIC(CAM)TLSEK	3	y9	0.720	0.631	0.743	0.532	0.571	0.595	0.656	0.597	0.596
37325	ALVLIAFAQYLQQC(CAM)PFEDHVK	3	y7	0.747	0.637	0.745	0.464	0.623	0.563	0.626	0.608	0.587
7212	C(CAM)C(CAM)AAADPHEC(CAM)YAK	3	y7	0.677	0.600	0.685	0.618	0.447	0.663	0.539	0.677	0.555
5581	TC(CAM)VADESAENC(CAM)DK	2	y10	0.389	0.391	0.451	0.395	0.381	0.288	0.264	0.468	0.355
Average coefficient of determination (r ²)				0.848	0.844	0.829	0.809	0.806	0.800	0.794	0.790	

^a Abbreviations: z, charge; (CAM), Carbamidomethylated; (Ox), Oxidized; (P), Phosphorylated

[0327] Next, an iterative process was employed to remove peptides with low correlations. The peptide with the lowest average correlation was excluded. Then, the correlation matrix was resorted. This was repeated 6 times until the lowest average correlation was >0.78. After each poorly correlated peptide was removed from the matrix, the average correlations for the remaining peptides increased. Table 12 presents a portion of the data from this matrix.

Table 12

Peak Area	Peptide sequence ^a	z	Frag Ion	LVNEV	QTALV	DDNPN	FQNAL	VFDEF	LVAAS	SLHTL	AEFAE	Ave r ²
739992	LVNEVTEFAK	2	y8		0.958	0.934	0.950	0.905	0.906	0.918	0.875	0.910
218306	QTALVELVK	2	y5	0.958		0.916	0.918	0.941	0.933	0.866	0.883	0.899
28179	VHTEC(CAM)C(CAM)HGDLLC(CAM)ADDR	4	y4	0.944	0.913	0.944	0.905	0.910	0.850	0.896	0.905	0.895
116552	DDNPNLPR	2	y5	0.934	0.916		0.869	0.876	0.927	0.858	0.884	0.884
19645	VPQVSTPTLVEVSR	2	y8	0.971	0.920	0.897	0.960	0.841	0.865	0.855	0.843	0.881
98707	QNC(CAM)ELFEQLGEYK	2	y4	0.942	0.935	0.899	0.920	0.899	0.906	0.895	0.794	0.880
501275	FQNALLVR	2	y6	0.950	0.918	0.869		0.842	0.850	0.865	0.821	0.876
366357	VFDEFKPLVEEPQNLIK	3	y6	0.905	0.941	0.876	0.842		0.859	0.911	0.902	0.873
441926	RPC(CAM)FSALEVDETYVPK	3	b6	0.903	0.950	0.860	0.896	0.932	0.888	0.883	0.927	0.871
131057	TYETTLEK	2	y6	0.921	0.889	0.953	0.927	0.821	0.877	0.805	0.830	0.871
419165	LC(CAM)TVATLR	2	y6	0.900	0.890	0.855	0.868	0.940	0.860	0.946	0.886	0.871
7661	EC(CAM)C(CAM)EKPLLEK	3	y5	0.946	0.943	0.898	0.935	0.838	0.848	0.810	0.823	0.866
53763	LVAASQAALGL	2	b8	0.906	0.933	0.927	0.850	0.859		0.809	0.857	0.862
69408	SLHTLFGDK	2	y7	0.918	0.866	0.858	0.865	0.911	0.809		0.853	0.857
243203	AEFAEVSK	2	y6	0.875	0.883	0.884	0.821	0.902	0.857	0.853		0.847
22522	YIC(CAM)ENQDSISSK	2	y10	0.872	0.823	0.939	0.838	0.797	0.823	0.818	0.829	0.834
21573	AAFTEC(CAM)C(CAM)QAADK	2	y7	0.866	0.821	0.891	0.833	0.807	0.816	0.811	0.900	0.827
15226	HPDYSVLLLLR	3	y4	0.849	0.877	0.842	0.800	0.886	0.928	0.864	0.776	0.818
2995	NEC(CAM)FLQHK	2	y6	0.903	0.821	0.831	0.828	0.780	0.833	0.825	0.669	0.804
206640	YLYEIAR	2	y5	0.830	0.878	0.725	0.817	0.901	0.752	0.789	0.835	0.780
Average coefficient of determination (r ²)				0.910	0.899	0.884	0.876	0.873	0.862	0.857	0.847	

^a Abbreviations: z, charge; (CAM), Carbamidomethylated; (Ox), Oxidized; (P), Phosphorylated

[0328] The final matrix of pairwise correlations between serum albumin peptides (Table 13) was created by excluding 10 additional peptides that contained cysteine residues and/or had an average peak area of <20,000.

Table 13

Peak Area	Peptide sequence ^a	z	Frag Ion	LVNEV	QTALV	VFDEF	DDNPN	FQNAL	LVAAS	AEFAE	TYETT	SLHTL	YLYEI	Ave r ²
739992	LVNEVTEFAK	2	y8		0.958	0.905	0.934	0.950	0.906	0.875	0.921	0.918	0.830	0.911
218306	QTALVELVK	2	y5	0.958		0.941	0.916	0.918	0.933	0.883	0.889	0.866	0.878	0.909
366357	VFDEFKPLVEEPQNLIK	3	y6	0.905	0.941		0.876	0.842	0.859	0.902	0.821	0.911	0.901	0.884
116552	DDNPNLPR	2	y5	0.934	0.916	0.876		0.869	0.927	0.884	0.953	0.858	0.725	0.883
501275	FQNALLVR	2	y6	0.950	0.918	0.842	0.869		0.850	0.821	0.927	0.865	0.817	0.873
53763	LVAASQAALGL	2	b8	0.906	0.933	0.859	0.927	0.850		0.857	0.877	0.809	0.752	0.863
243203	AEFAEVSK	2	y6	0.875	0.883	0.902	0.884	0.821	0.857		0.830	0.853	0.835	0.860
131057	TYETTLEK	2	y6	0.921	0.889	0.821	0.953	0.927	0.877	0.830		0.805	0.711	0.859
69408	SLHTLFGDK	2	y7	0.918	0.866	0.911	0.858	0.865	0.809	0.853	0.805		0.789	0.853
206640	YLYEIAR	2	y5	0.830	0.878	0.901	0.725	0.817	0.752	0.835	0.711	0.789		0.804
Average coefficient of determination (r ²)				0.911	0.909	0.884	0.883	0.873	0.863	0.860	0.859	0.853	0.804	

^a Abbreviations: z, charge; (CAM) and *, Carbamidomethylated; (Ox) and ^{Ox}, Oxidized; (P) and ^P, Phosphorylated

[0329] As undesirable and poorly correlated peptides were progressively excluded, the percentage of correlations with r² > 0.85 increased from 21.4% to 72.2%. Additional metrics

showing increased correlations throughout the peptide selection process are presented in Table 14.

Table 14.

	All Peptides	Remove Miscleaved and Met ^{ox}	Remove $r^2 < 0.78$	Remove Cys and Peak <20,000
Peptides	63	26	20	10
Total correlations	1953	325	190	45
$r^2 > 0.90$ (%)	9.2	18.5	31.6	37.8
$r^2 > 0.85$ (%)	21.4	35.1	58.9	72.2
Highest mean r^2	0.799	0.848	0.910	0.911
Lowest mean r^2	0.218	0.355	0.780	0.804

[0330] Validation of the serum albumin signature peptides. The resulting collection of 10 serum albumin signature peptides was compared with results from previously validated SRM assays.

[0331] Two of the peptides, LVNEVTEFAK and DDNPNLPR, were targeted in the SRM assays on 42 urine samples described in Example 1. Three transitions were monitored for each peptide. The assay included SIL peptide internal standards corresponding to the two serum albumin peptides. The correlation between normalized peak areas of the two serum albumin peptides was >98%, regardless of which transitions were compared.

[0332] Beasley-Green and colleagues selected 11 serum albumin peptides on the basis of retention time reproducibility, peak intensity, and the degree of sequence coverage. They built an SRM assay with SIL internal standards that targeted two transitions for each peptide. The linearity, precision, repeatability and accuracy of this SRM assay were extensively validated.

[0333] Eight out of the 10 highly correlated signature peptides shown in Table 13 were also targeted in Beasley-Green's SRM assay (Table 15). Two of the three Beasley-Green peptides that are not also found in Table 13 contain a cysteine amino acid residue. The third was not among the 63 quantifiable peptides in the PDAY SWATH data. The two Table 13 peptides that were also not targeted by Beasley-Green had the second and third lowest peak areas among Table 13 peptides.

Table 15

Highly-correlated signature peptides	Beasley-Green signature peptides
LVNEVTEFAK	LVNEVTEFAK
QTALVELVK	QTALVELVK
VFDEFKPLVEEPQNLIK	VFDEFKPLVEEPQNLIK
DDNPNLPR	
FQNALLVR	FQNALLVR
LVAASQAALGL	LVAASQAALGL
AEFAEVSK	AEFAEVSK
TYETTLEK	TYETTLEK
SLHTLFGDK	
YLYEIAR	YLYEIAR
	DLGEENFK
	LCTVATLR
	RPCFSALEVDETYVPK

[0334] The broad applicability of the signature peptide selection method of the present invention is highlighted by the observation that serum albumin signature peptides selected from the SWATH data yield reliable results in SRM assays. This was true despite differences in sample origin (aortic tissue v urine), sample preparation (harsh extraction and denaturation with urea v gentle treatment with RapiGest), and MS instruments (Triple-TOF v triple quadrupole).

Signature peptide selection from blood and tissue proteins

[0335] The SWATH dataset for the PDAY extracts includes data on 1,121 proteins. Six blood proteins and two tissue proteins were selected as exemplary proteins for the identification of highly-correlated signature peptides (Table 16). Several of these proteins have been implicated as biomarkers.

Table 16

Principle Location	Protein	UniProt ID	Quantifiable Peptides	Correlated Signature Peptides
Blood	Hemoglobin delta	P02042	14	4
	Hemopexin	P02790	14	7
	Apolipoprotein A-I	P02647	22	7
	Alpha-1-antitrypsin	P01009	35	12
	Serotransferrin	P02787	45	10
	Complement C3	P01024	68	36
Tissue	Mimecan	P20774	16	4
	Filamin-A	P21333	104	51

[0336] For each protein, data from 6 transitions for all quantifiable peptides was imported into a Microsoft Excel spreadsheet. The average peak area was calculated for each transition and the transition with the strongest peak area was selected to represent the peptide. All pairwise correlations (r^2) between the peptides were calculated with Prism and transferred to the Excel spreadsheet to create a correlation matrix. Peptides within the matrix were sorted according to the average of their correlations. Peptides having an average of correlations of less than 0.5 were removed. The peptides were resorted according their average of correlations and peptides having an average of correlations of less than 0.6 were removed. The process was repeated a third time to exclude peptides having an average of correlations of less than 0.7. Peptides with missed cleavages or methionine residues were then removed, and the remaining peptides were again sorted according to their average of correlations. A summary of these results is presented in Table 17.

