Depletion of Abundant Plasma Proteins and Limitations of Plasma Proteomics

Chengjian Tu[1,2], Paul A. Rudnick[3], Misti Y. Martinez[2], Kristin L. Cheek[2], Steven E. Stein[3], Robbert J. C. Slebos[1.4], and Daniel C. Liebler[1,2]

[1]The Jim Ayers Institute for Precancer Detection and Diagnosis, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37232; [2]Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN 37232; [3]National Institute of Standards and Technology, Gaithersburg, MD 20899; [4]Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN 37232

Keywords: plasma, high-abundance protein depletion, multiple affinity removal system, isoelectric focusing, shotgun proteomics, futility

FOOTNOTES

[†]Abbreviations: CPTAC, Clinical Proteomic Technologies Assessment for Cancer; CV, coefficient of variation; FDR, false discovery rate; IEF, isoelectric focusing; LC-MS/MS, liquid chromatography-tandem mass spectrometry; MARS, multiple affinity removal system; TFA, trifluoroacetic acid.

[‡]MARS™ is a registered trademark of Agilent Technologies, Inc.

ABSTRACT

Immunoaffinity depletion of high abundance plasma proteins is frequently employed to enhance detection of lower abundance proteins in both shotgun and targeted proteomic analyses. Here we present a detailed evaluation of the Agilent Multiple Affinity Removal System™‡ (MARS-7 and MARS-14) in the shotgun proteomic analysis of human plasma. We used a multidimensional analysis platform that combines peptide isoelectric focusing (IEF) with liquid chromatography-tandem mass spectrometry (LC-MS/MS) to analyze unfractionated and MARS-7 and MARS-14-depleted plasma proteomes. The MARS columns afforded highly repeatable and efficient plasma protein depletions and a global enrichment in non-target plasma proteins of 2-4-fold, as assessed by MS/MS spectral counting. Immunodepletion resulted in a 30-40% increase in identified proteins compared to unfractionated plasma. Relatively few non-target proteins were captured by the MARS-7and MARS-14 columns . Although some low abundance (<10 ng mL^{-1}) plasma proteins were detected, they accounted for only 5-6% of total protein identifications in MARS-depleted plasma. Analyses of unfractionated or immunodepleted plasma yielded only 10-20% of the identifications made from identical amounts of the human colon tumor RKO cells. This disparity in proteome detection is due to a much steeper protein abundance distribution in plasma than in cells. These findings demonstrate the consistent performance of abundant protein depletion with MARS columns, but also illustrate the limitations of shotgun proteomics to detect disease biomarkers in plasma proteomes.

INTRODUCTION

Human blood plasma is an attractive biomaterial source for clinical diagnosis, since it is easily obtained through minimally invasive procedures and because many disease states produce changes in blood biomolecules. Indeed, better clinical diagnostics have been widely anticipated from application of proteomic technologies to analyze plasma proteomes [1]. A principal barrier to realizing this goal is the wide dynamic range of human plasma protein concentrations, which is thought to exceed ten orders of magnitude. The 22 most abundant proteins in human plasma comprise ~99% of the total protein mass, while lower abundance proteins derived from tissue secretion and leakage are typically present at concentrations in the nanogram per milliliter range and in combination do not make up more than 1% of the plasma protein mass [1]. Hence, depletion of these major proteins is one potential strategy for improving proteomic analyses of plasma. Abundant protein depletion has been applied mainly to increase the breadth and depth of proteome identifications in unbiased discovery, but depletion also can increase sensitivity for targeted analyses of specific proteins.

Published work describes a variety of methods to remove high abundance proteins from plasma and serum (for a review, see Pernemalm [2]). The lipophilic dye Cibacron blue was introduced to facilitate albumin purification [3] and then adapted to albumin depletion for proteomics [4, 5]. Similarly, Protein A or G can selectively deplete immunoglobulins [6, 7]. Pieper and colleagues provided a key conceptual advance by employing multiple immobilized antibodies to deplete high abundance proteins from plasma [8]. This approach is now available through over a dozen commercially-available products [2].

Among the most widely used products (based on literature citations) is the Multiple Affinity Removal System (MARS†; Agilent Technologies, Santa Clara, CA), which offers

removal of 7 (MARS-7) or 14 (MARS-14) abundant plasma proteins. An earlier product (MARS-6) sold prior to 2008 was directed at 6 abundant proteins. Our interest in the MARS-7 and MARS-14 columns is due in large part to the adoption of this system for abundant plasma protein depletion in ongoing interlaboratory studies by the National Cancer Institute Clinical Proteomic Technologies Assessment for Cancer (CPTAC) program, in which we are participants. Several previous reports have characterized the performance of the MARS-6 or MARS-7 columns for abundant protein depletion in analyses of plasma and serum. Of these, 5 used two-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis to analyze the low abundance protein fraction [8-12]. These studies yielded limited proteome inventories (<100 protein identifications) and provided an essentially qualitative assessment of the impact of depletion on detection of plasma and serum proteins.

Two more recent studies [13, 14] employed liquid chromatography-tandem mass spectrometry (LC-MS/MS) to generate larger proteome inventories (>200 protein identifications). The primary metric for assessing the impact of high abundance protein depletion in both studies was increased numbers of proteins detected. Whiteaker et al. [13] also reported increased coverage of some proteins in the depleted samples, whereas Gong et al.[14] reported detection of 10 low-abundance proteins (concentrations <100 ng mL^{-1} based on literature annotations) only in depleted samples. The only quantitative assessment of enhanced detection of low abundance proteins as a result of MARS depletion was reported by Brand et al., who reported ELISA measurements on 11 proteins [15]. We note that other publications report the use of MARS columns in sample preparation for proteome analyses, but did not specifically evaluate the impact of this step (e.g., [16, 17]).

