

Global Proteomics: Pharmacodynamic Decision Making via Geometric Interpretations of Proteomic Analyses

Paul Kearney^{1*}, Nathan L. Currier², Daniel Chelsky², Clarissa Desjardins², Patrice Hugo², Joanna Hunter², Eustache Paramithiotis², Marc Riviere², Olivier Maes³, Howard M. Cherkow³, Hyman M. Schipper³

¹Olivier Maes, Howard M. Cherkow, Hyman M. Schipper

²Caprion Proteomics, Montreal, QC, Canada, H4S 2C8

³Centre for Neurotranslational Research, Lady Davis Institute for Medical Research, Sir Mortimer B. Davis Jewish General Hospital, and McGill University, Montreal, QC, Canada, H3T 1E2

*Corresponding author: Paul Kearney, Institute for Systems Biology, Seattle, 1441 North 34th Street, WA, USA, Fax: 98103-8904, 206 302 9929, 206 732 1299; E-mail: pkearney@systemsbiology.org

Received August 22, 2008; Accepted October 02, 2008; Published October 03, 2008

Citation: Paul K, Nathan LC, Daniel C, Clarissa D, Patrice H, et al. (2008) Global Proteomics: Pharmacodynamic Decision Making via Geometric Interpretations of Proteomic Analyses. *J Proteomics Bioinform* 1: 315-328. doi:10.4172/jpb.1000040

Copyright: © 2008 Paul K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Disease and drugs can modulate the concentrations of hundreds of proteins in the blood which can be accurately measured using contemporary proteomic methods. Nevertheless, it is common practice to reduce the plurality of disease and drug effects by a few proteins for the pragmatic purposes of immunoassay development. The vast majority of putative biomarkers discovered by this reductionist approach never reach the clinic for two reasons: the prohibitive time and cost of assay development and the acute risk of a reduced protein panel failing when validated on a broader cross-section of the population.

Global Proteomics is an alternate methodology where all blood proteins modulated by disease or drug are used to resolve pharmacodynamic questions without the time, cost, and risk of developing an immunoassay. The Global Proteomic approach was applied to an Alzheimer study where it was demonstrated that a large panel of plasma proteins is predictive of disease severity (as measured by the Mini Mental State Examination). Furthermore, a subset of this panel was shown to be modulated by disease treatment (donepezil), thereby providing a means to quantify response to treatment. Finally, to establish that the Global Proteomics methodology has broad utility, it was also applied to a Hypertension study, illustrating that a panel of plasma proteins can also be derived that are correlated with disease severity (as measured by blood pressure). In particular, the Global Proteomics methodology can readily distinguish patients responsive and non-responsive to hypertension therapies.

The Global Proteomics approach is based upon a bioinformatics analysis approach which clusters samples by proteomic similarity and then uses a geometric representation of sample similarity to answer common pharmacodynamic questions.

Keywords: Alzheimers; Bioinformatics; Biomarker; Hypertension; Pharmacodynamic; Plasma proteomics

Introduction

Blood-based biomarkers of disease and drug treatment have been the focus of intense interest in recent years. It is widely believed that most diseases and drugs modulate protein concentration in the blood. Indeed, the promise of per-

sonalized disease treatment assumes the existence of (predictive) markers of drug response in blood. Biomarker strategies are being widely adopted by the pharmaceutical industry (Mattingly and Saxberg, 2005) and play a central role in the

FDA's Critical Path Initiative (FDA, 2004). Finally, and perhaps most importantly, blood is the most accessible and least invasive sample for biomarker and diagnostic assays.

Contemporary proteomic methods are able to accurately measure the modulation of low abundance proteins in the blood. Multiple studies demonstrate that proteomic platforms have low sample to sample variation (CV ~ 14%) and high correlation between proteomic measurements and actual differential protein abundance in plasma ($R^2 \sim 0.99$) (Roy et al., 2004; Wiener et al., 2004; Silva et al., 2005). These figures of merit approach the precision and accuracy of ELISA technology. State of the art proteomic platforms routinely track 50000 plasma peptides reproducibly and accurately (Roy et al., 2004; Follettie et al., 2006). Furthermore, advances in protein and peptide separation technologies coupled with mass and retention time fingerprinting methods for protein identification enable proteomic platforms to identify plasma proteins at the ng/ml level (Conrads et al., 2000; Strittmatter et al., 2003; Adkins et al., 2005; Chen et al., 2005; Lekpor et al., 2007; Anderson and Anderson, 2002). These advancements allow proteomics to address the wide dynamic range of protein concentrations in blood, estimated to be 10 to 12 orders of magnitude (i.e. the diameter of the sun compared to the diameter of an orange) (Anderson and Anderson, 2002). An excellent overview of the state of the plasma proteomics can be found in the summary publication of the HUPO Plasma Proteome Project wherein standardized plasma and serum samples were analyzed by 18 participating labs using a variety of proteomic platforms (Omenn et al., 2005).

Despite the availability of blood samples and proteomic technologies for blood analysis, plasma biomarkers have not had the widespread impact in drug development anticipated. Recent publications have emphasized that plasma biomarker assays are not reaching the clinic because of the daunting post-discovery tasks of assay development and validation

(Aebersold et al., 2005; Cottingham, 2006; Anderson, 2005; Rifai et al., 2006; Anderson and Hunter, 2006). Although disease and drug may modulate the concentration of hundreds of blood proteins, only a few proteins can be developed into an immunoassay panel due to time and cost factors. Consequently, this reductionist approach attempts to quantify the widespread effects of disease and drug in blood using only a few peptides or proteins. Furthermore, small peptide or protein panels have an increased risk of failure when validated on a larger cross-section of the population. Unfortunately, the time and cost of assay development will have already been borne before it is known whether the panel passes or fails the validation test. There are many examples of the reductionist approach using SELDI technology (Petricoin et al., 2002a; Gillette et al., 2005). These biomarker panels are anonymous peptides in that they are defined by mass but not sequence. A well-known example of this approach generated a panel of peptides that distinguished ovarian cancer samples from healthy controls (Petricoin et al., 2002b). Re-examination of the data revealed design flaws in this study (Baggerly et al., 2004) which has had the positive effect of greater attention to study design in the proteomics community (Coombes et al., 2005; Boguski et al., 2003).

Global Proteomics is an alternative methodology where the proteomic analysis itself is the assay. This has three important advantages over the reductionist approach. First, the same technology is used for discovery and assay. Consequently, there are no development costs in the Global Proteomics approach. Second, there is no need to reduce the set of all modulated proteins for the purposes of cost-effective assay development. This ensures that the Global Proteomics assay does not compromise on sensitivity and specificity during the biomarker validation phase. Third, the Global Proteomics assay provides system-wide mechanism of action insights since the assay profiles the entirety of detectable proteins modulated by disease and drug.