Table 17

Protein / Sequence	Charge (z)	Fragment Ion	Average r^2	Average Peak Area
Hemoglobin subunit delta				
VNVDAVGGEALGR	3	y4	0.911	2043
LLGNVLVC(Cam)VLAR	2	y7	0.911	3704
GTFSQLSELHC(Cam)DK	2	y3	0.855	1408
VNVDAVGGEALGR	2	y7	0.878	18575
Hemopexin				
LLQDEFPGIPSPLDAAVEC(CAM)HR	3	y12	0.880	6013
SGAQATWTLPWPHEK	3	y4	0.849	6078
QGHNSVFLIK	2	y8	0.806	323
EVGTPHGIILDSVDAAFIC(CAM)PGSSR	3	y5	0.823	10191
NFPSPVDAAFR	2	y9	0.864	12780
GGYTLVSGYPK	2	y6	0.836	2963
GEC(CAM)QAEGVLFFQGDR	2	y7	0.796	1419
Apolipoprotein A-I				
VSFLSALEEYTK	2	y8	0.927	13501
THLAPYSDELRL	2	b3	0.920	3093
QGLLPVLESFK	2	y7	0.918	25313
DYVSQFEFSALGK	2	y10	0.896	6389
EQLGPVTQEFWDNLEK	2	y4	0.891	2299
LLDNWDSVTSTFSK	2	y6	0.888	6381
DLATVYVDVLK	2	y6	0.848	4235

Protein / Sequence	Charge (z)	Fragment Ion	Average r^2	Average Peak Area
Alpha-1-antitrypsin				
FLENEDR	2	y5	0.902	1033
VFSNGADLSGVTEEAPLK		y3	0.890	10629
LSITGTYDLK	2	y7	0.889	15263
AVLTIDEK	2	y6	0.875	28127
LQHLENELTHDIITK	4	b3	0.874	3774
VFSNGADLSGVTEEAPLK		y3	0.871	1443
SASLHLPK	2	y6	0.867	2358
SVLGQLGITK	2	y8	0.847	28919
TDTSHHDQDHPTFNK	4	y5	0.832	575
DTEEDFDHVDQVTTVK	3	y7	0.824	1724
LYHSEAFVTNFGDTEEAK		y7	0.744	1267
LQHLENELTHDIITK	3	b3	0.742	4044
Serotransferrin				
DGAGDVAFAVK	2	y7	0.815	15560
ASYLDC(Cam)IR	2	y4	0.796	7916
SVIPSDGPSVAC(Cam)VK	2	y11	0.791	11944
IEC(Cam)VSAETTEDC(Cam)IAK	2	b3	0.772	5663
DSAHGFLK	2	y4	0.769	898
SASDLTWDNLK	2	y6	0.753	7257
DDTVC(Cam)LAK	2	y6	0.753	3883
FDEFFSEGC(Cam)APGSK	2	y4	0.736	24477
EFQLFSSPHGK	3	y6	0.728	4093
C(Cam)DEWSVNSVGK	2	b3	0.710	926
Complement C3				
FYYIYNEK	2	y6	0.994	646
DTWVEHWPEEDEC(Cam)QDEENQK	3	y6	0.948	1379
EPGQDLVVLPLSITTFIPSR	2	y4	0.938	1234
SSLSVPYVIVPLK	2	y8	0.931	4660
NTLIYLDK	2	y6	0.930	2245
QLYNVEATSYALLALLQLK	3	y6	0.929	311
IHWESASLLR	3	y4	0.925	2419
DIC(Cam)EEQVNSLPGSITK	2	y6	0.924	6151
FISLGEAC(Cam)K	2	y7	0.924	3834
VFLDC(Cam)C(Cam)NYITELR	2	y4	0.924	2124
QGALELIK	2	y4	0.923	2305
DSC(Cam)VGSLVVK	2	y6	0.918	3994
GLEVTITAR	2	y5	0.918	2041
EYVLPSFEVIVPEK	2	b3	0.917	4462

Protein / Sequence	Charge (z)	Fragment Ion	Average r^2	Average Peak Area
EVVADSVWVDVK	2	y5	0.913	1435
VSHSEDDC(Cam)LAFK	3	y3	0.913	1815
SGSDEVQVGQQR	2	y4	0.913	587
LVAYYTLIGASGQR	3	y6	0.912	1596
TIYTPGSTVLYR	2	y8	0.911	3207
GYTQQLAFR	2	y5	0.910	1039
DAPDHQELNLDVSLQLPSR	3	y7	0.909	2002
VELLHNPAFC(Cam)SLATTK	3	b6	0.906	1436
AC(Cam)EPGVDYVYK	2	y8	0.899	2000
IPIEDGSGEVVLSR	2	y11	0.890	2511
SNLDEIIAEENIVSR	2	y9	0.886	3618
VYAYYNLEESC(Cam)TR	2	y6	0.885	421
VTIKPAPETEK	3	y7	0.876	1163
DFDFVPPVVR	2	y5	0.876	18784
TGLQEVEVK	2	y6	0.861	2191
APSTWLTAYVVK	2	y7	0.859	365
VHQYFNVELIQPGAVK	3	y5	0.835	4037
VPVAVQGEDTVQSLTQGDGVAK	2	b4	0.833	5289
SGIPIVTSPYQIHFTK	3	y4	0.831	918
QPSSAFAAFVK	2	y6	0.829	1977
ADIGC(Cam)TPGSGK	2	y5	0.801	858
AAVYHHFISDGVR	3	y5	0.748	650
Mimecan				
LNNLTFLYLDHNALESVPLNPESLR	3	y5	0.842	269781
LDFTGNLIEDIEDGTFSK	2	y5	0.839	242667
LSLLEELSLAENQLLK	3	y7	0.824	280665
DFADIPNLR	2	y4	0.757	59975
Filamin-A				
EGPYSISVLYGDEEVPR	2	y11	0.871	15559
EATTEFSVDAR	2	y6	0.869	19572
FNEEHIPDSPFVVPVASPSGDAR	3	y10	0.861	72916
AFGPGLQGGSAGSPAR	2	y9	0.858	17271
VSGQGLHEGHTFEPAEFIIDTR	3	y9	0.849	2064
VANPSGNLTETYVQDR	2	y8	0.849	11286
SPFSVAVSPSLDLSK	2	y7	0.845	15559
FNGTHIPGSPFK	3	y6	0.847	5564
VGEPGHGGDPGLVSAYGAGLEGGVTGNP AEFVVNTSNAGAGALSVTIDGPSK	4	y4	0.837	6776
VGSAADIPINISETDLSLLTATVVPPSGR	3	y5	0.839	76329

Protein / Sequence	Charge (z)	Fragment Ion	Average r^2	Average Peak Area
ENGVYLIDVK	2	y6	0.829	10363
DGSC(CAM)SVEYIPYEAGTYSLNVTYGGH QVPGSPFK	3	y6	0.831	13569
YNEQHVPGPSPTAR	2	y8	0.826	2831
VKETADFK	2	y6	0.819	1289
YGGQPVPNFPSK	2	y6	0.823	14928
DAGEGLLAVQITDPEGKPK	2	y6	0.817	12196
NGHVGISFVPK	2	y7	0.818	676
GTVEPQLEAR	2	y6	0.818	41503
ASGPGLNTTGVPAASLPVEFTIDAK	2	y13	0.816	6897
IANLQTDLSGLR	2	y8	0.817	32516
GLVEPVDVVDNADGTQTVNYVPSR	3	y3	0.811	39841
EAGAGGLAIAVEGPSK	2	y4	0.812	19210
TGVAVNKPAEFTVDAK	2	y9	0.810	2720
DGSC(CAM)GVAYVVQEPGDYEVSVK	2	y9	0.809	2266
EEGPYEVEVITYDGVVPGPSFPLEAVAPTK PSK	3	y6	0.808	13995
FGGEHVPNSPFQVTALAGDQPSVQPPLR	3	y4	0.798	30580
VEPGLGADNSVVR	2	y11	0.798	38786
LYSVSYLLK	2	y7	0.794	17335
SPFEVYVDK	2	y7	0.799	16092
SADFVVEAIGDDVGTGLGFSVEGPSQAK	3	y6	0.789	15162
AGVAPLQVK	2	y5	0.791	38022
AEISC(CAM)TDNQDGTC(CAM)SVSYLPV LPGDYSILVK	3	y9	0.774	15308
DAGEGGLSLAIEGPSK	2	y4	0.771	21033
AHVVPC(CAM)FDASK	2	y7	0.775	7011
LPQLPITNFSR	2	y7	0.771	95463
AWGPGLEGGVVGK	2	y11	0.760	15328
YTPVQQGPVGNNVTYGGDPIPK	2	y4	0.764	7271
FADQHVPGPSFVSVK	3	y8	0.758	4813
DQEFTVK	2	y5	0.753	1425
AEISFEDR	2	y5	0.755	18604
VNQPASFAVSLNGAK	2	y12	0.752	3160
TFSVWYVPEVTGTHK	2	y8	0.747	3825
C(CAM)APGVVGPAAEDIDFDIIR	2	y5	0.740	4577
LDVQFSGLTK	2	y6	0.728	9797
NGQHVASSPIPVVISQSEIGDASR	3	y10	0.739	1156
VTAQGPGLPSGNIANK	2	y8	0.729	11129

Protein / Sequence	Charge (z)	Fragment Ion	Average r^2	Average Peak Area
DAGYGGLSLSIEGPSK	2	y4	0.722	2508
WGDEHIPGSPYR	2	y6	0.710	3121
DVDIIDHHDNTYTVK	3	y13	0.710	6653
GAGTGGLGLAVEGPSEAK	2	y6	0.707	17618
THIQDNHDGTYTVAYVPDVTGR	3	y6	0.712	1930

[0337] Persons of ordinary skill will recognize that this process can be repeated to identify correlated signature peptides for all 1,121 identified proteins in the PDAY SWATH data. The resulting correlated signature peptides will provide accurate and reproducible quantitative results for this and other MS datasets. Persons of ordinary skill will also realize that this approach will allow signature peptides to be selected from any database for every human (or other species) protein. Reproducibility can be enhanced by incorporating SIL peptides matching the sequence of the correlated signature peptides. Correlated signature peptides identified in SWATH data can also be targeted in higher sensitivity SRM assays.

References

- [0338] Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*. 2008;24:2534–2536.
- [0339] Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*. 2007; 4:207–214.
- [0340] Craig R, Beavis R. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004; 20:1466–1467.
- [0341] Eng JK, Jahan TA, Hoopmann MR. Comet: an open source tandem mass spectrometry sequence database search tool. *Proteomics*. 2012 Nov 12. doi: 10.1002/pmic.201200439
- [0342] Agger SA1, Marney LC, Hoofnagle AN. Simultaneous quantification of apolipoprotein A-I and apolipoprotein B by liquid-chromatography-multiple-reaction-monitoring mass spectrometry. *Clin Chem*. 2010 Dec;56(12):1804-13.
- [0343] Keller, A., Eng, J., Zhang, N., Li, X. J., Aebersold, R., A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol*. 2005, 1, 2005 0017.
- [0344] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal.*

Chem. 2002, 74, 5383–5392.

[0345] Shteynberg D., Deutsch E.W., Lam H., Eng J.K., Sun Z., Tasman N., Mendoza L., Moritz R.L., Aebersold R., Nesvizhskii A.I. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates.

Mol Cell Proteomics. 2011, 10:M111.007690

[0346] Collins BC, Gillet LC, Rosenberger G, Rost HL, Vichalkovski A, Gstaiger M, Aebersold R. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. Nat Methods. 2013; 10:1246–1253.

[0347] Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, Aebersold R. Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics. 2007;7:655–667.

[0348] scher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, Rinner O. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics. 2012;12:1111–1121.

[0349] Rost HL, Rosenberger G, Navarro P, Gillet L, Miladinovic SM, Schubert OT, Wolski W, Collins BC, Malmstrom J, Malmstrom L, Aebersold R. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotech. 2014;32:219–223.

[0350] Weisser H, Nahnsen S, Grossmann J, Nilse L, Quandt A, Brauer H, Sturm M, Kenar E, Kohlbacher O, Aebersold R, Malmstrom L. An automated pipeline for high-throughput label-free quantitative proteomics. J Proteome Res. 2013;12:1628–1644.

[0351] Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, Kuster B, Aebersold R. Computational prediction of proteotypic peptides for quantitative proteomics. Nat Biotechnol. 2007;25:125–131

[0352] Beasley-Green A, Burris NM, Bunk DM, Phinney KW, Multiplexed LC-MS/MS assay for urine albumin, J Proteome Res. 2014 Sep 5;13(9):3930-9.

[0353] SEQ ID NOs for all the peptide sequences described herein are listed in Table 18 below.