The available literature indicates that abundant protein depletion using the MARS system enhances detection of lower abundance proteins, but the impact has not been characterized on a global proteomic scale. Moreover, previous reports did not assess the reproducibility of the depletion step or the degree to which non-target proteins were captured by the MARS columns. Most of these aspects were characterized, however, in recent studies of the IgY-14 and the Seppro SuperMix products (Genway, San Diego, CA), which deplete either 14 or more than 50 high abundance plasma proteins, respectively [18-20]. In light of the widespread use of the MARS system, the somewhat limited quantitative characterization of its performance in previous studies, and the planned use of MARS columns in CPTAC studies, we felt that a detailed characterization of this system was justified.

One striking characteristic of plasma proteome datasets is the relatively small number of proteins identified in comparison to analyses of cells or tissues. Application of one dimensional LC-MS/MS, together with commonly accepted criteria for protein identification ($\geq$ 2 peptides at <5% false discovery rate (FDR)) yielded approximately 150-250 proteins [13, 20], whereas addition of multidimensional LC-MS/MS yielded 200-700 proteins [14, 20]. Even the multilaboratory HUPO plasma proteome study yielded less than 900 high confidence protein identifications [21]. These totals are approximately 2-5-fold less than have been reported for analyses of cell and tissue extracts using similar protein input amounts and similar LC-MS/MS platforms [22-26]. Although this effect is commonly attributed to the overwhelming predominance of a small number of high abundance proteins, this disparity has not been explored with both tissue and plasma proteomes in the same experimental system.

Here we present a detailed evaluation of the MARS-7 and MARS-14 systems in the analysis of human plasma by shotgun proteomics using LC-MS/MS. We have employed a

standardized analysis platform that combines narrow-range isoelectric focusing (IEF) of peptides followed by LC-MS/MS to analyze unfractionated and MARS-7 and MARS-14-depleted plasma proteomes [24]. We evaluated the repeatability and efficiency of plasma protein depletions and evaluated the global enrichment in lower abundance protein detection by MS/MS spectral counting. We also evaluated the nature and consistency of capture of non-target proteins by the MARS-7 and MARS-14 columns. Our results document the high repeatability of target and non-target protein capture as well as an average 2-4-fold global enrichment of low abundance proteins by the MARS-7 and MARS-14 columns. Our analyses also demonstrate a nearly 10-fold greater number of detected proteins in a human colon tumor cell line than in depleted plasma at the same level of protein input. Further, we demonstrate unambiguously that the plasma proteome detection disparity is due to a steep protein abundance distribution in plasma versus cells. These findings demonstrate the consistent performance of abundant protein depletion with MARS columns, but also illustrate the limitations of shotgun proteomics to detect disease biomarkers in plasma proteomes.

MATERIALS AND METHODS

Plasma samples

Human blood plasma samples were obtained from three anonymous healthy volunteers through the Jim Ayers Institute for Precancer Detection and Diagnosis under an Institutional Review Board-approved protocol. Informed consent for research was obtained from all volunteers. Blood was drawn by venipuncture into 7 ml Vacutainer tubes using EDTA as anticoagulant (BD Worldwide, Franklin Lakes, NJ). Plasma was obtained by centrifugation of whole blood for 15 min at 380g. The initial plasma protein concentration was ~82 mg mL^{-1} as determined with the BCA protein assay (Pierce Chemicals, Rockford, IL). Unless otherwise noted, all protein sample processing was performed at 4°C.

MARS-7 and MARS-14 immunoaffinity chromatography

MARS-7 columns (4.6 x 100 mm) designed to deplete 7 abundant proteins (albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen) and MARS-14 columns (4.6 x 100 mm) designed to deplete 14 abundant proteins (albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen, alpha2-macroglobulin, alpha1-acid glycoprotein, IgM, apolipoprotein AI, apolipoprotein AII, complement C3, and transthyretin) were purchased from Agilent (Palo Alto, CA). Depletion was performed at room temperature with an Agilent 1100 series HPLC system. Plasma samples were diluted fourfold using the load/wash buffer supplied by the manufacture and remaining particulates in the diluted plasma were removed by centrifugation through a 0.22-μm spin filter 1 min at 16,000 × g.

After equilibration with the load/wash buffer, the MARS-7 column was loaded with 160 μL of the diluted plasma at a low flow rate (0.5 mL min^{-1}) for 10 min. Flow-through fractions,

representing depleted plasma, were collected and stored at -20 °C. The bound proteins were released with elution buffer at 1.0 mL min^{-1} for 7 min. The column was then washed with the load/wash buffer for 11 min at a flow rate of 1 mL min^{-1}. Each depletion cycle took 28 min of total run time. The MARS-14 column was equilibrated with the load/wash buffer and 160 μL of the diluted plasma was loaded at a low flow rate (0.125 mL min^{-1}) for 18 min and then for an additional 2 min at a flow rate of 1 mL min^{-1}. The other binding and elution steps were identical to those used for the MARS-7 column. Thus, each depletion run cycle of the MARS-14 column took 38 min of total run time. Both depleted (flow-through fraction) and abundant plasma proteins (bound fraction) were collected and stored at -20 °C until further analysis.