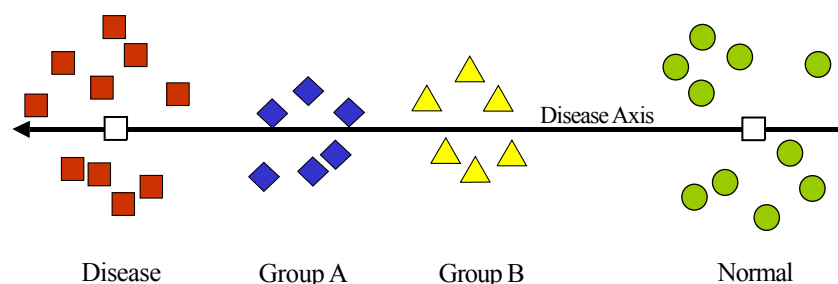


Figure 1: The concept of the disease axis and disease severity.

Geometric tools for visualization and quantitation are required to perform Global Proteomic assays. Specifically, a clustering technique such as multidimensional scaling (MDS) (Cox et al., 2001) or principal component analysis (PCA) is applied to the proteomic dataset to obtain geometric relationships among the samples. This geometric positioning of samples is based upon overall sample similarity and dissimilarity. Pharmacodynamic questions are then resolved by interpreting the geometric relationships among sample groups. To illustrate, consider the hypothetical clinical study presented in Figure 1.

Figure 1 presents a study with four sample groups: Normal, Disease, Group A and Group B where samples have been geometrically clustered by similarity. The empty boxes represent the geometric medians or centroids of the Normal and Disease groups. The line from the Normal centroid to the Disease centroid is called the disease axis. The disease axis hypothesis is that the location of samples along the disease axis correlates to disease severity. For example, samples closer to the Normal centroid (i.e. closer to normality) are healthier than those closer to the Disease centroid. Depending on the groups in the study, this permits various pharmacodynamic interpretations of the data. For example:

Dose Optimization: If Group A and Group B are two doses of the same drug treatment then the dosage administered to Group B is more efficacious since Group B samples are closer to normality.

Compound Selection: If Group A and Group B are two clinical compounds then the compound administered to Group B is more efficacious since Group B samples are closer to normality.

Patient Segregation: If 12 patients in a clinical study are administered the same drug then those in Group B had a better response to the drug than those in Group A.

Peptides that contribute most significantly to disease severity can be readily obtained and standard mass spectrometry (MS) techniques can be applied to identify proteins from which these peptides are derived. These proteins can be classified into biological processes, pathways, cellular locations, etc. using tools such as DAVID (Denis, 2003). This enables drug mechanism of action and disease biology insights. If desired, proteins or peptides can even be selected for immunoassay or MRM (Multiple Reaction Monitoring) development.

In this paper we introduce the Global Proteomics analysis technique and apply it to an Alzheimer proteomic study with

33 healthy controls, 19 untreated, early stage Alzheimer patients and 25 donepezil-treated, early stage Alzheimer patients. Alzheimer's diagnosis and severity is performed using a collection of tests including the Mini Mental State Examination (MMSE) (Folstein et al., 1975). To date, there is not an approved blood test for the diagnosis of Alzheimer's disease. This is of considerable concern as an estimated 4.5 million Americans have Alzheimer's disease which has more than doubled since 1980 and is expected to continue growing as the population ages (Hebert et al., 2003). National direct and indirect annual costs of caring for individuals with Alzheimer's disease are at least \$100 billion, according to estimates used by the Alzheimer's Association and the National Institute on Aging (Ernst and Hay, 1994).

The primary results include the discovery of 282 plasma peptides that predict Alzheimer disease severity as measured by the Mini Mental State Examination (MMSE). Furthermore, a subset of this panel is shown to be modulated by disease treatment (donepezil) thereby providing a means to quantify response to treatment. Along the way, novel visualization and data analysis techniques are introduced to enable this new time and risk efficient approach to pharmacodynamic biomarker development.

The final goal of this paper is to demonstrate that the Global Proteomics methodology is applicable to many indications, not only Alzheimer's. To achieve this, Global Proteomics was applied to a hypertension plasma study with 14 controls, 15 hypertensive patients responsive to treatment and 10 hypertensive patients not responsive to treatment. Results for the hypertension study demonstrate that blood pressure and global proteomic disease severity are highly correlated. Furthermore, the Global Proteomics approach clearly segregated responders and non-responders to hypertension treatment.

Materials and Methods

Alzheimer's Clinical Human Plasma

Patients meeting the National Institute of Neurological and Communicative Disorders and Stroke - Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable sporadic Alzheimer's disease (AD) patients were recruited from the Sir Mortimer B. Davis Jewish General Hospital (JGH)/McGill University Memory Clinic, a tertiary care facility for the evaluation of memory loss in Montreal, Canada. All AD patients were administered the Folstein Mini-Mental State Examination (MMSE) and underwent comprehensive neuropsychological testing by Memory Clinic neuropsychologists (Schipper).

Healthy elderly controls aged 60 years and over were recruited from JGH Family Practice clinics. Healthy subjects had no memory complaints and scored within one standard deviation of age- and education-standardized normal values on a series of memory and attention tests. Clinical histories for each patient were obtained after written informed consent was obtained from all subjects or their primary caregivers with approval by the Research and Ethics Committee of the JGH. After screening samples for age matching and therapy regimens, 33 healthy (i.e. control) patient samples, 19 untreated, mild-Alzheimer's (early stage) patient samples and 25 donepezil-treated, mild-Alzheimer's (early stage) patients samples were analyzed using the Global Proteomic method. Supplementary table 1 provides clinical information for the 77 study samples.

Hypertension Clinical Human Plasma

Hypertensive and hypertensive-controlled patient plasma samples were purchased from SeraCare Life Sciences Inc.'s (Gaithersburg, MD, USA) line of BioRepository (BioBank) Clinical Specimens. All specimens were collected under strict adherence to relevant HIPAA, IRB and informed consent procedures and were accompanied with demographics and basic medical histories. Samples were age and gender matched and screened for hypertension mono-therapy. In total, 14 healthy control samples, 15 hypertensive controlled patient samples and 10 hypertensive uncontrolled patient samples participated in the study. Controlled hypertensive is defined to be diastolic and systolic blood pressure below 90 and 140, respectively, whereas uncontrolled hypertensive is defined to be above 90 and 140, respectively (Chobanian et al., 2003). Treatment therapies included Ace Inhibitors and Beta Blockers.

Study Design

Several general rules were employed to construct an unbiased study design. First, healthy, untreated Alzheimer and treated Alzheimer samples were interleaved during sample processing and analysis. Second, randomization or explicit definition of the order of sample processing ensured that the order in which the samples were processed at each step was never repeated. Third, a sufficient number of replicates were analyzed to overcome processing bias and the unforeseen loss of samples not passing quality control checks. Samples to be compared were processed and analyzed on the same instruments with the same lot of reagents, whenever possible.