Table 18

SEQ ID NO:	Sequence	SEQ ID NO:	Sequence
1	ACAHW	135	DLGEENFK
2	AFSSL	136	VNVDVGGGEALGR

SEQ ID NO:	Sequence	SEQ ID NO:	Sequence
3	DGPCG	137	LLGNVLVCVLAR
4	DSTIQ	138	GTFSQLSELHCDK
5	DWVSV	139	LLQDEFPGIPSPLDAAVECHR
6	FAGNY	140	SGAQATWTELPWPHEK
7	FALLM	141	QGHNSVFLIK
8	FSVQM	142	EVGTPHGIILDSVDAAFICPGSSR
9	FVGQG	143	NFPSPVDAAFR
10	GDGWH	144	GGYTLVSGYPK
11	GVQAT	145	GECQAEGLVFFQGDR
12	INFAC	146	VSFLSALEEYTK
13	KACAH	147	THLAPYSDELRL
14	KGVQA	148	QGLLPVLESFK
15	LECGA	149	DYVSQFEGSALGK
16	MAETC	150	EQLGPVTQEFWDNLEK
17	NETHA	151	LLDNWDSVTSTFSK
18	QDFNI	152	DLATVYVDVLK
19	SGSVI	153	FLENEDR
20	SLGFD	154	VFSNGADLSGVTEEAPLK
21	STEYG	155	LSITGTYDLK
22	TALQP	156	AVLTIDEK
23	TLDEY	157	LQHLENELTHDIITK
24	VFMYL	158	SASLHLPK
25	VGGTG	159	SVLGQLGITK
26	VLNLG	160	TDTSHHDQDHPTFNK
27	VWLPL	161	DTEEDFHVDQVTTVK
28	YFIIQ	162	LYHSEFTVNFGDTEEAK
29	WHCQC	163	DGAGDVAFAVK
30	CSGFN	164	ASYLDCIR
31	CKPTC	165	SVIPSDGPSVACVK
32	RTLDEYWRS	166	IECVSAETTEDCIAK
33	RSTEYGEYACDTDLRG	167	DSAHGFLK
34	RFVGQGGARM	168	SASDLTWDNLK
35	RMAETCVPVLRC	169	DDTVCLAK
36	KACAHWSGHCCLDASVQVKA	170	FDEFFSEGCAPGSK
37	RWHCQCKQ	171	EFQLFSSPHGK
38	KQDFNITDISLLEHRL	172	CDEWSVNSVGK
39	RLECGANDMKV	173	FYYIYNEK
40	KSLGFDKV	174	DTWVEHWPEEDECQDEENQK
41	KVFMYLSDSRC	175	EPGQDLVVLPLSITDFIPSFR
42	RCSGFNDRD	176	SSLSVPYVIVPLK

SEQ ID NO:	Sequence	SEQ ID NO:	Sequence
43	RDWVSVVTPARD	177	NTLIYLDK
44	RDGPCGTVLTRN	178	QLYNVEATSYALLALLQLK
45	RNETHATYSNTLYLADEIIRD	179	IHWESASLLR
46	KINFACSYPLDMKV	180	DICEEQVNSLPGSITK
47	KTALQPMVSALNIRV	181	FISLGEACK
48	RVGGTGMMFTVRM	182	VFLDCCNYITELR
49	RFALLMTNCYATPSSNATDPLKY	183	QGALELIK
50	KYFIQDRC	184	DSCVGSLLVVK
51	RDSTIQVVENGESSQGRF	185	GLEVTITAR
52	RFSVQMFRF	186	EYVLPSEFVIVEPEK
53	KCKPTCSGTRF	187	EVVADSVWVDVK
54	RSGSVIDQSRV	188	VSHSEDDCLAFK
55	RVLNLGPITRK	189	SGSDEVQVGQQR
56	KGVQATVSRA	190	LVAYYTLIGASGQR
57	RAFSSLGLLKV	191	TIYTPGSTVLYR
58	KVWLPLLLSATLTFTQ	192	GYTQQLAFR
59	TLDEYWR	193	DAPDHQELNLDVSLQLPSR
60	DGPCGTVLTR	194	VELLHNPAFCSLATTK
61	YFIQDR	195	ACEPGVDYVYK
62	FVGQGGAR	196	IPIEDGSGEVLSR
63	DWVSVVTPAR	197	SNLDEIIAEENIVSR
64	SGSVIDQSR	198	VYAYYNLEESCTR
65	FSVQMFR	199	VTIKPAPETEK
66	STEYGEYACDTRLR	200	DFDFVPPVVR
67	VFMYLSDSR	201	TGLQEVEVK
68	MAETCVPVLR	202	APSTWLTAYVVK
69	DSTIQVVENGESSQGR	203	VHQYFNVELIQPGAVK
70	TALQPMVSALNIR	204	VPVAVQGEDTVQSLTQGDGVAK
71	YSQQQLMETSHR	205	SGIPIVTSPIYQIHFTK
72	RDWENPGVTQLNR	206	QPSSAFAAFVK
73	GDFQFNISR	207	ADIGCTPGSGK
74	IDPNAWVER	208	AAVYHHFISDGVR
75	DVSLHLKPTTQISDFHVATR	209	LNNLTFLYLDHNALESVPLNLPESLR
76	VDEDQFPAPVK	210	LDFTGNLIEDIEDGTFSK
77	DWENPGVTQLNR	211	LSLLEELSLAENQLLK
78	APLDNDIGVSEATR	212	DFADIPNLR
79	WVGYGQDSR	213	EGPYSISVLYGDEEVPR
80	GDFQFNIS	214	EATTEFSVDAR
83	QTALVELVK	215	FNEEHIPDSPFVVPVASPSGDAR
84	SHCIAEVENDEMPADLPPLAADFVESK	216	AFGPGLQGGSGAGSPAR

SEQ ID NO:	Sequence	SEQ ID NO:	Sequence
85	LVNEVTEFAK	217	VSGQGLHEGHTFEPAEFIIDTR
86	RPCFSALEVDETYVPK	218	VANPSGNLTETYVQDR
87	VFDEFKPLVEEPQNLIK	219	SPFSVAVSPSLDLSK
88	QNCELFEQLGEYK	220	FNGTHIPGSPFK
89	AVMDDFAAFVEK	221	VGEPGHGGDPGLVSAYGAGLEGGVTGNPAEFVV NTSNAGAGALSVTIDGPSK
90	RMPCAEDYLSVVLNQLCVLHEK	222	VGSAADIPINISETDLSELLTATVPPSGR
91	KQTALVELVK	223	ENGVYLIDVK
92	VHTECCHGDLLECADDR	224	DGSCSVEYIPYEAGTYSLNVTYGGHQVPGSPFK
93	LCTVATLR	225	YNEQHVPGPSPTAR
94	LVRPEVDVMCTAFHDNEETFLKK	226	VKETADFK
95	ECCEKPLLEK	227	YGGQPVPNFPSK
96	LVAASQAALGL	228	DAGEGLLAVQITDPEGKPK
97	DDNPNLPR	229	NGHVGISFVPK
98	FQNALLVR	230	GTVEPQLEAR
99	VPQVSTPTLVEVSR	231	ASGPGLNTTGVPASLPVEFTIDAK
100	RHPYFYAPELLFFAK	232	IANLQTDLSDDL
101	AEFAEVSK	233	GLVEPVDVVDNADGTQTVNYVPSR
102	SLHTLFGDK	234	EAGAGGLAIAVEGPSK
103	KVPQVSTPTLVEVSR	235	TGVAVNKPAEFTVDAK
104	KYLYEIAR	236	DGSCGVAYVVQEPGDYEVSVK
105	TYETTLEK	237	EEGPYEVEVTDGVPVPGSPFPLEAVPTKPSK
106	RHPDYSVVLRL	238	FGGEHVPNSPFQVTALAGDQPSVQPPLR
107	HPDYSVVLRL	239	VEPGLGADNSVVR
108	KLVAASQAALGL	240	LYSVSYLLK
109	LVRPEVDVMCTAFHDNEETFLK	241	SPFEVYVDK
110	FKDLGEENFK	242	SADVFVEAIGDDVGTGLGFSVEGPSQAK
111	YLYEIAR	243	AGVAPLQVK
112	VHTECCHGDLLECADDRADLAK	244	AEISCTDNQDGTCSVSYLPVLPDYSILVK
113	NECFLQHKDDNPNLPR	245	DAGEGGLSLAIEGPSK
114	AAATECCQAADK	246	AHVVPDADASK
115	ETYGEMADCCAK	247	LPQLPITNFSR
116	QEPERNECFLOHKDDNPNLPR	248	AWGPGLEGGVVGK
117	YICENQDSISSK	249	YTPVQQGPVGVNVTYGGDPIPK
118	ADDKETCFAEEGK	250	FADQHVPGSPFSVK
119	NECFLOHK	251	DQEFTVK
120	CCTESLVNR	252	AEISFEDR
121	LAKTYETTLEKCCAAADPHECYAK	253	VNQPASFAVSLNGAK
122	MPCAEDYLSVVLNQLCVLHEK	254	TFSVWYVPEVTGTHK
123	HPYFYAPELLFFAK	255	CAPGVVGPAEADIDFDIIR

SEQ ID NO:	Sequence	SEQ ID NO:	Sequence
124	SLHTLFGDKLCTVATLR	256	LDVQFSGLTK
125	LKECCEKPLLEK	257	NGQHVASSPIPVVISQSEIGDASR
126	ALVLIAFAQYLQQCPFEDHVK	258	VTAGPGGLEPSGNIANK
127	ADDKETCFAEEGKK	259	DAGYGGLSLSIEGPSK
128	EFNAETTFHADICTLSEK	260	WGDEHIPGSPYR
129	DVFLGMFLYEYAR	261	DVDIIDHHDNTYTVK
130	LDEL RDEGK	262	GAGTGGLGLAVEGPSEAK
131	CCAAADPHECYAK	263	THIQDNHDGTYTVAYVPDVTGR
132	NYAEAKDVFLGMFLYEYAR		
133	TCVADESAENCDK		
134	EFNAETTFHADICTLSEKER		

[0354] The various methods and techniques described above provide a number of ways to carry out the application. Of course, it is to be understood that not necessarily all objectives or advantages described can be achieved in accordance with any particular embodiment described herein. Thus, for example, those skilled in the art will recognize that the methods can be performed in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other objectives or advantages as taught or suggested herein. A variety of alternatives are mentioned herein. It is to be understood that some preferred embodiments specifically include one, another, or several features, while others specifically exclude one, another, or several features, while still others mitigate a particular feature by inclusion of one, another, or several advantageous features.

[0355] Furthermore, the skilled artisan will recognize the applicability of various features from different embodiments. Similarly, the various elements, features and steps discussed above, as well as other known equivalents for each such element, feature or step, can be employed in various combinations by one of ordinary skill in this art to perform methods in accordance with the principles described herein. Among the various elements, features, and steps some will be specifically included and others specifically excluded in diverse embodiments.

[0356] Although the application has been disclosed in the context of certain embodiments and examples, it will be understood by those skilled in the art that the embodiments of the application extend beyond the specifically disclosed embodiments to other alternative embodiments and/or uses and modifications and equivalents thereof.

[0357] Preferred embodiments of this application are described herein, including the best mode known to the inventors for carrying out the application. Variations on those preferred embodiments will become apparent to those of ordinary skill in the art upon reading the foregoing description. It is contemplated that skilled artisans can employ such variations as appropriate, and the application can be practiced otherwise than specifically described herein. Accordingly, many embodiments of this application include all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the application unless otherwise indicated herein or otherwise clearly contradicted by context.

[0358] All patents, patent applications, publications of patent applications, and other material, such as articles, books, specifications, publications, documents, things, and/or the like, referenced herein are hereby incorporated herein by this reference in their entirety for all purposes, excepting any prosecution file history associated with same, any of same that is inconsistent with or in conflict with the present document, or any of same that may have a limiting affect as to the broadest scope of the claims now or later associated with the present document. By way of example, should there be any inconsistency or conflict between the description, definition, and/or the use of a term associated with any of the incorporated material and that associated with the present document, the description, definition, and/or the use of the term in the present document shall prevail.

[0359] It is to be understood that the embodiments of the application disclosed herein are illustrative of the principles of the embodiments of the application. Other modifications that can be employed can be within the scope of the application. Thus, by way of example, but not of limitation, alternative configurations of the embodiments of the application can be utilized in accordance with the teachings herein. Accordingly, embodiments of the present application are not limited to that precisely as shown and described.

[0360] Various embodiments of the invention are described above in the Detailed Description. While these descriptions directly describe the above embodiments, it is understood that those skilled in the art may conceive modifications and/or variations to the specific embodiments shown and described herein. Any such modifications or variations that fall within the purview of this description are intended to be included therein as well. Unless specifically noted, it is the intention of the inventors that the words and phrases in the

specification and claims be given the ordinary and accustomed meanings to those of ordinary skill in the applicable art(s).

[0361] The foregoing description of various embodiments of the invention known to the applicant at this time of filing the application has been presented and is intended for the purposes of illustration and description. The present description is not intended to be exhaustive nor limit the invention to the precise form disclosed and many modifications and variations are possible in the light of the above teachings. The embodiments described serve to explain the principles of the invention and its practical application and to enable others skilled in the art to utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. Therefore, it is intended that the invention not be limited to the particular embodiments disclosed for carrying out the invention.

[0362] While particular embodiments of the present invention have been shown and described, it will be obvious to those skilled in the art that, based upon the teachings herein, changes and modifications may be made without departing from this invention and its broader aspects and, therefore, the appended claims are to encompass within their scope all such changes and modifications as are within the true spirit and scope of this invention.

CLAIMS

1. A method of identifying signature peptides for quantifying a polypeptide in a sample, comprising:
 - acquiring mass spectrometry (MS) data on multiple candidate peptides derived from the polypeptide in multiple samples;
 - using the MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and
 - identifying highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide.
2. The method of claim 1, wherein the MS data is collected by a targeted acquisition method.
3. The method of claim 1, wherein the MS data is collected by a data independent acquisition method.
4. The method of claim 1, wherein the correlation values are coefficient of determination (r^2) values.
5. The method of claim 1, wherein the multiple candidate peptides are derived by proteolysis or chemical cleavage of the polypeptide.
6. The method of claim 1, wherein acquiring MS data comprises operating a mass spectrometer.
7. The method of claim 1, wherein the sample is derived from food, water, cheek swab, blood, serum, plasma, urine, saliva, semen, cells, tissue, tumor, or a combination thereof.
8. The method of claim 1, further comprising ranking the correlation values of the multiple candidate peptides.
9. The method of claim 1, wherein the highly correlated peptides have correlation values ranked in the top 2, 3, 4, 5, 6, 7, 8, 9, or 10 among the multiple candidate peptides.
10. The method of claim 1, wherein the highly correlated peptides have correlation values ranked in the top 80%, 70%, 60%, 50%, 40%, 30% or 20% among the multiple candidate peptides.
11. The method of claim 1, further comprising ranking the mean or median correlation values of the multiple candidate peptides.