Protein Digestion

The depleted and bound protein fractions prepared from each sample were concentrated using Amicon 5 KDa molecular weight filters (Millipore, MA), followed by buffer exchange to 50 mM NH_4HCO_3 in the same unit. Protein concentration was then determined by the BCA protein assay (Pierce-Thermo Scientific, Rockford, IL). Proteins were denatured and reduced in 50 mM ammonium bicarbonate buffer, pH 8.0, 8 M urea, 10 mM dithiothreitol for 1 hour at 37 °C. Protein cysteines were alkylated with 50 mM iodoacetamide for 60 min at room temperature in the dark. The reduced and alkylated protein mixture was diluted 10-fold with 50 mM ammonium bicarbonate, pH 8.0, followed by the addition of trypsin (Promega, Cat#TB309, Madison, WI) at a trypsin/protein ratio of 1:50 (w/w). The digestion mixture was incubated overnight at 37 °C and then frozen at -80 °C and lyophilized. Samples were resuspended in 1 mL of deionized water and applied to SEP-Pak vac 1 cc (100 mg) C-18 cartridges (Waters Corp., Milford, MA), which were prewashed with 1 mL of acetonitrile and equilibrated with 2 mL of

deionized water. The flow-through was discarded, the cartridges were washed with 1 mL of deionized water, the bound peptides were eluted with 80% acetonitrile in deionized water, and the eluate was evaporated *in vacuo*.

IEF of peptides

Tryptic peptides from 200 μg of protein were redissolved in 500 μL of 6 M urea and loaded in an IPGphor rehydration tray (GE Healthcare, Piscataway, NJ). Carrier ampholyte (2% (v/v) was included in the loading solution (IPG buffer pH 3-10 NL (GE catalog #17-6000-88)). IPG strips (24 cm, pH 3-10 NL) (GE Healthcare) were placed over the samples and allowed to rehydrate overnight at room temperature. The loaded strips were focused at 20 °C on an Ettan IPGPhor-III IEF system (GE Healthcare) using the following focusing program: 500 V for 1 hr; a gradient to 1000 V for 3 hr; a gradient to 10000 V for 3 hr; a step to 10000 V for 5 hr. The strips were then cut into 20 pieces and placed in separate wells of a 96-well Falcon flat bottom polystyrene ELISA plate (Fisher Scientific). Peptides were eluted from the strips with 200 μL of 0.1% formic acid for 15 min, followed by 200 μL of acetonitrile/0.1% formic acid (1:1, v/v) for 15 min, then with 200 μL of acetonitrile containing 0.1% formic acid for 15 min. The combined eluates for each IPG strip fraction were evaporated *in vacuo* and then redissolved in 0.1% trifluoroacetic acid (TFA) and applied to a 96 well C-18 Oasis HLB plate (30 μm particle size, 10 mg packing) (Waters Corp., Milford, MA) prewashed with 1 mL of acetonitrile and equilibrated with 2 mL of 0.1% TFA. The flow-through was discarded and the cartridges were washed with 1 mL of 0.1% TFA. The bound peptides were eluted with 0.3 mL each of 30% acetonitrile/0.1% TFA, 70% acetonitrile/0.1% TFA and 100% acetonitrile/0.1% TFA and the

combined eluate was evaporated *in vacuo* and redissolved in 100 μL of 0.1% formic acid for LC-MS/MS analysis, as previously described [24].

Reverse-phase LC-MS/MS analysis

LC-MS/MS analyses were performed on a Thermo LTQ linear ion trap mass spectrometer equipped with a Thermo MicroAS autosampler and Thermo Surveyor HPLC pump, Nanospray source and Xcalibur 1.4 instrument control (Thermo Electron, San Jose, CA). Peptides were separated on a packed capillary tip (Polymicro Technologies, 100 μm X 11 cm) with Jupiter C18 resin (5 μm, 300 Å, Phenomenex) using an in-line solid-phase extraction column (100 μm X 6 cm) packed with the same C18 resin (using a frit generated with liquid silicate Kasil 1 [27]) similar to that previously described[28]. The flow from the HPLC pump was split prior to the injection valve to achieve flow-rates of 700 nL-1000 μL min-1 at the column tip. Mobile phase A consisted of 0.1 % formic acid and Mobile phase B consisted of 0.1% formic acid in acetonitrile. A 95 min gradient was performed with a 15 min washing period (100 % A for the first 10 min followed by a gradient to 98% A at 15 minutes) to allow for solid-phase extraction and removal of any residual salts. Following the washing period, the gradient was increased to 25% B by 50 min, followed by an increase to 90 % B by 65 min and held for 9 min before returning to the initial conditions. Liquid chromatography was carried out at ambient temperature at a flow rate of 0.6 μL min^{-1}. Centroided MS/MS scans were acquired on the LTQ using an isolation width of 2 *m*/*z*, an activation time of 30 ms, an activation *q* of 0.250 and 30% normalized collision energy using 1 microscan with a max ion time of 100 ms for each MS/MS scan and 1 microscan with a max ion time of 500 ms for each full MS scan. The mass spectrometer was tuned prior to analysis using the synthetic peptide TpepK (AVAGKAGAR), so

that some parameters may have varied slightly from experiment to experiment, but typically the tune parameters were as follows: spray voltage of 2 kV, a capillary temperature of 150 °C, a capillary voltage of 50 V and tube lens of 120 V. The MS/MS spectra were collected using data-dependent scanning in which one full MS spectrum was followed by four MS-MS spectra. MS/MS spectra were recorded using dynamic exclusion of previously analyzed precursors for 60 s with a repeat count of 1 and a repeat duration of 1.

Data processing and analysis

The LC-MS/MS raw data were converted into mzData file format by ScanSifter v2.0, an in-house developed software, and the MyriMatch algorithm (version 1.1.0)[29] was used to independently search all the MS/MS spectra against the human International Protein Index (IPI) database (version 3.37) with a total of 69,164 protein entries. The search parameters used were as follows: 1.25 Da tolerance for precursor ion masses and 0.5 Da for fragment ion masses. Candidate peptides were required to feature semitryptic cleavages, which allow one non-tryptic end, and any number of missed cleavages was permitted. Carbamidomethylation of cysteines was specified as a fixed modification, and variable modifications of methionine oxidation, N-terminal pyro-Glu from glutamine were allowed during the database search. The sequence database was doubled to contain each sequence in both forward and reversed orientations, enabling false discovery rate estimation.