The platform variation has been measured several times with median coefficient of variation estimated to be 14.1%

(data not included). Both Alzheimer and Hypertension samples illustrate median coefficient of variation across all peptides measured in the 30% to 40% range depending on the cohort. A power analysis indicates that at least 10 samples per cohort are required to reliably detect 25% differences among cohorts in the two studies.

Plasma Sample Preparation.

Rigid sample processing and analysis procedures including quality control checks were controlled by a series of standard operating procedures (SOPs). The SOPs cover every step of the sample processing and analysis, beginning with shipment of samples. To ensure sample integrity, plasma samples were shipped on dry ice with a WarmMark[®] temperature tag (VWR, Mississauga, Ontario, Canada) included in the shipping container. Upon reception, frozen plasma samples were bar-coded, entered into the Laboratory Information Management System (Nautilus LIMS, Thermo Electron, Woburn, MA) and stored at -80°C.

To begin the sample preparation, samples were thawed, passed through 0.22 µm filters and then transferred to 24-well plates. Several standard plasma samples are processed with the study samples to monitor each step of the procedure and ensure reproducibility. Plasma samples were depleted of high abundance proteins using the Multiple Affinity Removal System[™] (MARS, Agilent Technologies, Palo Alto, CA) on an Agilent 1100 HPLC fitted with a refrigerated (4°C) autosampler and fraction collector (Bjorhall). The depletion method was a modified version of the Agilent MARS protocol (Sitnikov). Plasma samples were loaded onto the column in 150 mM ammonium bicarbonate (pH 7.8) for 10 minutes, and the unbound proteins were collected. The column was then washed for three minutes in Agilent buffer A. Bound proteins were eluted over eight minutes in Agilent buffer B. The column was then re-equilibrated in 150 mM ammonium bicarbonate (pH 7.8).

Depleted plasma samples were proteolyzed under denaturing conditions (8 M urea / 400 mM ammonium bicarbonate, pH = 8.0) with endo-LysC (Princeton Separations, Adelphia, NJ) (1:50, enzyme: total protein) for two hours, and then diluted (4:1) and proteolyzed with trypsin (Promega, Madison, WI) (1:50, enzyme: total protein) for an additional 16 hours. Following proteolysis, the peptides were desalted on a 10x10mm C18 HPLC guard column (Phenomenex, Torrance, CA). Buffer A was water/0.1% TFA, and buffer B was acetonitrile/0.1% TFA. After a two-minute wash in 2% B, the samples were eluted by a one-minute ramp up to 90% B. The column was then re-equilibrated in 2% B.

Following desalting, the samples were fractionated by SCX chromatography using a 4.6x150 mm BioBasic column (Thermo Electron, Bellefonte, PA). The Agilent 1100 HPLC was operated at a flow rate of 800 μ l/min. The mobile phase A was 5 mM ammonium formate/15% acetonitrile, and mobile phase B was 1 M ammonium formate/15% acetonitrile. The gradient was developed by moving from 2.5% to 75% B over the course of 20 minutes. Prior to injecting plasma samples, the system was verified by separating a mixture of peptides. The measured retention times of two standard peptides must be within 6 seconds of the accepted retention time. Eight fractions were collected from the separated peptides. The fractionated samples were then freeze-dried in bar-coded 24-well plates and stored at -80°C.

The distribution of fractions into 96-well plates for mass spectrometry analysis was accomplished on a Multiprobe II HT Plus (Packard, Meriden, CT) four channel liquid handler. Sample plates were then lyophilized and stored at -80°C.

Liquid Chromatography-Mass Spectrometry (LC-MS)

The LC-MS system consisted of a CapLC (Waters, Milford, MA) with a cooled autosampler and a QTOF Ultima (Waters, Milford, MA) controlled by MassLynx version 4.0 software. Samples were reconstituted in 15 μ l of water/10% acetonitrile/0.1% formic acid solution and injected onto a reversed-phase (Jupiter C18, Phenomenex, Torrance, CA) column. For the reversed-phase HPLC separation, buffer A was water/0.2% formic acid, and buffer B was acetonitrile/0.2% formic acid. The gradient started at 10% B and was ramped up to 60% B in 55 minutes. After holding at 60% B for two minutes, B was decreased to 10% for column re-equilibration before the next injection. For LC-MS survey scans, the mass spectra were acquired over 400-1600 Da at a rate of 1 spectrum/second.

Instrument performance was verified by injecting 5 μ l of a peptide standards mixture. Performance characteristics were automatically generated by the platform. The sensitivity was recorded in terms of the number of multiply-charged ions. The retention time and mass accuracy of two peptides in the standard samples were also recorded.

Sample lists were generated by the LIMS and imported into MassLynx. Samples were injected sequentially by fraction. As data was acquired from the mass spectrometer, it was automatically retrieved from the instrument computer to a central database where it was registered. Registration includes study name, sample number, fraction, and condi-

tion (e.g. healthy, disease, drug treated). The raw data was then converted into a three-dimensional isotope map format containing m/z, retention time and intensity information.

Peptide Detection and Alignment

The first step in the LC-MS data analysis is peak detection, which is the process of detecting isotopic peaks (either peptidic or non-peptidic) in the LC-MS data. Peak detection is automatically applied to every LC-MS analysis, represented as isotope maps. Isotope maps are converted into peptide maps by Savitzky-Golay smoothing in both the m/z and retention time dimensions followed by peak fitting to a four dimensional (m/z, retention time, charge and intensity) peptide isotope model. This model utilizes the difference in mass between peptide isotope peaks, retention time coincidence of peptide isotopes and the expected intensity profile of a peptide's isotopes as a function of peptide mass. The peptide map output is a listing of the m/z, charge, retention time and intensity of all peptides.

The peptide maps undergo normalization of retention time and intensity to correct for analytical variability. A dynamic and nonlinear correction algorithm for normalizing retention time across all LC-MS injections of a study is applied. First, a standard injection is selected by sorting all injections by their overall retention time offset and selecting the injection with median offset. Then the retention times of all of the other injections are normalized to the retention time of the standard injection. This software tool allows tracking between two or more LC-MS injections, independent of the LC column or mass spectrometer, or the time of the analysis. This dynamic function is able to reduce the retention time variability to less than seven seconds.

Intensity normalization is performed for each LC-MS injection. First, the intensity ratios of matched peptides are determined and the median of the distribution of ratios is calculated. A standard sample of average median intensity across all samples is selected. The intensities of all other samples are then normalized to the median intensity.

Following normalization, peptides are matched across all samples in a study. Peptides are clustered according to fraction, mass, retention time and charge using standard hierarchical clustering techniques adapted to the proteomics context. The process of peptide clustering, or grouping, of the same peptide observed in different samples across a study enables the detection of peptides that are differentially expressed. Once peptide clusters have been formed, a representative median mass and median retention time are calculated to represent the peptide cluster.