12. The method of claim 1, wherein the highly correlated peptides have mean or median correlation values ranked in the top 2, 3, 4, 5, 6, 7, 8, 9, or 10 among the multiple candidate peptides.
13. The method of claim 1, wherein the highly correlated peptides have mean or median correlation values ranked in the top 80%, 70%, 60%, 50%, 40%, 30% or 20% among the multiple candidate peptides.
14. The method of claim 1, wherein the multiple candidate peptides are obtained from data-dependent MS screen, data-independent MS data, targeted peptides data, MS spectral database, or proteotypic peptide prediction, or a combination thereof.
15. The method of claim 1, further comprising eliminating peptides that satisfy one or more of the following criteria:
 - i. not previously detected by MS;
 - ii. not unique to the polypeptide;
 - iii. absent from the polypeptide's mature form;
 - iv. containing an uncleaved protease recognition site;
 - v. susceptible to post-translational modification (PTM);
 - vi. containing methionine and/or cysteine residues;
 - vii. sensitive to endogenous proteases;
 - viii. having m/z values lower than an m/z bottom cutoff value;
 - ix. having m/z values higher than an m/z top cutoff value; and
 - x. having signal intensities lower than an intensity bottom cutoff value in the acquired MS data.
16. The method of claim 1, wherein the identified signature peptides have high and reproducible signal intensities in the acquired MS data.
17. A method of quantifying a polypeptide in a sample, comprising:
 - cleaving the polypeptide to yield a signature peptide identified according to the method of claim 1;
 - analyzing the sample on a mass spectrometer;
 - detecting MS signals of the signature peptide; and
 - quantifying the polypeptide based on the detected MS signals.
18. The method of claim 17, wherein multiple polypeptides in a complex sample are quantified.

19. The method of claim 17, further comprising spiking the sample with an internal standard of the signature peptide and detecting the internal standard's MS signals.
20. The method of claim 19, wherein the internal standard comprises the signature peptide labeled with a stable isotope.
21. The method of claim 19, further comprising normalizing the signature peptide's MS signals to the internal standard's MS signals.
22. A method of quantifying a polypeptide in a sample, comprising:
 - cleaving the polypeptide in the sample to yield a signature peptide identified according to the method of claim 1;
 - spiking the sample with an internal standard of the signature peptide;
 - capturing the signature peptide and internal standard with a capturing reagent specifically binding to the signature peptide;
 - analyzing the captured signature peptide and internal standard on a mass spectrometer;
 - detecting MS signals of the signature peptide the internal standard; and
 - quantifying the signature peptide based on the detected MS signals.
23. The method of claim 22, wherein the capturing reagent is an antibody or an antigen-binding fragment thereof specifically binding to the signature peptide.
24. The method of claim 22, wherein the capturing reagent is an aptamer specifically binding to the signature peptide.
25. A kit for quantifying a polypeptide in a sample, comprising:
 - an internal standard of a signature peptide identified for the polypeptide according to the method of claim 1; and
 - instructions for using the internal standard to quantify the polypeptide in the sample.
26. The kit of claim 25, wherein the kit comprises multiple internal standards.
27. The kit of claim 25, wherein the kit quantifies multiple polypeptides in a complex sample.
28. The kit of claim 25, further comprising an antibody specifically binding to the signature peptide.
29. The kit of claim 25, further comprising a protease for cleaving the polypeptide to yield the signature peptide.

30. The kit of claim 25, wherein the internal standard comprises the signature peptide labeled with a stable isotope.
31. A method of identifying signature fragments for quantifying a macromolecule in a sample, comprising:
- acquiring mass spectrometry (MS) data on multiple candidate fragments of the macromolecule from multiple samples;
 - using the MS data to calculate correlation values for pairwise comparisons between each of the multiple candidate fragments; and
 - identifying the highly correlated fragments among the multiple candidate fragments as the signature fragments for quantifying the macromolecule.
32. A system for identifying signature peptides for quantifying a polypeptide, comprising:
- a mass spectrometer configured for acquiring mass spectrometry (MS) data on multiple candidate peptides derived from the polypeptide in multiple samples; and
 - a computer configured for using the MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide,
- wherein the mass spectrometer and the computer are connected via a communication link.
33. The system of claim 32, wherein the computer comprises:
- a memory configured for storing a program; and
 - a processor configured for executing the program,
- wherein the program comprises instructions for using the MS data to calculate correlation values for pairwise comparisons among the multiple candidate peptides; and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide.
34. A non-transitory computer-readable storage medium,
- wherein the non-transitory computer-readable storage medium is configured for storing a program,
 - wherein the program is configured for execution by a processor of a computer,
- and

wherein the program comprises instructions for using mass spectrometry (MS) data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide.

35. A computer, comprising:

- a memory configured for storing a program; and
- a processor configured for executing the program,

wherein the program comprises instructions for using mass spectrometry (MS) data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide.

36. A computer implemented method, comprising:

- providing the computer of claim 35;
- inputting mass spectrometry (MS) data into the computer; and

- operating the computer to use the MS data to calculate correlation values for pairwise comparisons between each of multiple candidate peptides for quantifying a polypeptide, and for identifying the highly correlated peptides among the multiple candidate peptides as the signature peptides for quantifying the polypeptide.

37. A non-transitory computer-readable storage medium,

- wherein the non-transitory computer-readable storage medium is configured for storing a program,

- wherein the program is configured for execution by a processor of a computer,
- and

- wherein the program comprises instructions for processing MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and for quantify the polypeptide based on the signature peptide.

38. A computer, comprising:

- a memory configured for storing a program; and
- a processor configured for executing the program,

wherein the program comprises instructions for processing MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and for quantifying the polypeptide based on the signature peptide.

39. A computer implemented method, comprising:
 - providing the computer of claim 38;
 - inputting MS data into the computer; and
 - operating the computer to process MS data to identify, analyze and/or quantify a signature peptide of a polypeptide and to quantify the polypeptide based on the signature peptide.
40. A method of producing an antibody, comprising
 - providing a signature peptide identified according to the method of claim 1;
 - immunizing an animal using the signature peptide, thereby producing the antibody.
41. An antibody specifically binding to a signature peptide identified according to the method of claim 1, or an antigen-binding fragment thereof.
42. A method of quantifying a polypeptide in a sample, comprising:
 - contacting the sample with the antibody of claim 41 or an antigen-binding fragment thereof;
 - detecting the binding between the polypeptide and the antibody or the antigen-binding fragment thereof; and
 - quantifying the polypeptide based on the detected binding.
43. A method of producing a capturing reagent, comprising:
 - providing a signature peptide identified according the method of claim 1; and
 - producing the capturing reagent specifically binding to the signature peptide.
44. A capturing reagent specifically binding to a signature peptide identified according to the method of claim 1.