The IDPicker algorithm[30] (version XXX) filtered the identifications for each LC-MS/MS run to include the largest set for which a 5% peptide identification FDR could be maintained. These identifications were pooled for each sample. The peptide filtering employed reversed sequence database match information to determine thresholds that yielded an estimated 5% FDR

for the identifications of each charge state by the formula FDR = (2 × reverse) / (forward + reverse)[31]. Proteins were required to have at least two distinct peptide sequences observed in the analyses. Indistinguishable proteins were recognized and grouped. Parsimony rules were applied to generate a minimal list of proteins that explained all of the peptides that passed our entry criteria[30].

Relative peptide/protein intensity analysis

Relative peptide intensities were calculated in the following way. First, MS/MS spectra were extracted to MGF files using ReAdWRaw4Mascot.exe (version 20091016a) and identified by searching the NIST library of peptide fragmentation mass spectra (Human IT Rel. 3, 02/04/09) with MSPepSearch (Version 0.9, NIST), both were downloaded from http://peptide.nist.gov. Mass tolerances were set to 0.7 *m/z* and 0.8 *m/z* for precursor and fragment masses, respectively. Additionally, MSPepSearch was set to pre-search all spectra in fast, peptide mode (fiPv), and all top-ranked matches were filtered to a 'Score' threshold of >450 (<1% FDR according to previous decoy searches, data not shown). Then, for each identified peptide ion, the maximum abundance was estimated from extracted ion chromatograms (also with ReAdW4Mascot.exe). Relative peptide intensity was then calculated as the fraction of all peptide intensity values for each peptide. To calculate relative protein intensity values, proteins were reassembled by first mapping the peptides to protein sequences. Only proteins with more than one distinct peptide sequence were counted and peptide intensities were assigned to only one protein. When several proteins contained a given sequence, assignment was made to the protein having the largest

number of assigned peptides. Peptide ion intensities for each protein then were summed and the relative abundance for each protein was calculated as the fraction of all protein abundances.

RESULTS

Sample recoveries and repeatability for MARS-7 and MARS-14 column separations

Three separate samples of human plasma (3.3 mg plasma protein each) were fractionated with MARS-7 and MARS-14 columns and recoveries were determined for both the flow-through and bound fractions (Table 1). An average of 0.40 mg (12%) and 2.09 mg (63%) of applied protein was recovered from the MARS-7 flow-through and bound fractions, respectively. From the MARS-14 an average of 0.25 mg (8%) and 2.23 mg (68%) of the applied protein was recovered from the flow-through and bound fractions, respectively (Table 1). CV values for recoveries of proteins from the flow-through and bound fractions across the three replicates were between 2 and 6%, indicating overall repeatability of the analysis. Ultraviolet detection (220 nm) traces for each run demonstrated consistent chromatography for both the MARS-7 column and MARS-14 column (Figure S1).

Analysis of MARS-7 and MARS-14 plasma protein fractions by LC-MS/MS

We evaluated the repeatability of MARS-7 and MARS-14 using the same plasma sample and reverse phase LC-MS/MS. Separations with the MARS-7 and MARS-14 were done three times and the corresponding flow-through and bound fractions were individually analyzed. A total of 43, 36 and 40 proteins were identified from the MARS-7 bound fractions and 107, 106 and 99 proteins were identified from flow-through fractions, respectively. Not included in these totals are many spectral matches to a large number of different immunoglobulin sequences; the extensive sequence overlap between these database entries made it impossible to discern the total number of proteins present based on peptide identifications. A total of 52, 52 and 50 proteins were identified from the MARS-14 bound fraction and 114, 118 and 113 proteins were identified

from flow-through fraction, respectively. As for the MARS-7 analyses, spectral matches to immunoglobulins were excluded.

The numbers of proteins identified in the triplicate analyses were largely consistent in both fractions from each column. We used spectral count data for quantitative comparisons, as this method provides robust estimates of differential protein concentrations in shotgun proteomic analyses [32, 33]. An illustration of the consistency of proteome inventories based on spectral counting is shown in Figure 1. Nonzero spectral count measurements are plotted for all combinations of the 3 replicate analyses of MARS14 flow-through fractions. Spearman Rank correlation coefficients were 0.976 (1 versus 2), 0.971 (1 versus 3) and 0.977 (2 versus 3), respectively, indicating that spectral counts for the large majority of proteins proved highly reproducible across replicate depletion experiments using the MARS system.

Efficiency of high abundance protein depletion by MARS-7 and MARS-14 columns

The premise of the MARS system is that enrichment of lower abundance proteins can be achieved through efficient extraction of high abundance proteins by the column. We measured efficiency of high abundance protein extraction on the basis of spectral counts from LC-MS/MS analyses of bound and flow-through fractions from three replicate MARS-7 or MARS-14 depletions. Although the bound and flow-through proteins are different in peptide composition, the spectral count data still allowed us to identify those proteins that are efficiently captured by the columns. The binding efficiency was calculated for each protein by dividing the spectral counts for each protein from the bound fraction (B) by the sum of spectral counts from the bound fraction and flow-through fraction (FT), *i.e.* B/(B+FT) [19]. In this study, albumin, IgG, antitrypsin, IgA, transferrin and haptoglobin were captured with >99% efficiency both by the MARS-7 and

MARS-14 column, as shown in Table 2. Fibrinogen was captured with 71% and 76% efficiency by MARS-7 and MARS-14, respectively. Not all of the targeted proteins were captured efficiently by the MARS-14 column. For example, apolipoprotein AI and AII were captured only with 47% and 31% efficiency by the MARS-14 column.