Global Proteomics Data Analysis

The Global Proteomics method consists of the following steps.

1. Unsupervised clustering of the samples is performed by applying MultiDimensional Scaling (MDS) to the peptide intensity data using Pearson correlation as the distance measure. The data is then reduced to three dimensions for visualization and subsequent data analyses. Although we assume reduction to three dimensions, any number of dimensions can be used.
2. For each study group, a group centroid is defined by determining the median value in each of the three dimensions. For visualization purposes only, a group centroid can be *gravitized*. For example, if study groups overlap significantly in three dimensional space, samples within each group can be moved a percentage distance closer to the group centroid. However, all data analyses and results are based on non-gravitized data.
3. The disease axis is the unique line that intersects the Normal and Disease group centroids. It is oriented from the Normal centroid towards the Disease centroid. For each sample in the study, its disease severity is the distance from the Normal centroid to the closest point on the disease axis to the sample (i.e. the *projection* of the sample onto the disease axis). The disease severity profile is the collection of disease severities for each sample in a group or study.
4. Peptides that correlate to the disease severity profile can be obtained by measuring the Pearson correlation between each peptide expression profile in the study and the disease severity profile and keeping those above a specified threshold.
5. Peptides correlated to the disease severity profile are submitted to mass and retention time fingerprinting (Lekpor et al., 2007) with tolerances of 18 ppm and 7 min, respectively. The database searched is version 3.14 of the Human IPI Database (Kersey et al., 2004). False Detection Rate (FDR) rates for mass and retention time searches of full human IPI protein database have been shown to be approximately 10% (Lekpor et al., 2007). Only proteins with 3 or more peptide hits are retained. Proteins identified are further filtered down to those associated to plasma, plasma membrane or extracellular localizations by the Gene Ontology (Ashburner, 2000) to focus on those proteins most likely to be secreted or shed into the blood.
6. Clustering of identified proteins into pathways, biological processes, etc. is performed using the DAVID online service (Denis, 2003).
7. All False Detection Rates (FDR) calculations are obtained using permutation tests on the raw peptide expression data by permuting samples independently of group assignment (Benjamini and Hochberg, 1995).

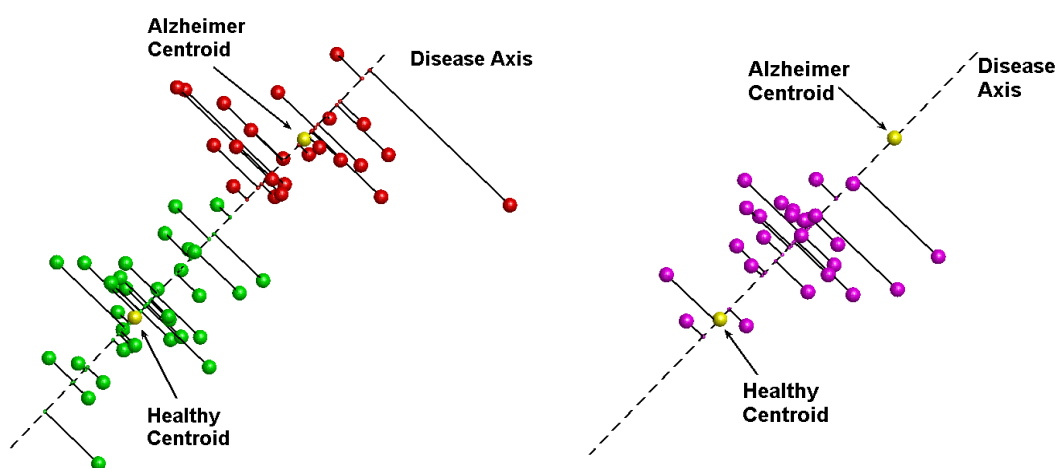


Figure 2: The results of the global proteomic analysis of the Alzheimer Study. In both plots, the disease axis (dashed line) runs from the centroid of the healthy patients to the centroid of the untreated Alzheimer patients (yellow circles). On the left, the untreated Alzheimer patients (red) and the Healthy patients (green) are ordered along the disease axis. On the right, the treated Alzheimer patients (purple) are ordered along the same disease axis. The distribution of treated patients is shifted toward the healthy centroid.

Results

Visualization of Global Proteomics

Visualization of the global proteomic analysis of the healthy (green), untreated Alzheimer (red) and treated Alzheimer (purple) patients appears in Figure 2. The treated Alzheimer patients are visually closer to normality (i.e. the healthy centroid) than the untreated Alzheimer patients. Statistically, the significance of this reversion to a healthier state has p-value 0.02.

Disease Severity Profile

Using the normal controls as a reference, the disease severity profile across the 19 Alzheimer patients was obtained and matched against the 48429 peptides profiled in the study. In total, 282 peptides matched the disease severity profile with a Pearson correlation score of at least 0.75. To ensure that a set of 282 correlating disease peptides would not occur by chance alone, 20 permutation tests were performed resulting in a FDR estimate of $6.2/282 = 2.2\%$. The 282 disease peptides and the disease severity profile appear in Figure 3. The distribution of all correlation scores appear in Figure 4.