FIG. 1A

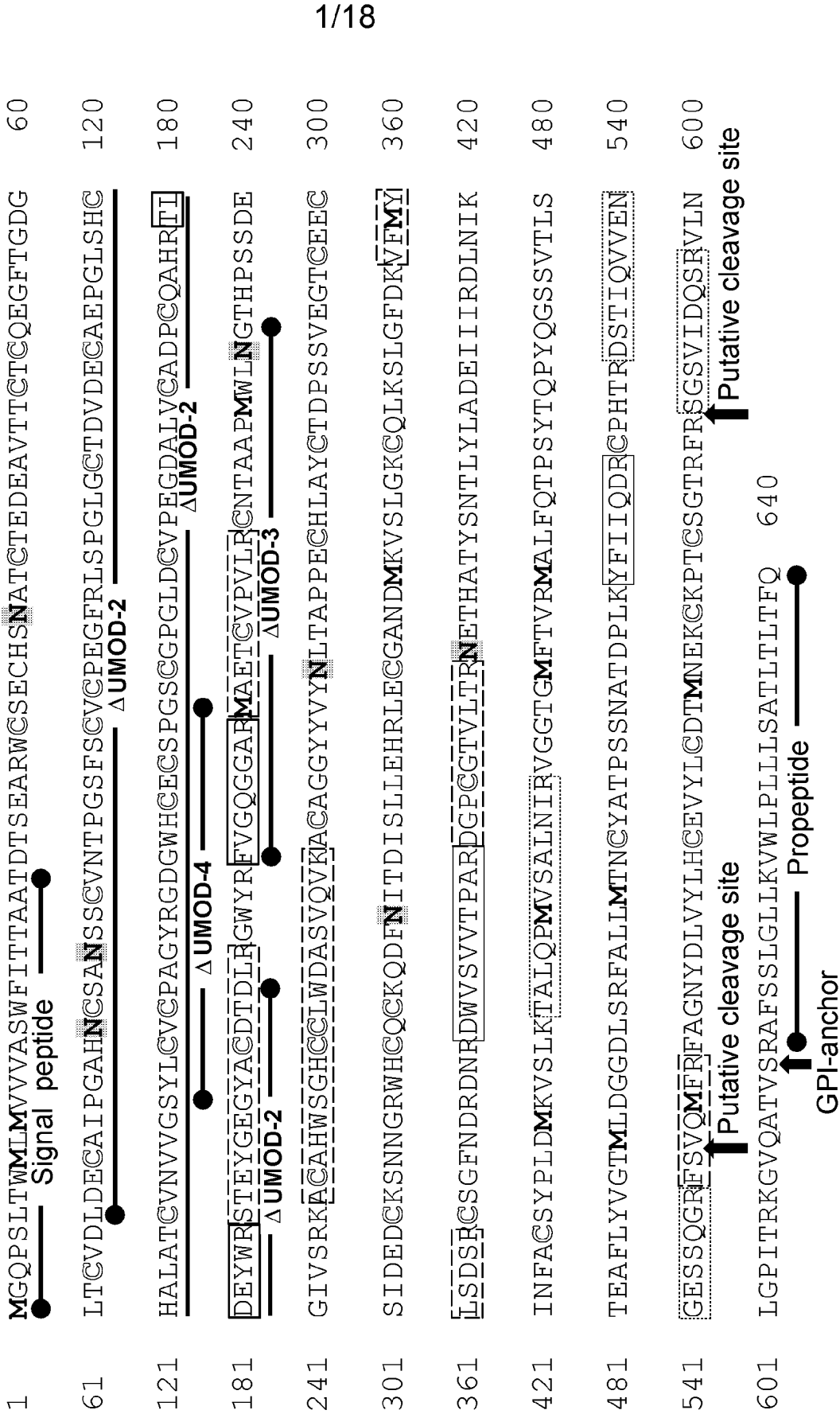
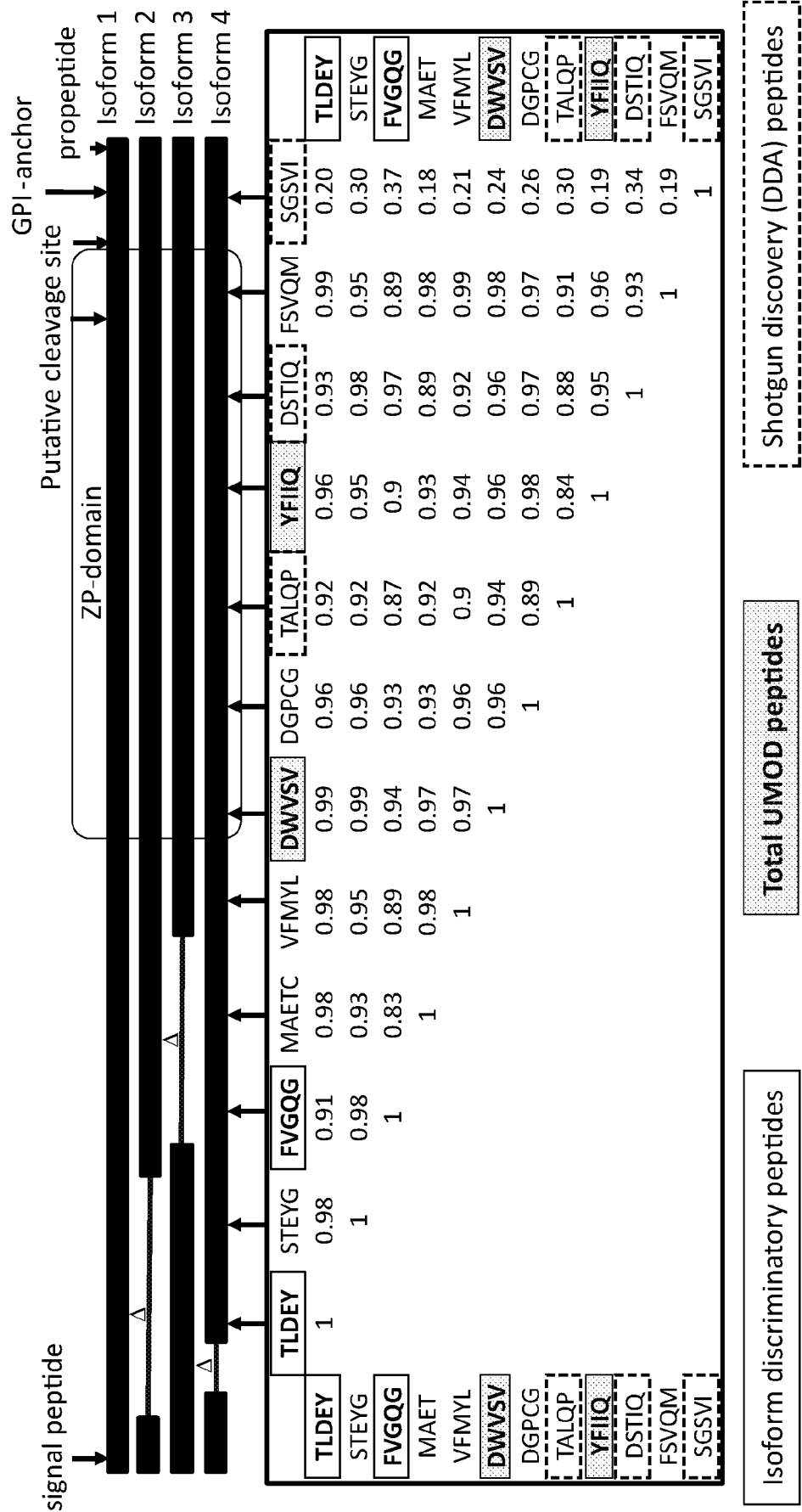
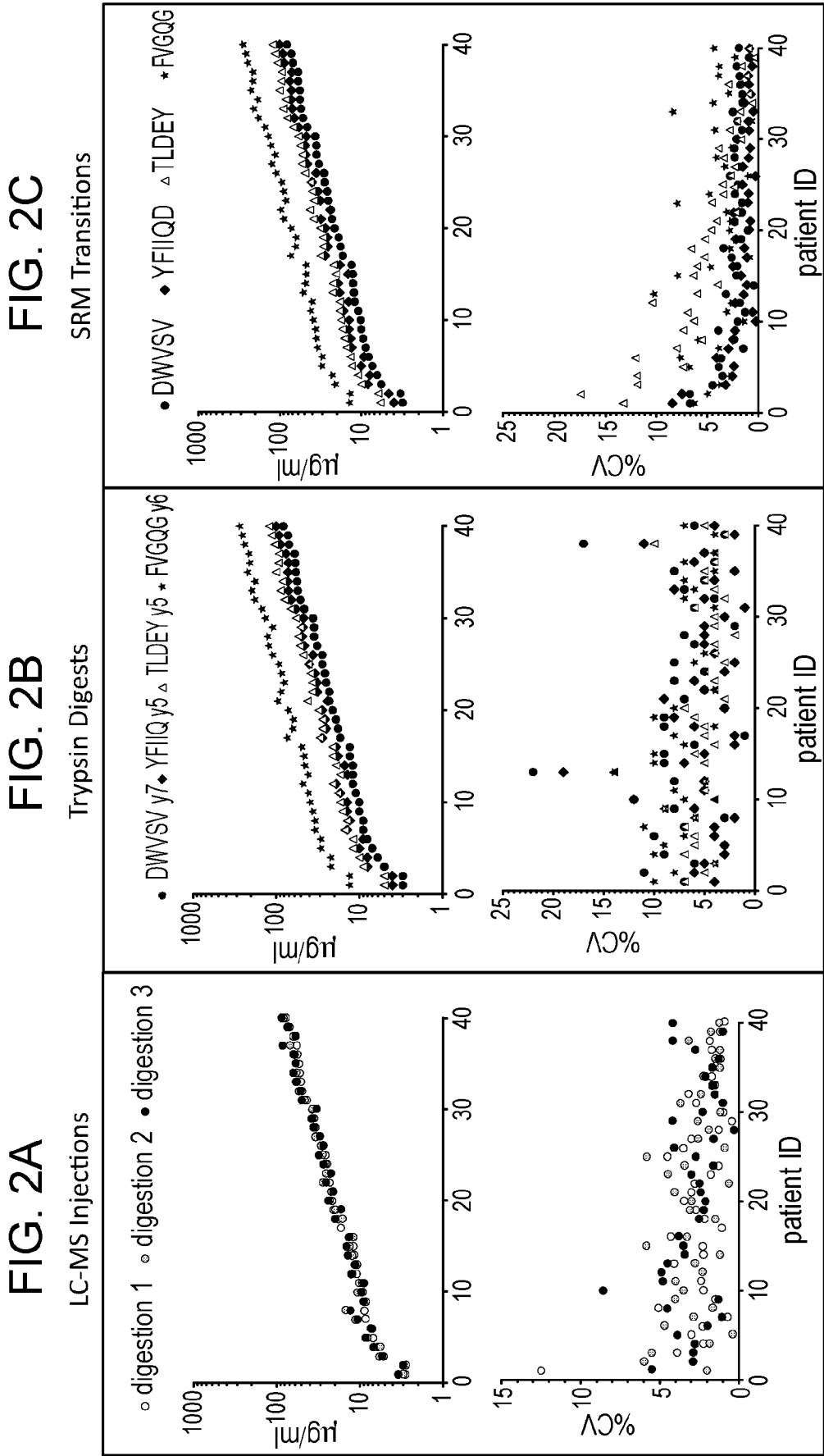


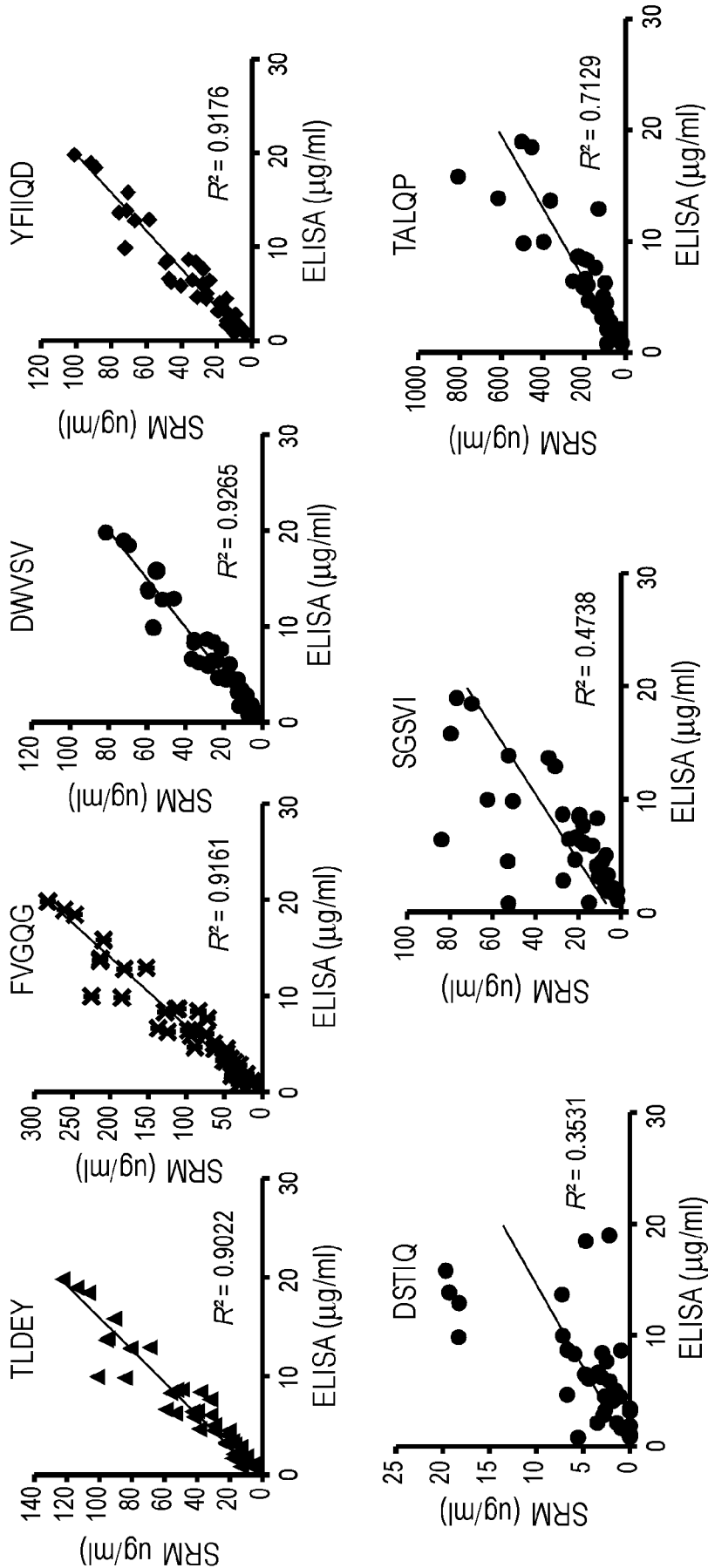
FIG. 1B



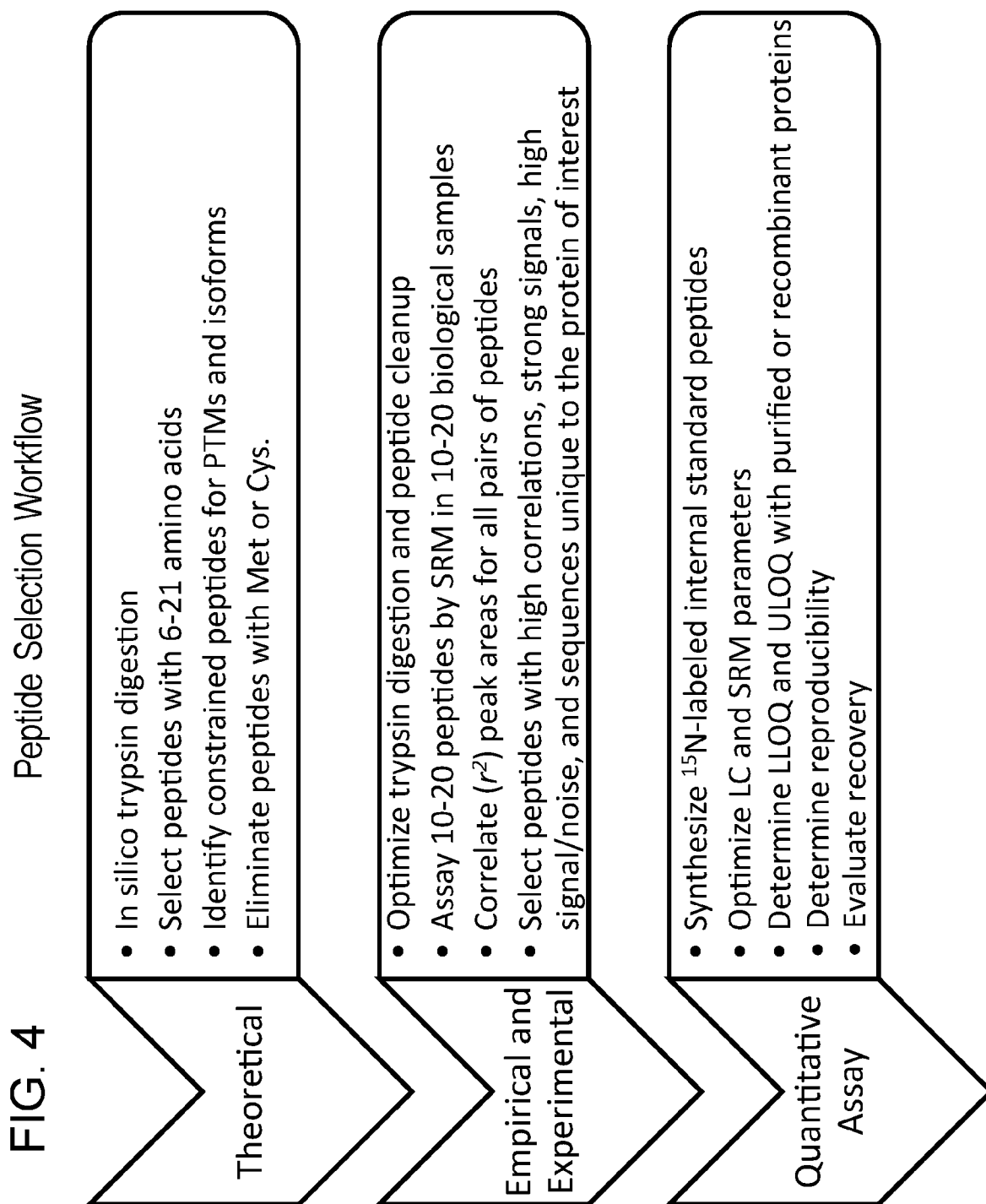


Comparison of empirically selected signature peptides and ELISA

	ELISA	TLDEY	FVGQG	DWVSV	YFIQ	DSTIQ	SGSVI	TALQP
ELISA	1.00	0.91	0.92	0.93	0.92	0.35	0.47	0.71
TLDEY		1.00	1.00	0.99	0.99	0.38	0.48	0.74
FVGQG			1.00	0.99	0.98	0.38	0.49	0.74
DWVSV				1.00	0.99	0.39	0.45	0.73
YFIQ					1.00	0.37	0.40	0.70
DSTIQ						1.00	0.28	0.52
SGSVI							1.00	0.59
TALQP								1.00