Binding of non-target plasma proteins to MARS-7 and MARS-14 columns

A potential problem with immunoaffinity protein depletion is that some non-target proteins might be removed along with the targeted proteins, either because they associate with target proteins or because they nonspecifically interact with the column. We refer to these as "non-target proteins". To evaluate the potential binding of non-target proteins to the MARS in more detail, we performed LC-MS/MS analyses of the bound and flow-through fractions from three replicate MARS-7 and MARS-14 column depletion experiments. A total of 20 non-target proteins were identified in LC-MS/MS analyses of the bound fractions from the MARS-7 and MARS-14 columns, respectively. Only proteins with two or more unique peptide identifications from all three experiments were considered and all immunoglobulin identifications were removed. To compare the extent of nonspecific binding in an individual high-abundance protein separation, a "non-target capture" value was calculated for each detected non-target protein by dividing the spectral counts for each protein from the bound fraction (B) by the sum of spectral counts for corresponding protein from the bound fraction and from the flow-through fraction (FT), *i.e.* B/(B + FT). A larger value indicates that a high proportion of the protein is observed in the bound fraction. The capture values for each non-target protein with at least 5 spectral counts observed in the triplicate LC-MS/MS analyses of the bound fractions from MARS-7 and MARS-14 are listed in Figure 2. (Spectral counts for all non-target proteins identified from the bound

fraction and from the flow-through fraction are provided in Table S1.) Several of the higher abundance non-target proteins bound to the MARS-7 column are target proteins for the MARS-14 column (α-2 macroglobulin, apolipoprotein AI, α1-acid glycoprotein 1, α1-acid glycoprotein 2, and transthyretin). Others were consistently detected only as non-target proteins on MARS-14 columns (pregnancy zone protein, apolipoprotein B-100, apoliproprotein CII, apolipoprotein CIII). Of the proteins we detected as having significant non-target binding to the MARS columns, CD5 antigen-like protein, zinc-α_2–glycoprotein, hemoglobin, apolipoprotein C-III, apolipoprotein L1 and serum amyloid A-4 protein also were reported as non-target proteins for the IgY12 column[19].

Impact of high-abundance protein depletion on detection of lower abundance proteins

We used multidimensional LC-MS/MS to evaluate the extent to which lower abundance proteins could be detected in depleted versus undepleted plasma. The same quantities of peptides from unfractionated plasma, the MARS-7 FT fraction and the MARS-14 FT fraction were resolved by IEF into 20 fractions prior to LC-MS/MS. Only those proteins identified with at least two unique peptides from the entire study set were considered confident identifications. (The full IDPicker report, which contains an .html summary of all identified peptides and proteins is supplied as Supplementary Information.)

Additionally, we filtered out all proteins that had less than 10 spectra identified in the combined dataset (N=417) and collapsed an additional 19 proteins with more than one IPI identifier from protein groups that represented isoforms of the same protein. The remaining dataset contained 238 IPI entries with matching HUGO identifiers, 113 IPI entries without HUGO identifiers (mostly IgGs), 4 contaminant entries and 8 proteins identified by reverse

peptide matches. Thus, protein FDR for this filtered dataset is 4.6% (2 x 8/351). Removal of all 131 IPI entries clustering in IgG clusters (#2, 20, 155 and 242 in the IDPicker report) reduced the final dataset to 220 non-overlapping protein groups, of which 213 had matching HUGO identifiers. Figure 3 lists a VENN diagram of the identifications in crude plasma, MARS-7 FT and MARS-14 FT for this set of 220 proteins.

Spectral count data provide a robust means to estimate protein abundances from shotgun proteomic datasets [32, 33]. Figure 4 illustrates the rank correlation ($r^2 = 0.921$) between the spectral counts for 100 proteins detected in crude plasma and with previously reported protein concentrations in plasma. We compared the spectral counts for proteins detected in the MARS FT fractions to the counts for the same proteins in crude plasma. Only the 100 proteins observed with at least 10 spectral counts in the entire study were used and these were ranked by spectral counts in the crude plasma. (The immunoglobulin identifications were excluded). To assess the level of enrichment for lower abundant proteins, the $\log_2$ of the ratio of counts in the MARS FT fraction to crude (unfractionated) plasma was used to estimate the increase (or decrease) in detection as a function of MARS fractionation (Figure 5). For proteins detected only in MARS FT fractions, the denominator of the ratio is zero; thus, for these proteins a single count was used. A positive $\log_2$ ratio means that the spectral counts for the protein are higher in MARS flow-through than in crude plasma.

Analysis of the MARS-7 FT indicated that 205 proteins (95% of the total detected) increased in spectral counts by at least 2-fold compared to unfractionated plasma (Figure 5). Similarly, in the MARS-14 FT, 193 proteins (94% of the total detected) increased the spectral counts by at least 2-fold. For both the MARS-7 and MARS-14 columns, enrichment relative to undepleted plasma ranged from 2-4 fold. Figure 5 also illustrates the depletion of some proteins

by the MARS-7 and MARS-14 columns; these include the target proteins and some non-target proteins described above and listed in Table 2 and Figure 2.

Among the proteins with enhanced detection in the MARS-7 and MARS-14 FT fractions were 23 proteins present at plasma concentrations of 10 ng mL^{-1} or lower Table S4. Concentrations for these proteins were obtained from previously reported data for plasma or serum levels in normal human subjects [1]. These very low abundance proteins accounted for only 5.6% and 6.0% of the proteins detected in the MARS-7 and MARS-14 FT fractions, respectively.

Comparison of protein identifications between crude plasma, depleted plasma and RKO cells

Analysis of undepleted plasma and the MARS-7 and MARS-14 depleted plasma by IEF-LC-MS/MS collectively yielded 659 protein identifications (with at least 2 identified peptides). This total was significantly lower than we had identified in analyses of human colon adenomas and tumor cells with a similar IEF-LC-MS/MS platform [24, 25]. To compare these proteomes in the same analysis, we analyzed 200 μg of protein each from RKO colon tumor cells, from unfractionated plasma and from the MARS FT fractions. The analyses were done in three replicates each with the IEF-LC-MS/MS platform described above. We confidently identified 3411 proteins in RKO cells, 371 proteins from unfractionated plasma, 342 proteins in MARS-7 depleted plasma and 288 proteins in MARS-14 depleted plasma. Our analyses of IEF fractions from undepleted plasma acquired 287,223 MS/MS spectra, which is nearly identical to the total acquired (289,010 spectra) in analysis of RKO cells.