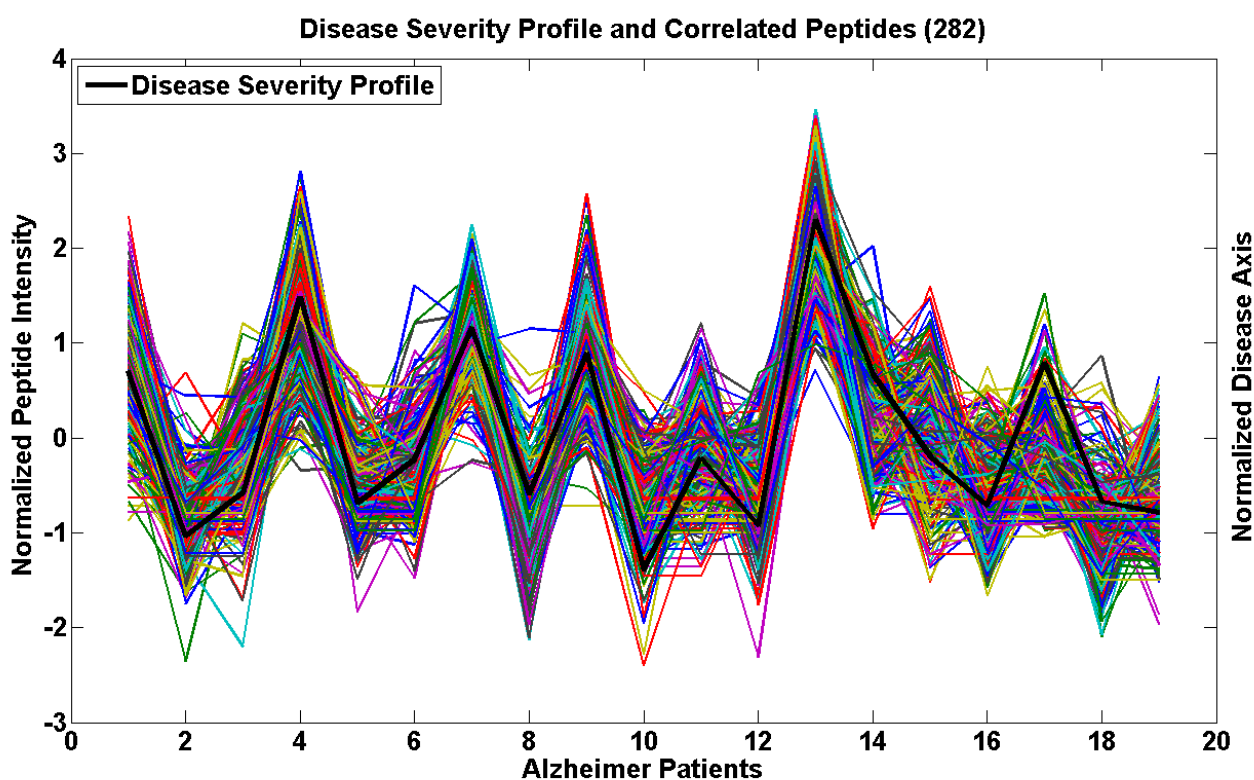


Figure 3: The disease severity profile (solid black line) for the 19 untreated Alzheimer's patients. Profiles in color represent the 282 peptides with a Pearson correlation score of at least 0.75 to the disease severity profile. The FDR for this set of 282 profiles is 2.2%.

Correlation to MMSE

To assess the relevance of the disease severity measurement, it was correlated to the MMSE scores of the 19 AD patients. The resulting Pearson correlation score is 0.75 which has p-value 0.00022 by the Student's t distribution test. This correlation appears in Figure 5. Note that MMSE test is largely language-based and is known to be affected

by education level, sensory ability and first language. More specifically, in a 331 patient study, MMSE scores were estimated to have a standard variation of 2.8 (Clark et al., 1999). Hence, a Pearson correlation score of 0.75 given the inherent variability in the MMSE scores is quite high.

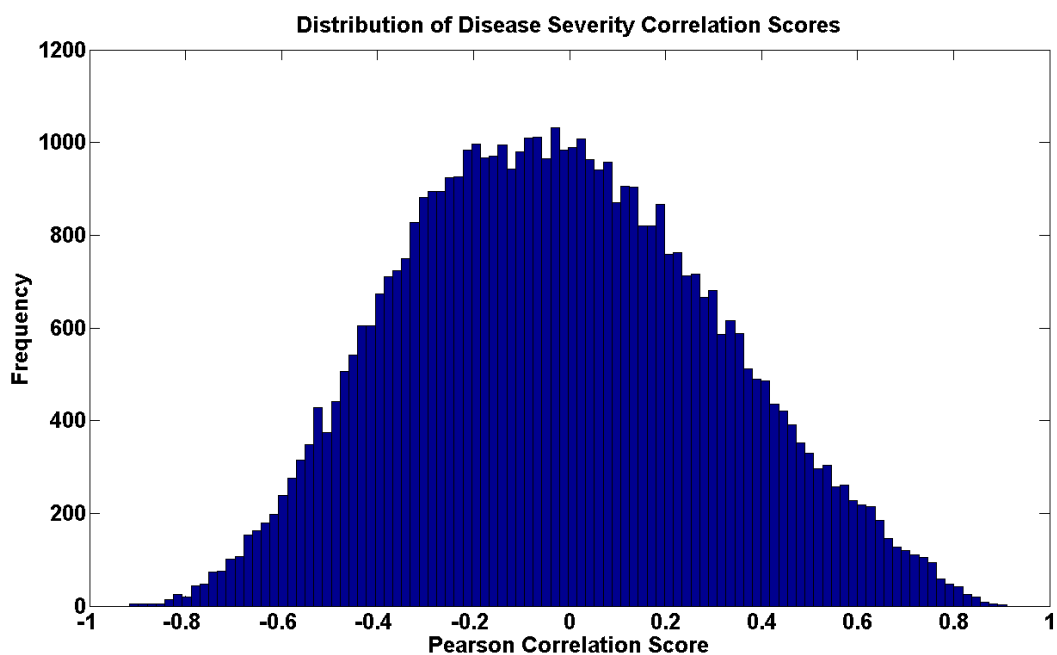


Figure 4: The distribution of peptide Pearson correlation scores to the disease severity profile. Scores close to -1 indicate a high negative correlation, scores near 0 indicate no correlation and scores near 1 indicate high positive correlation.

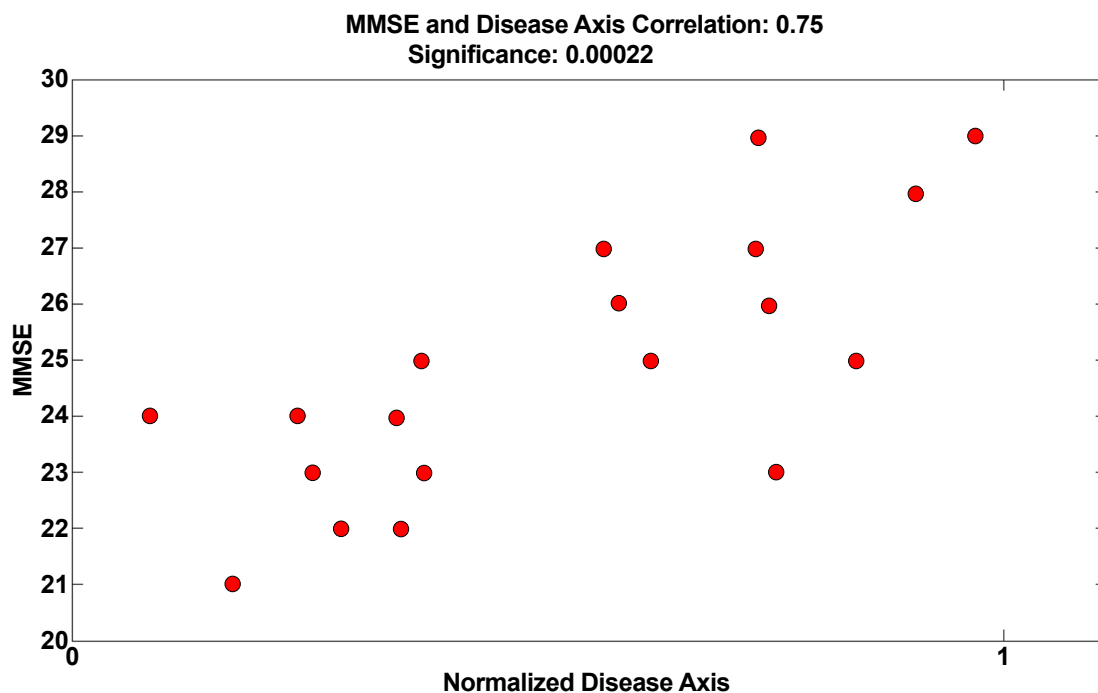


Figure 5: The correlation of patient MMSE to disease severity as measured along the disease axis. The two disease severity measures are highly correlated with significance 0.00022.