5/18



6/18

FIG. 5

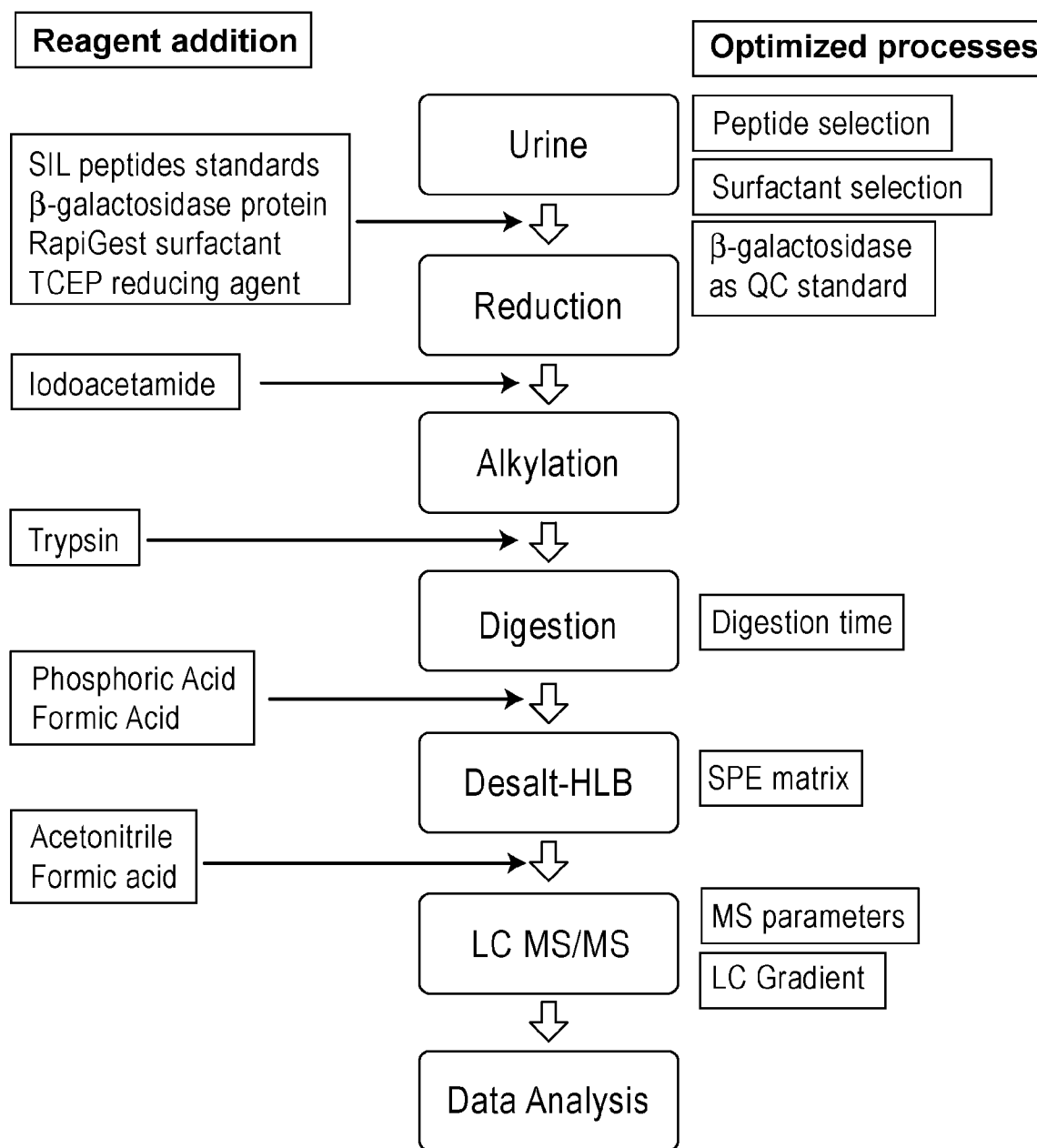
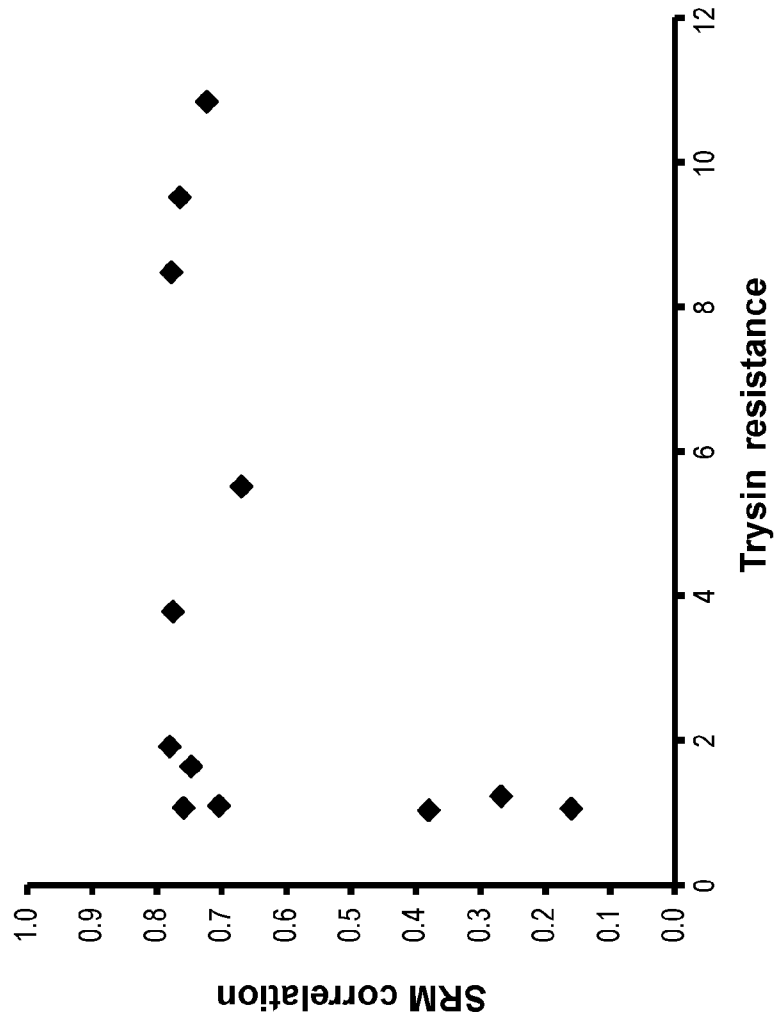
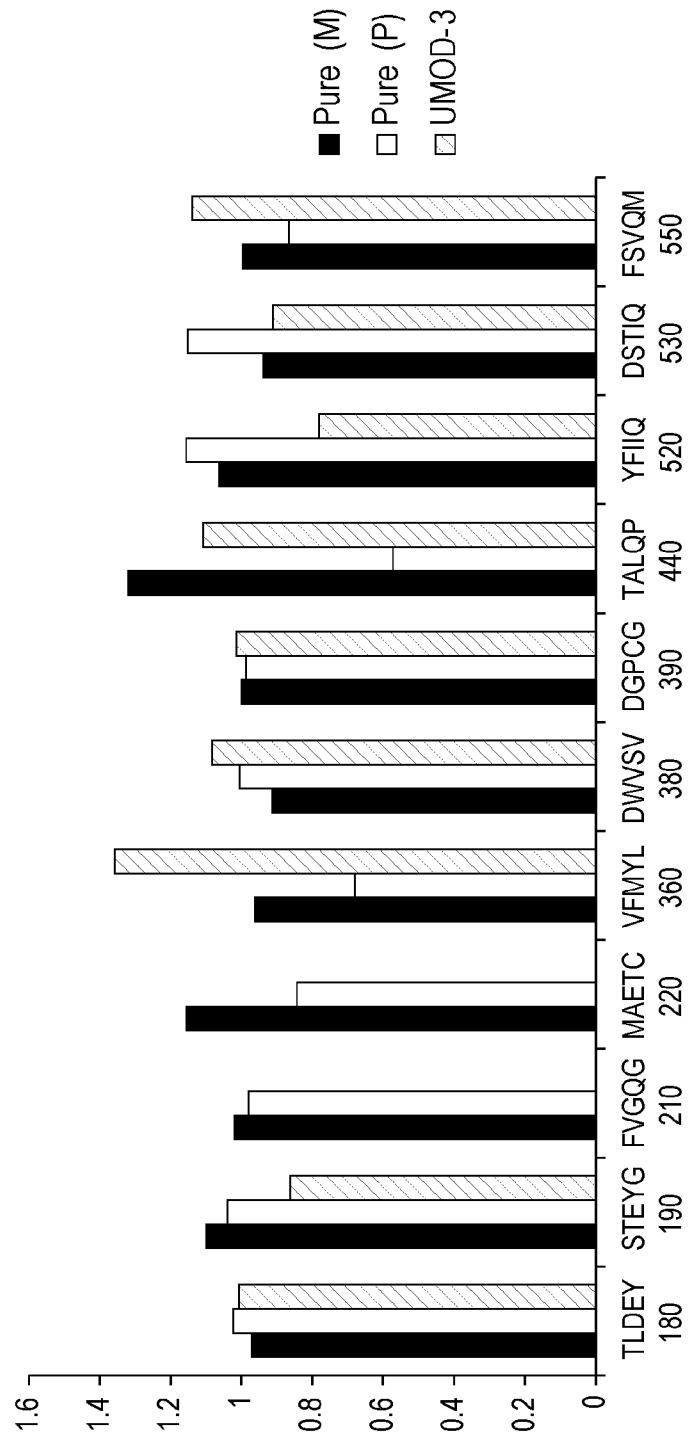