We hypothesized that protein identifications in plasma may be diminished by significant numbers of peptides (e.g., variant immunoglobulin sequences) that are not represented in the

search database and thus cannot be identified by database searching. If so, we reasoned that addition of plasma to the RKO cell lysate might decrease the number of spectral counts for RKO protein without producing a concomitant increase in plasma protein spectral counts.

We performed a plasma spike experiment, in which the total amount of plasma and RKO cell protein mixture was fixed. We spiked 0, 1%, 2%, 5%, 10%, and 20% plasma protein into RKO cell lysates to create 7 mixed samples; we also prepared a 100% plasma sample (Figure 6). All were analyzed by reverse phase LC-MS/MS (no IEF). "Confident IDs" refers to MS/MS spectra mapped to peptides according to the criteria described in Experimental Procedures (see above). Proteins found only in unspiked RKO cell lysates were assumed to be derived only from RKO cells, whereas other proteins were assumed to be derived from plasma. Identifications either in RKO lysates or in plasma comprised "total confident IDs" and "total proteins" in Figure 6. All of the analyses yielded essentially identical numbers of MS/MS spectra. Numbers of confidently identified RKO peptides did not decrease with increasing plasma spike concentration, nor did total numbers of identified peptides. Similar results were obtained for identifications at the protein level. Nevertheless, the sample containing 100% plasma yielded many fewer protein identifications than the RKO lysate and the RKO/plasma mixtures. These data suggest that lower numbers of plasma protein identifications reflects the presence of unidentifiable peptide sequences not represented in the database.

Impact of protein abundance distribution on proteome analyses

A characteristic feature of the plasma proteome is the high concentration of a few dozen blood proteins, which comprise approximately 99% of total plasma protein [1]. This fact provides the driving rationale for immunodepletion strategies, but it is not clear whether it also explains

the disparity in proteome inventories from plasma and cells. Cells also display a broad range of protein abundances [34, 35], which could similarly affect detection of lower abundance proteins. We reasoned that any bias due to oversampling of highly abundant peptides would be evident in spectral count data for the detected peptides and proteins. Plots of protein order as a function of spectral counts for RKO cells, undepleted plasma, MARS-7 FT, and MARS-14 FT proteomes are shown in Figure 7. In each plot, the spectral count curve for detected peptides and proteins is considerably steeper for plasma and the MARS FT fractions than for RKO cells. For example, the (blue) curve for plasma in Figure 7A shows that the top 10 proteins account for 70% of the spectral counts, whereas the curve for RKO cells shows that the top 10 proteins account for less than 205 of the total. These relationships are consistently observed in plots where spectral counts for multiple peptides are combined to generate protein spectral counts (Figures 7A and 7B) or when spectral counts for individual peptides are plotted (Figures 7C and 7D). These relationships held for both reverse phase LC-MS/MS analyses (Figures 7A and 7C) and for IEF-LC-MS/MS analyses (Figures 7B and 7D). In analyses of plasma, the MS instrument is presented with a selection of peptide ions that is most heavily biased in favor of the highest abundance proteins. This bias is lessened somewhat in the MARS FT fractions, but the curves for the MARS-7 and MARS-14 FT fractions are still considerably steeper than for RKO cells.

If bias in peptide ion sampling drives the plasma/cell disparity in protein identifications, then spectral counting provides only a surrogate measure of this effect. Direct measurement of peptide ion MS1 signal intensities would most directly represent the basis for dynamic sampling by the MS instrument. We therefore extracted MS1 signal intensities for peptides from the LC-MS/MS datasets for RKO cells, undepleted plasma, the MARS-7 FT and MARS-14 FT fractions. We further calculated protein MS1 intensities from the combined peptide MS1 intensity data

from reverse phase LC-MS/MS analyses (no IEF). The distributions of the peptide and protein intensities for each sample type are shown in Figure 8. Relative abundance values were plotted according to intensity rank (from highest to lowest). Figure 8A and 8B show large differences in the fraction of total abundance accounted for by the same number of proteins and peptides. For example, in Figure 8A (reverse phase LC-MS/MS), the top protein and top 20 proteins account for 56% and 89% of the total MS1 signal intensity in crude plasma, whereas in RKO cells, the top protein and top 20 proteins account for only 3.7% and 30% of the total intensity. These plots also illustrate the relative effectiveness of the MARS depletion columns. Depletion of the top 14 proteins (with MARS-14) reduced the fractional intensity at protein rank 20 by about 19%, presumably allowing for the detection of additional lower abundance proteins. The protein and peptide signal distributions are similar for IEF-LC-MS/MS analyses (Figure 8C and 8D), with unfractionated plasma displaying the steepest signal intensity curve and RKO cells displaying the flattest curve. These data quantitatively support the hypothesis that the disparity in proteome inventories between plasma and cells is due to the steep protein abundance relationship for plasma.

DISCUSSION

Depletion of abundant plasma proteins with immunoaffinity columns was originally applied in the context 2D gel analyses to observe changes in the lower abundance proteome that were literally obscured by high abundance proteins [8]. Subsequent work confirmed the ability of abundant protein depletion to facilitate detection of lower abundance plasma and serum proteins [2]. However, much of this work was done with older proteomic technologies (i.e., 2D gels) and did not involve quantitative assessment of either protein depletion or the enhanced detection of lower abundance proteins. We undertook our study to quantitatively assess the performance of the MARS system with a current generation LC-MS/MS analysis platform. Our results demonstrate that the MARS-7 and MARS-14 systems provide largely specific and highly repeatable depletion of their target proteins in human plasma and that they broadly enhance detection of plasma proteins by 2-4 fold. Our studies also highlighted the disparity in protein identifications between plasma and cell proteomes. This disparity is due to the uniquely steep concentration curve for plasma proteins. The results further imply that shotgun proteomic analysis of plasma is poorly suited to discover low abundance proteins as candidate disease biomarkers, even in samples depleted of the most abundant proteins.