Biological Significance

The 282 disease peptides were submitted to protein identification and the resulting proteins clustered into biological processes using the online DAVID tool. DAVID clusters proteins by biological process, cellular location, molecular function, pathway, etc.

Genes associated with the processes listed in Table 1 appear in Table 2. Supplementary Tables 1 and 2 present the raw protein identification results. Note that some genes appear multiple times due to their participation in multiple biological processes.

GO Terms Biological Process	GO ID	Number of Proteins
Nervous system development	GO:0007399	6
Synaptic transmission	GO:0007268	1
Neurophysiological process	GO:0050877	5
Neurogenesis	GO:0022008	2

Table 1: Genes associated with these biological processes appear in Table 2.

Gene Symbol	Gene Name	Biological Process
UNC5C	UNC-5 HOMOLOG C (C. ELEGANS)	Nervous System Development
PCDH18	PROTOCOLADHERIN 18	Nervous System Development
PARD3	PAR-3 PARTITIONING DEFECTIVE 3 HOMOLOG (C. ELEGANS)	Nervous System Development
BDNF	BRAIN-DERIVED NEUROTROPHIC FACTOR	Nervous System Development
COL4A4	COLLAGEN, TYPE IV, ALPHA 4	Nervous System Development
ADAM23	ADAM METALLOPEPTIDASE DOMAIN 23	Nervous System Development
CHRNA1	CHOLINERGIC RECEPTOR, NICOTINIC, BETA 1 (MUSCLE)	Synaptic Transmission
CRB1	CRUMBS HOMOLOG 1 (DROSOPHILA)	Neurophysiological Process
TYR	TYROSINASE (OCULOCUTANEOUS ALBINISM IA)	Neurophysiological Process
GRIN2A	GLUTAMATE RECEPTOR, IONOTROPIC, N-METHYL D-ASPARTATE 2A	Neurophysiological Process
COL1A2	COLLAGEN, TYPE I, ALPHA 2	Neurophysiological Process
CHRNA1	CHOLINERGIC RECEPTOR, NICOTINIC, BETA 1 (MUSCLE)	Neurophysiological Process
UNC5C	UNC-5 HOMOLOG C (C. ELEGANS)	Neurogenesis
PARD3	PAR-3 PARTITIONING DEFECTIVE 3 HOMOLOG (C. ELEGANS)	Neurogenesis

Table 2: List of proteins assigned to the four biological processes in Table 1.

Markers of Drug Efficacy

The global proteomics approach was also applied to assess the effect of Alzheimer treatment with the drug donepezil on 25 Alzheimer patients. The disease severity profile across the 19 Alzheimer patients and the 25 Alzheimer treated patients was obtained and matched against the 48429 peptides profiled in the study. In total, 75 peptides matched the disease severity profile with a Pearson correlation score of

at least 0.75. To ensure that the set of 75 treatment response peptides could not occur by chance alone, 20 permutation tests were performed resulting in a FDR estimate of $0/282 = 0\%$. The 75 treatment response peptides and the disease severity profile appear in Figure 6. This set of 75 peptides is a subset of the 282 disease peptides.

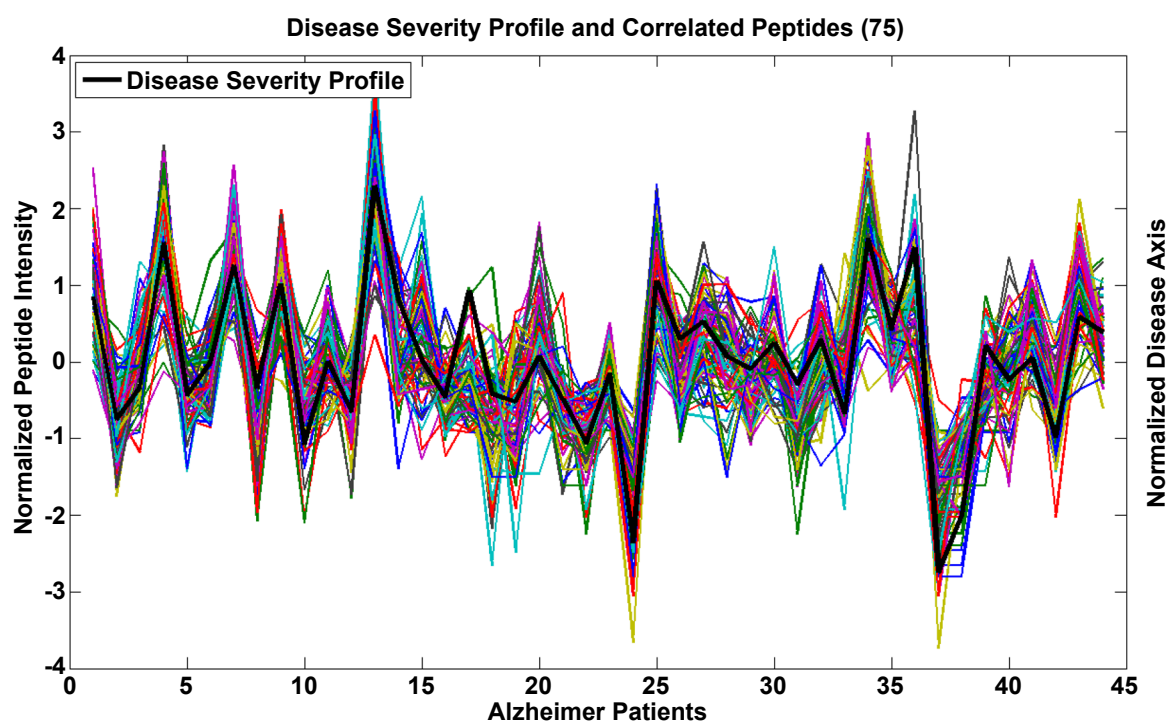


Figure 6: The disease severity profile for the 19 Alzheimer patients and the 25 Alzheimer patients treated with donepezil.