FIG. 6



8/18

FIG. 7



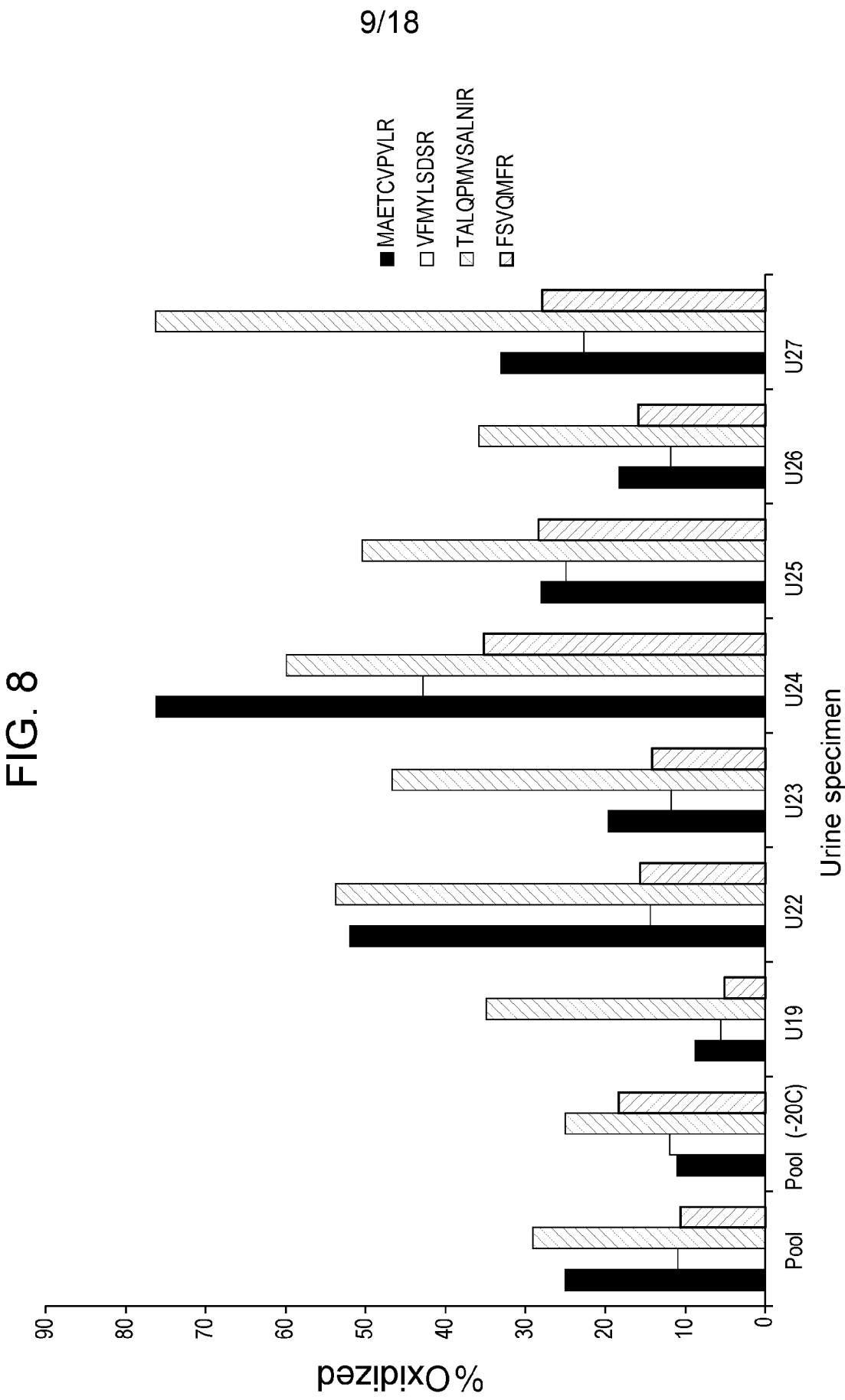
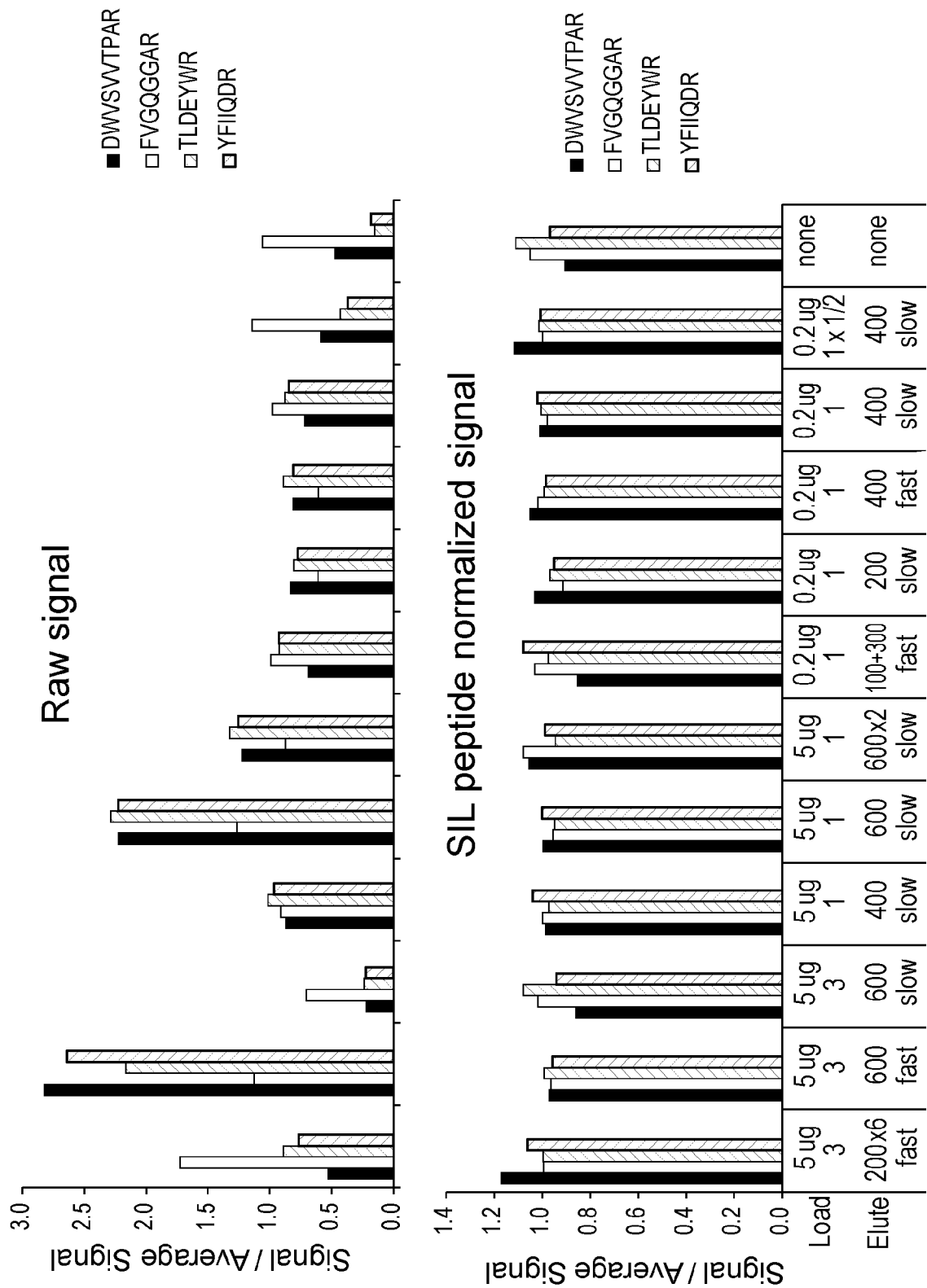


FIG. 9



11/18

FIG. 10

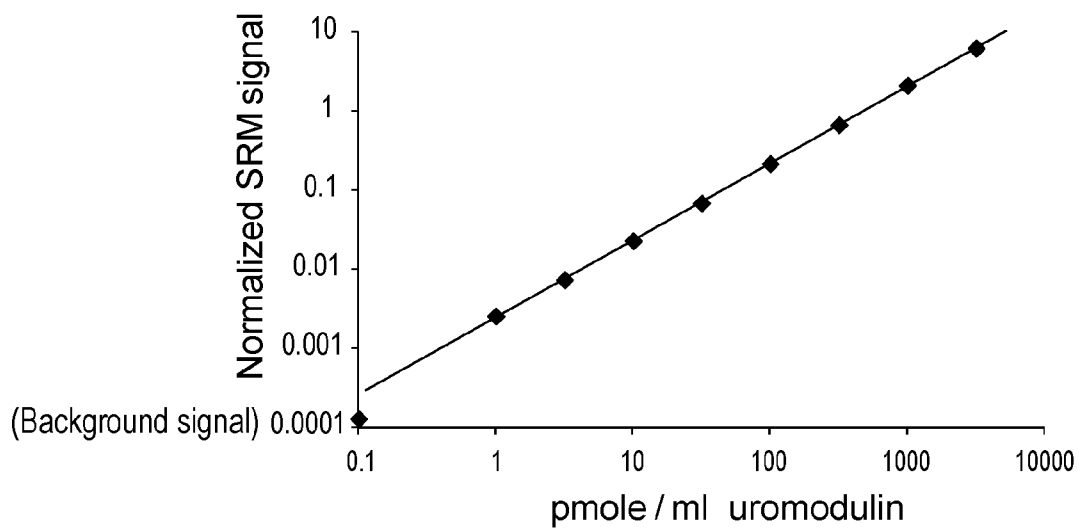
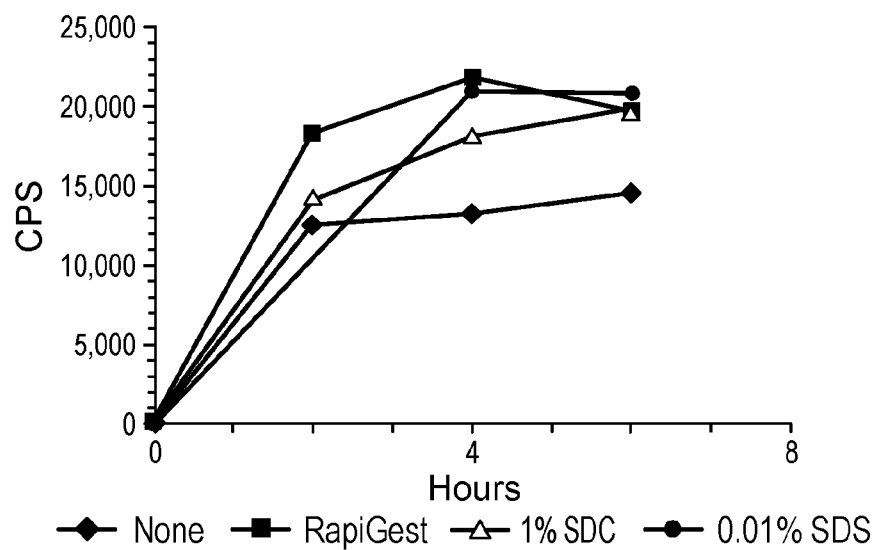


FIG. 11



12/18

FIG. 12B

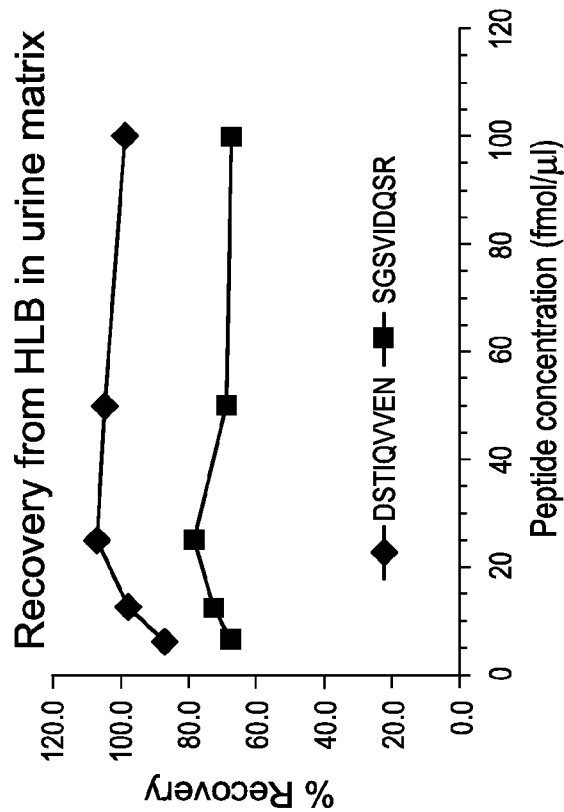
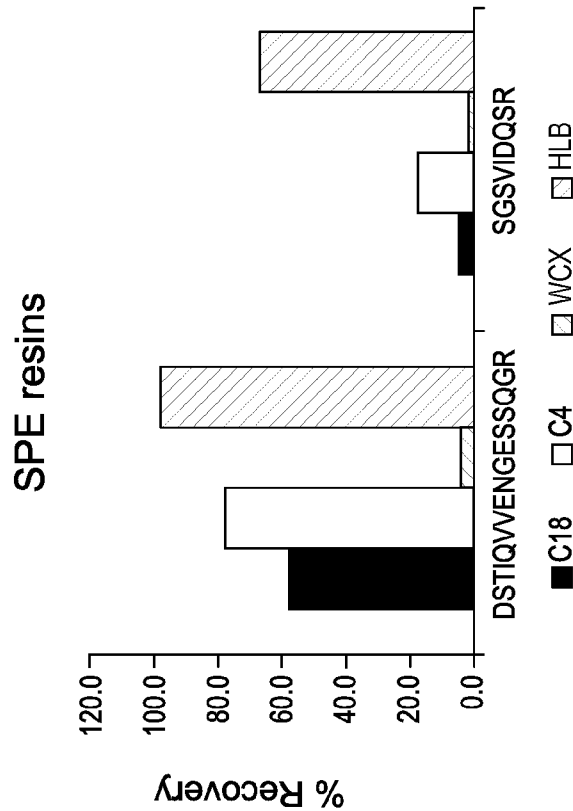