The MARS-7 and MARS-14 columns displayed consistent performance in depleting target proteins, although not all targets were captured with equal efficiency. Albumin, IgG, antitrypsin, IgA, transferrin and haptoglobin were captured with >99% efficiency both by the MARS-7 and MARS-14 column and as previously reported for the MARS-6 column [15]. Fibrinogen was captured only with 71% and 76% efficiency by MARS-7 and MARS-14, respectively. Apolipoprotein AI and apolipoprotein AII were captured only with 47% and 31% efficiency by the MARS-14 column (these proteins were not targeted by the MARS-7 column).

In addition to consistently depleting target proteins and causing minimal depletion of non-target proteins, the MARS columns broadly enhanced detection of plasma proteins, typically by 2-4-fold, based on spectral counts. In our IEF-LC-MS/MS analyses MARS immunodepletion typically increased detection of plasma proteins by 30-40% relative to non-depleted plasma. Although the MARS-14 column removed 7 additional proteins and a greater fraction of the plasma proteome mass, this conferred no real advantage over the MARS-7 column in enhancing global plasma protein identifications—indeed, the MARS-14 FT fraction yielded slightly fewer identifications (208 vs. 216) than the MARS-7 FT. Our data and that of Qian et al. [20] demonstrate increased detection of several plasma proteins with reported concentrations <10 ng mL^{-1}. Nevertheless, it appears that these increased detections represent exceptions to the rule that proteins at concentrations below μg mL^{-1} are sporadically detected in shotgun proteome analyses, even in immunodepleted plasma [21, 36].

A frequently expressed concern regarding immunodepletion strategies is the removal of non-target proteins through non-specific association with the depletion columns or with captured proteins. This has been described as the "albumin sponge effect"[37, 38]. We observed that the bound fraction associated with MARS-7 and MARS-14 columns contained 30 non-target proteins, most of which were captured at low levels based on spectral counts (Table S3). This finding is similar to the previous report of 38 non-target proteins captured by IgY-12 columns [19]. For those non-target proteins captured at higher levels, binding appeared to be reproducible (Figure 2). We note that specific interaction with target proteins may explain at least a few of the non-target captures. For example, hemoglobin is reported to bind to haptoglobin and zinc-α_2-glycoprotein may interact with the targeted apolipoproteins and interact directly with the Fc region of IgG [39, 40]. Pregnancy zone protein, a relatively abundant pregnancy-associated plasma

protein, has strong similarity to alpha-2-macroglobulin [41], which perhaps explained why pregnancy zone protein was only captured by the MARS-14 column (which captures alpha-2-macroglobulin), but not by the MARS-7 column. Our observations regarding non-target protein capture are similar to those reported by Qian et al. [20], who performed similar evaluation of the IgY14 and SuperMix columns. Taken together, these observations with two widely used products suggest that inadvertent capture of non-target proteins is unlikely to interfere with the performance of analysis platforms that incorporate an abundant plasma protein depletion step.

Beyond the characterization of the MARS columns for enhancement of plasma proteome analysis, our second major goal was to address reasons for the dramatic disparity in protein identifications between plasma and cell proteomes. Given identical protein input and analysis on a standardized IEF-LC-MS/MS platform, the difference in identifications between plasma and cell lysates was striking—nearly 10-fold. We recognized that the presence of a relatively small number of highly abundant proteins has long been recognized as a barrier to plasma proteome analysis, but we considered other possible contributing factors. First, IEF-LC-MS/MS analyses of plasma, MARS-depleted plasma and RKO cells all yielded comparable numbers of MS/MS spectra. Our mixing experiments with plasma and RKO cell lysates tested the hypothesis that some components of plasma proteomes (e.g., uncharacterized immunoglobulin sequences) suppressed protein detection by generating MS/MS spectra that could not be matched to database sequences. These experiments showed that the decrement in detected RKO proteins resulting from plasma spikes was no more than would be expected based on simple competition from identifiable plasma proteins. These experiments don't eliminate the possibility that unidentified protein and peptide sequences are present in plasma, but do suggest that they do not significantly suppress protein detection or identification.

Our final experiments compared the quantitative hierarchy of plasma, depleted plasma and cell lysate proteomes based on spectral counts and MS1 signal intensities. Both analyses clearly demonstrated that plasma and immunodepleted plasma fractions display a much steeper concentration curve for the most abundant proteins. This steep abundance curve appears to be the major reason for the low numbers of plasma protein identifications. We note that in our analyses of the 100% plasma sample shown in Figure 6, numbers of confident peptide identifications were similar to those for RKO cells, but that plasma yielded fewer protein identifications. This is consistent with oversampling of the more abundant plasma peptides, which yields relatively few protein identifications.

The oversampling of relatively abundant plasma peptides provides an explanation for the poor representation of low abundance proteins (< 10 ng mL^{-1}), even in immunodepleted samples. Our analyses of the MARS-7 and MARS-14 FT fractions yielded only 23 proteins [Rob will check to make sure this is still 23 after we tightened up the protein FDR] in this concentration range—approximately 5-6% of the total identifications. These findings confirm published results with IgY14 and SuperMix immunodepletion columns [20] and indicate that, even with immunodepletion, detection of low abundance proteins is sporadic at best. Untargeted proteomic analyses using current LC-MS/MS platforms thus cannot be expected to discover low abundance, disease-specific biomarkers in plasma.