As donepezil is a clinically approved drug for Alzheimer's disease, it is expected that the plasma concentration of a subset of the 282 disease peptides would be modulated by treatment. This is indeed the case as illustrated by the 75 treatment response peptides and their FDR. Furthermore, when the log ratio of the non-treated and treated patient peptide abundances are compared, the distribution in Figure 7 is obtained. Testing against the null hypothesis that this

distribution is centered at 0 (i.e. no significant shift in peptide intensity is observed due to treatment), the null hypothesis is rejected with p -value $6.9E-$. Returning to the right panel of Figure 2, the effect of the cholinesterase inhibitor donepezil in terms of the disease axis can be seen visually. The magnitude of reversion toward the healthy centroid is significant but modest, which is consistent with the known effect of cholinesterase inhibitors (Trinh et al., 2003).

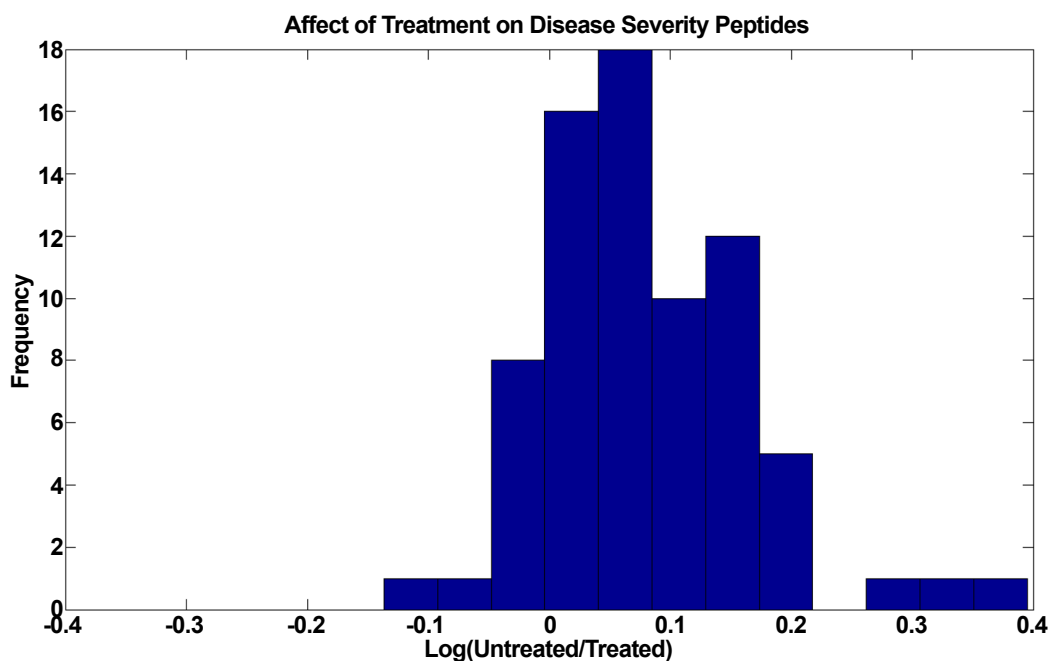


Figure 7: The distribution of the log ratios of untreated patient peptide abundances to treated patient peptide abundances in plasma. This distribution is strongly skewed to the right indicating that the plasma concentrations of the proteins associated with the 75 treatment response peptides are shifted towards healthy plasma levels.

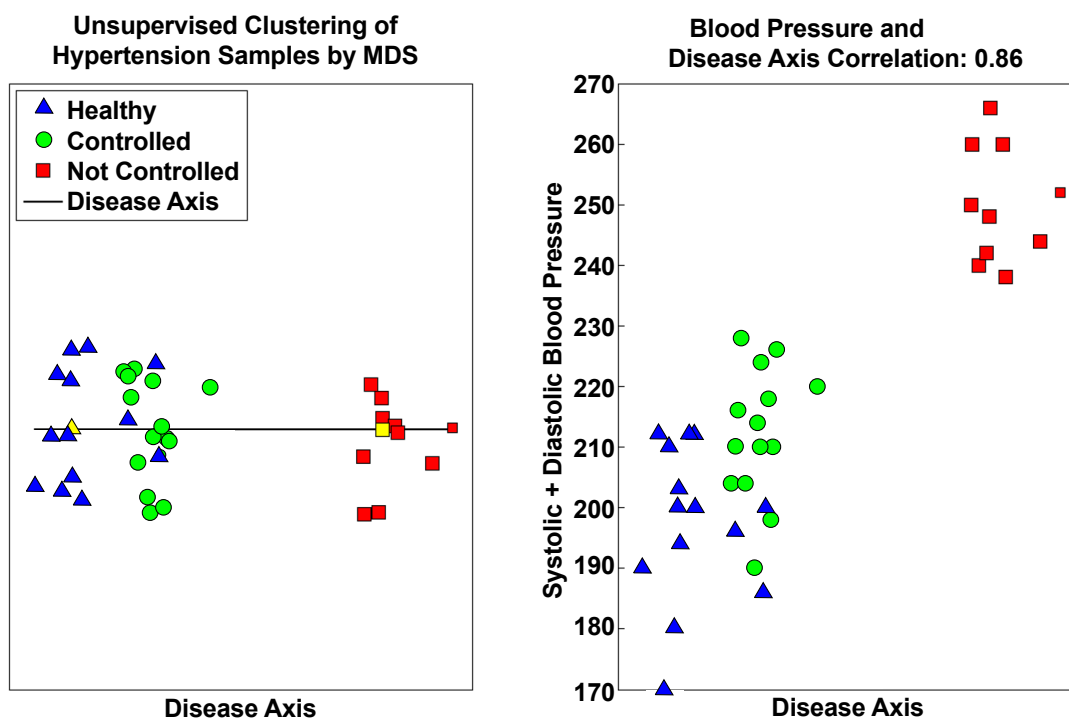


Figure 8: The Hypertension analysis as rendered by an unsupervised multidimensional scaling (MDS) analysis (left). Samples are ordered from left (normal) to right (diseased) based on their proteomic similarity. Quantification of the disease axis correlation to blood pressure is shown on the right. A high correlation (Pearson correlation = 0.86) exists between the 39 patient combined blood pressure values (systolic + diastolic) and their location on the disease axis. 1093 peptides were found to segregate the responsive and non-responsive patients.

Hypertension Study Results

To assess the accuracy of the Global Proteomic hypertension disease severity measurement, it was correlated to the combined diastolic and systolic blood pressure measurements of the 39 patients in the hypertension study. The resulting Pearson correlation score is 0.86 which has p-value $9.4e-10$ by the Student's t distribution test. This correlation appears in Figure 8. Note that the treatment-responsive and treatment-nonresponsive groups are clearly separated. 1093 peptides are primarily responsible for this segregation using the same techniques used for the Alzheimer's example above.

Discussion

The results of this work indicate that a blood-based objective measure of Alzheimer's disease severity is achievable. More generally, global proteomic techniques have broad applicability to pharmacodynamic questions: Which dose is better? Which compound is better? Which patients are more responsive to treatment? By including calibration samples in a study, quantitative classifications of samples can also be made.

Importantly, the results of the Hypertension study demonstrate that the Global Proteomics methodology can be applied broadly to studies involving different indications and drug treatments.