FIG. 12A



13/18

FIG. 13

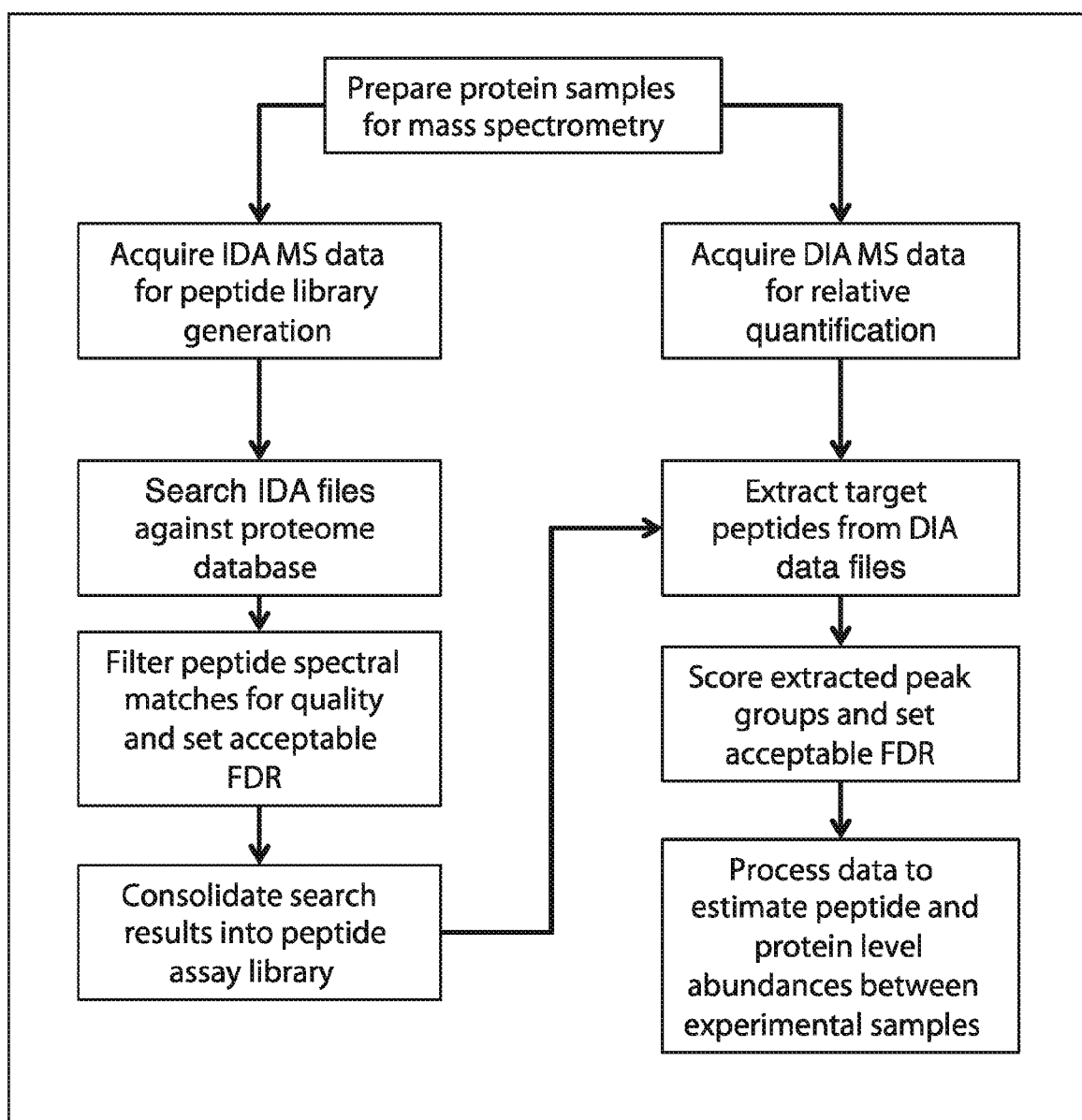


FIG. 14

MS		Advanced MS		Switch Criteria		Include/Exclude		IDA Advanced	
Experiment:		1		☒ IDA Experiment		Create IDA Exp		Create SWATH™ Exp	
Scan Type:		TOF MS		TOF Masses (Da)		Min: 400		Max: 1250	
Accumulation time:		0.249939 (secs)		Polarity		☒ Positive		☐ Negative	
								Edit Parameters	
Period		143.988 (mins)		Cycles:		3757		Delay Time: 0 (secs)	
Cycle time:		2.2995 (secs)		Period:		1			

Acquisition method

- ☒ Acquisition method
 - ☒ Mass Spectrometer 143.988 mins
 - ☒ Period 143.988 mins
 - ☒ TOF MS (+)
 - ☒ Product Ion (+) IDA
 - ☒ Gradient 1
 - ☒ Loading Pump
 - ☒ ekspert nanoLC 400

FIG. 15

Acquisition method

☒ MS Acquisition method

☒ Mass Spectrometer 143.989 mins
☒ Period 143.989 mins
☒ TOF MS (+)
☒ Product Ion (+) IDA
☒ Gradient 1
☒ Loading Pump
☒ ekspert nanoLC 400

MS

Advanced MS

Switch Criteria

Include/Exclude

IDA Advanced

Survey -> IDA Experiment

☒ For ions greater than: 400 m/z
☒ For ions smaller than: 1250 m/z
☒ With charge state 2 5
 Which exceeds 150 cps
☐ Exclude isotopes within 4 Da
 Mass Tolerance 50 ppm
 Maximum number of candidate ions to monitor per cycle: 20

Exclude former target ions

☐ Never
☒ Always For 12 seconds
☐ After 1 occurrence(s)

Mass Defect Fiber

☐ Exclusively
☒ Peptides
☐ Metabolism
 Tolerance: +/- 40 mDa

1

Formula

MW (Da)

Width (Da)

Mass Defect (mDa)

FIG. 16

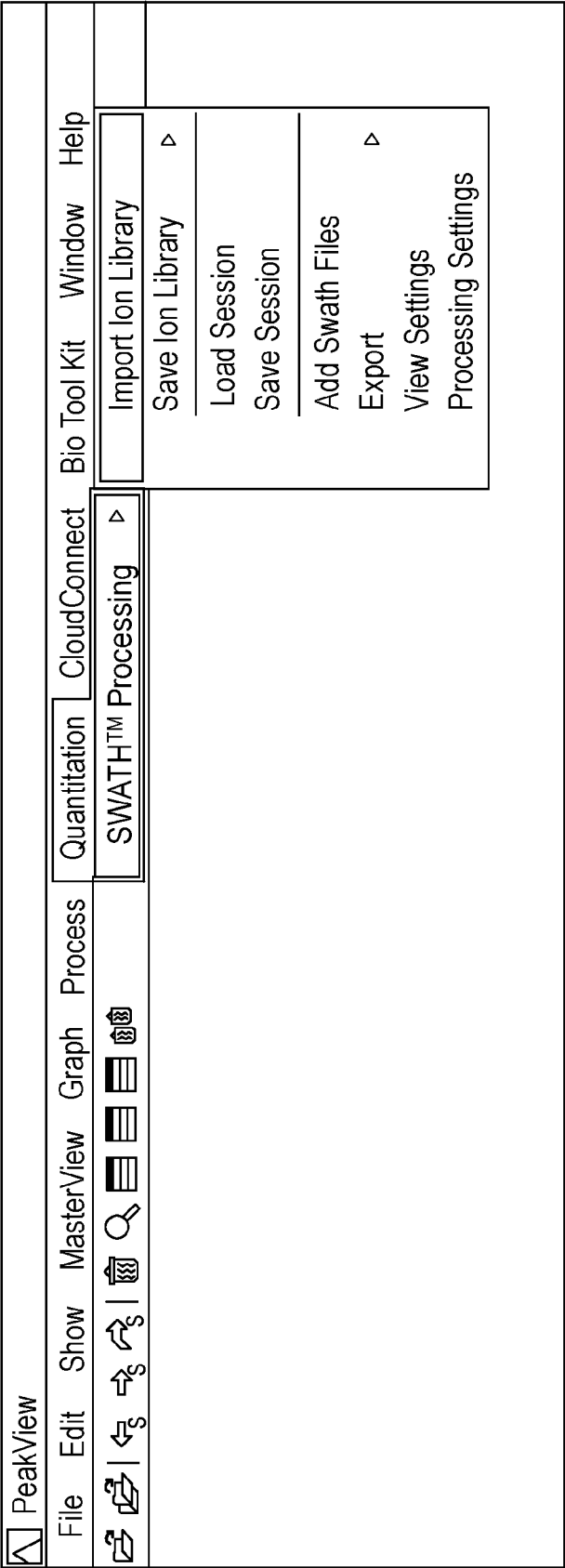


FIG. 17

Processing Settings

Peptide Filter:

Number of Peptides per Protein:

Number of Transitions per Peptide:

Peptide Confidence Threshold % (0-99):

False Discovery Rate Threshold % (0-100):

☒ Exclude Modified Peptides

☒ Exclude Shared Peptides

☐ Fix Rank

5

5

95

1.0

XIC Options

XIC Extraction Window (min):

☐ XIC width (ppm):

☒ XIC width (Da):

☐ Clear Manual Selections

5.0

75

0.050

OK

Cancel

FIG. 18

PeakView - [RH_150206_6600_NRVM_Cyto_030min_IonLibrary_GLO1-GAPDH-iRT_Test.swath]	
File Edit Show MasterView Graph Process	CloudConnect Bio Tool Kit Window Help
	Quantitation SWATH™ Processing
Detection rate 96.3% (26 of 27) for 27 targeted peptides < 1.0% F Current Sample: [TIC from RH_150209_SWATH_6600_NRVM_C]	
SWATH™ Processing	
Import Ion Library ▾ Save Ion Library ▾ Load Session Save Session Add Swath Files	
Processing Settings ▾ Process 150209_SWATH_6600_NRVM_Cyto_Control_2	
Export ▾ View Settings Processing Settings	
Areas Score FDR Observed Retention Time All	
Proteins ☐ Show Selected Only Text Search:	
☒ N	Accession
☐ -1	[RT-Cal protein] Retention time calibration protein
☒ 69	iRT_protein
☒ 107	sp P04406 G3P_HUMAN Glyceraldehyde-3-phosphate dehydrogenase OS-Homo sapiens GN-GAPDH PE-1 SV-3
☒ 270	sp Q04760 LGUL_HUMAN Lactoylglutathione lyase OS-Homo sapiens GN-GLO1 PE-1 SV-4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/35117

Box No. 1 Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/35117

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G01N 33/68; C12Q 1/68 (2016.01)

CPC - G01N 33/6803, 33/6848, 33/6851, 33/68; C12Q 1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8): G01N 33/68; C12Q 1/68 (2016.01)

CPC: G01N 33/6803, 33/6842, 33/6848, 33/6851, 33/68; C12Q 1/68

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data, RU, AT, CH, TH, BR, PH); Google Scholar; PubMed; EBSCO; identifi*, determin*, 'signature peptide', 'mass spectrometry', MS, correlat*, orthologous, quantif*, polypeptide, protein, pairwise, 'pair-wise', peptide, targeted, 'data independent acquisition', DIA, SWATH, 'sequential window acquisition', data, rank

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2002/0037532 A1 (REGNIER, FE et al.) March 28, 2002; paragraphs [0013], [0015], [0017], [0020], [0022], [0025], [0026], [0032], [0038]-[0039], [0062], [0069], [0106], [0145], [0229], [0233]-[0234]	1-44
Y	US 2007/0218505 A1 (KEARNEY, P) September 20, 2007; paragraphs [0020]-[0021], [0023], [0028], [0036], [0039], [0093], [0119] figures 6, 10	1-44
Y	US 2014/0236497 A1 (BIOGNOSYS AG) August 21, 2014; paragraphs [0003]-[0004]	2-3
Y	(BARK, SJ et al.) Linear and Accurate Quantitation of Proenkephalin-Derived Peptides by Isotopic Labeling with Internal Standards and Mass Spectrometry. Anal Biochem. June 1, 2009, Vol. 389, No. 1; page 4, paragraph 4; page 5, paragraphs 1-2; figure 2; DOI:10.1016/j.ab.2009.03.010.	4, 15-16
Y	(WORBOYS, JD et al.) Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. Nat Methods. October 2014. Vol. 11, No. 10; page 7, paragraph 4; DOI: 10.1038/nmeth.3072.	11-13
Y	(ENG, JK et al.) A Face in the Crowd: Recognizing Peptides Through Database Search. Molecular & Cellular Proteomics. 2011. Vol. 10, No. 11; page 2, column 1, paragraph 3; DOI: 10.1074/mcp.R111.009522-2.	14
Y	US 2006/0029977 A1 (ISTA SpA) February 9, 2006; paragraphs [0021]-[0022], [0048], [0075]	19-23, 25-30
Y	US 2013/010274 A1 (MENG, X et al.) November 21, 2013; paragraphs [0011], [0013]	40



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

07 August 2016 (07.08.2016)

Date of mailing of the international search report

16 SEP 2016

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/35117

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	(PARNAS, D et al.) Highly sensitive ELISA-based assay for quantifying protein levels in neuronal cultures. Brain Research Protocols. 1998. Vol. 2; abstract	42
Y	US 2005/0095649 A1 (AEBERSOLD, RH et al.) May 5, 2005; paragraph [0016]	24