Disease severity (and drug response) can be measured using the Global Proteomic method. More specifically, using a sufficiently large database of Alzheimer peptide profiles and healthy peptide profiles, new patients can be classified as Alzheimer or healthy based correlation to these profiles. Such a diagnostic would likely required regulatory approval through the recently implemented IVDMA (In Vitro Diagnostic Multivariate Index Assay) process. However, a more traditional diagnostic implementation using MRM technology is discussed below.

The Global Proteomic approach described here is particularly well-suited to early stage (preclinical or R&D) applications where the investigator wants to determine if there is drug response, disease stratification, patient stratification and dosing optimization, among other questions. However, Global Proteomics is not a quantitative assay that can be used for applications such as a companion diagnostic for drug therapy. The results of a Global Proteomics analysis can justify the development of such an assay. For example, the recent development of highly multiplexed MRM validation techniques (Stahl-Zeng et al., 2007; Anderson and Hunter, 2006) are well-suited to take the results of a Global Proteomics analysis and create a quantitative clinical assay

appropriate for disease diagnosis. In this sense, Global Proteomics is an efficient tool for identifying the peptides and proteins in blood that are modulated by drug and/or disease and providing statistically significant results justifying further validation work.

Acknowledgements

The authors would like to acknowledge the insightful comments provided by the referees.

References

1. Adkins JN, Monroe ME, Auberry KJ, Shen Y, Jacobs JM, et al. (2005) A proteomic study of the HUPO Plasma Proteome Project's pilot samples using an accurate mass and time tag strategy. *Proteomics* 5: 3454-3466. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Aebersold R, Anderson L, Caprioli R, Druker B, Hartwell L, et al. (2005) Perspective: a program to improve protein biomarker discovery for cancer. *J Proteome Res* 4:1104-1109. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
3. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1: 845-867. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
4. Anderson NL (2005) The roles of multiple proteomics platforms in a pipeline for new diagnostics. *Mol Cell Proteomics* 4:1441-1444. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
5. Anderson L, Hunter CL (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* 5:573-588. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
6. Baggerly KA, Morris JS, Coombes KR (2004) Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 20: 777-785. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
7. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Stat Methodol* 57:289-300. » [CrossRef](#) » [Google Scholar](#)
8. Boguski MS, McIntosh MW (2003) Biomedical informatics for proteomics. *Nature* 422: 233-237. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
9. Chen SS, Deutsch EW, Yi EC, Li XJ, Goodlett DR, et al. (2005) Improving mass and liquid chromatography based identification of proteins using bayesian scoring. *J Proteome Res* 4: 2174-2184. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
10. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, et al. (2003) Seventh report of the Joint National Committee

- on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* 42:1206-1252. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
11. Clark CM, Sheppard L, Fillenbaum GG, Galasko D, Morris JC, et al. (1999) Variability in annual Mini-Mental State Examination score in patients with probable Alzheimer's disease: A clinical perspective of data from the Consortium to Establish a Registry for Alzheimer's Disease. *Arch Neurol* 56: 857-862. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 12. Conrads TP, Anderson GA, Veenstra TD, Pasa-Tolic' L, Smith RD (2000) Utility of accurate mass tags for proteome-wide protein identification. *Anal Chem* 72:3349-3354. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 13. Coombes KR, Morris JS, Hu J, Edmonson SR, Baggerly KA (2005) Serum proteomics profiling – a young technology begins to mature. *Nat Biotechnol* 23:291-292. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 14. Cottingham K (2006) Speeding up biomarker discovery. *J Prot Res* 5:1047–1048.
 15. Cox MF, Cox MAA (2001) *Multidimensional Scaling*, Chapman and Hall.
 16. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, et al. V(2003) DAVID: Database for Annotation, isualization, and Integrated Discovery. *Genome Biol* 4:R60. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 17. Ernst RL, Hay JW (1994) The US economic and social costs of Alzheimer's disease revisited. *Am J Public Health* 84:1261-1264. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 18. FDA (2004) Challenge and opportunity on the critical path to new medical products.
 19. Follettie MT, Pinard M, Keith JC Jr, Wang L, et al. (2006) Organ messenger ribonucleic acid and plasma proteome changes in the adjuvant-induced arthritis model: responses to disease induction and therapy with the estrogen receptor-b selective agonist ERB-041. *Endocrinology* 147:714-723. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 20. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-Mental State". A practical method for grading the state of patients for the clinician. *J Psychiatr Res* 12:189-198. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25-29. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 22. Gillette MA, Mani DR, Carr SA (2005) Place of pattern in proteomic biomarker discovery. *J Proteome Res* 4:1143-1154. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 23. Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA (2003) Alzheimer disease in the US population: prevalence estimates using the 2000 census. *Arch Neurol* 60:1119 – 1122. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 24. Kersey PJ, Duarte J, Williams A, Karavidopoulou Y, Birney E, et al. (2004) The international protein index: an integrated database for proteomics experiments. *Proteomics* 4:1985-1988. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 25. Lekpor K, Benoit MJ, Butler H, Schirm M et al. (2007) An evaluation of multidimensional fingerprinting in the context of clinical proteomics. *Proteomics - Clinical Applications* 1: 457 – 466. » [CrossRef](#) » [Google Scholar](#)
 26. Mattingly SZ, Saxberg BEH (2005) Biomarkers come of age, *Pharmaceutical Executive*.
 27. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, et al. (2005) Overview of the HUPO Plasma Proteome Project : results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 5: 3226-45. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 28. Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA (2002a) Clinical proteomics: translating benchside promise into bedside reality. *Nat Rev Drug Discov* 1:683-695. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 29. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, et al. (2002b) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359:572-577. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 30. Rifai N, Gillette MA, Carr SA (2006) Protein biomarker discovery and validation : the long and uncertain path to clinical utility. *Nat Biotechnol* 24:971-983. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 31. Roy SM, Anderle M, Lim H, Becker CH (2004) Differential expression profiling of serum proteins and metabolites for biomarker discovery. *Int J Mass Spec* 238: 163-171. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 32. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, et al. (2005) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* 77: 2187-2200. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 33. Stahl-Zeng J, Lange V, Ossola R, Eckhardt K, Krek W, et al. (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* 6:1809-1817. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 34. Strittmatter EF, Ferguson PL, Tang K, Smith RD (2003)

- Proteome analysis using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J Am Soc Mass Spectrom* 14: 980-991. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
35. Trinh NH, Hoblyn J, Mohanty S, Yaffe K (2003) Efficacy of cholinesterase inhibitors in the treatment of neuropsychiatric symptoms and functional impairment in Alzheimer disease: a meta-analysis. *JAMA* 289:210-216. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
36. Wiener MC, Sachs JR, Deyanova EG, Yates NA (2004) Differential mass spectrometry: A label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal Chem* 76:6085-6096. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)