ACKNOWLEDGEMENTS

This work was supported by a cooperative agreement award 5U24CA126479 from the National Cancer Institute through the Clinical Proteomic Technology Assessment for Cancer (CPTAC) program and by an interagency agreement between the National Cancer Institute and the National Institute of Standards and Technology.

REFERENCES

1. Anderson, N. L.; Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **2002,** 1, (11), 845-67.

2. Pernemalm, M.; Lewensohn, R.; Lehtio, J., Affinity prefractionation for MS-based plasma proteomics. *Proteomics* **2009,** 9, (6), 1420-7.

3. Travis, J.; Bowen, J.; Tewksbury, D.; Johnson, D.; Pannell, R., Isolation of albumin from whole human plasma and fractionation of albumin-depleted plasma. *Biochem J* **1976,** 157, (2), 301-6.

4. Ahmed, N.; Barker, G.; Oliva, K.; Garfin, D.; Talmadge, K.; Georgiou, H.; Quinn, M.; Rice, G., An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum. *Proteomics* **2003,** 3, (10), 1980-7.

5. Gianazza, E.; Arnaud, P., A general method for fractionation of plasma proteins. Dye-ligand affinity chromatography on immobilized Cibacron blue F3-GA. *Biochem J* **1982,** 201, (1), 129-36.

6. Govorukhina, N. I.; Keizer-Gunnink, A.; van der Zee, A. G.; de Jong, S.; de Bruijn, H. W.; Bischoff, R., Sample preparation of human serum for the analysis of tumor markers. Comparison of different approaches for albumin and gamma-globulin depletion. *J Chromatogr A* **2003,** 1009, (1-2), 171-8.

7. Wang, Y. Y.; Cheng, P.; Chan, D. W., A simple affinity spin tube filter method for removing high-abundant common proteins or enriching low-abundant biomarkers for serum proteomic analysis. *Proteomics* **2003,** 3, (3), 243-8.

8. Pieper, R.; Su, Q.; Gatlin, C. L.; Huang, S. T.; Anderson, N. L.; Steiner, S., Multi-component immunoaffinity subtraction chromatography: An innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics.* **2003,** 3, (4), 422-432.

9. Bjorhall, K.; Miliotis, T.; Davidsson, P., Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* **2005,** 5, (1), 307-17.

10. Desrosiers, R. R.; Beaulieu, E.; Buchanan, M.; Beliveau, R., Proteomic analysis of human plasma proteins by two-dimensional gel electrophoresis and by antibody arrays following depletion of high-abundance proteins. *Cell Biochem Biophys* **2007,** 49, (3), 182-95.

11. Echan, L. A.; Tang, H. Y.; Ali-Khan, N.; Lee, K.; Speicher, D. W., Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* **2005,** 5, (13), 3292-303.

12. Roche, S.; Tiers, L.; Provansal, M.; Seveno, M.; Piva, M. T.; Jouin, P.; Lehmann, S., Depletion of one, six, twelve or twenty major blood proteins before proteomic analysis: the more the better? *J Proteomics* **2009,** 72, (6), 945-51.

13. Whiteaker, J. R.; Zhang, H.; Eng, J. K.; Fang, R.; Piening, B. D.; Feng, L. C.; Lorentzen, T. D.; Schoenherr, R. M.; Keane, J. F.; Holzman, T.; Fitzgibbon, M.; Lin, C.; Zhang, H.; Cooke, K.; Liu, T.; Camp, D. G., 2nd; Anderson, L.; Watts, J.; Smith, R. D.; McIntosh, M. W.; Paulovich, A. G., Head-to-head comparison of serum fractionation techniques. *J Proteome Res* **2007,** 6, (2), 828-36.

14. Gong, Y.; Li, X.; Yang, B.; Ying, W.; Li, D.; Zhang, Y.; Dai, S.; Cai, Y.; Wang, J.; He, F.; Qian, X., Different immunoaffinity fractionation strategies to characterize the human plasma proteome. *J Proteome Res* **2006,** 5, (6), 1379-87.

15. Brand, J.; Haslberger, T.; Zolg, W.; Pestlin, G.; Palme, S., Depletion efficiency and recovery of trace markers from a multiparameter immunodepletion column. *Proteomics* **2006,** 6, (11), 3236-42.

16. Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H.; Apweiler, R.; Haab, B. B.; Simpson, R. J.; Eddes, J. S.; Kapp, E. A.; Moritz, R. L.; Chan, D. W.; Rai, A. J.; Admon, A.; Aebersold, R.; Eng, J.; Hancock, W. S.; Hefta, S. A.; Meyer, H.; Paik, Y. K.; Yoo, J. S.; Ping, P.; Pounds, J.; Adkins, J.; Qian, X.; Wang, R.; Wasinger, V.; Wu, C. Y.; Zhao, X.; Zeng, R.; Archakov, A.; Tsugita, A.; Beer, I.; Pandey, A.; Pisano, M.; Andrews, P.; Tammen, H.; Speicher, D. W.; Hanash, S. M., 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* **2005,** 5, (13), 3226-45.

17. Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A., Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* **2007,** 6, (12), 2212-29.

18. Huang, L.; Harvie, G.; Feitelson, J. S.; Gramatikoff, K.; Herold, D. A.; Allen, D. L.; Amunngama, R.; Hagler, R. A.; Pisano, M. R.; Zhang, W. W.; Fang, X., Immunoaffinity separation of plasma proteins by IgY microbeads: meeting the needs of proteomic sample preparation and analysis. *Proteomics* **2005,** 5, (13), 3314-28.

19. Liu, T.; Qian, W. J.; Mottaz, H. M.; Gritsenko, M. A.; Norbeck, A. D.; Moore, R. J.; Purvine, S. O.; Camp, D. G., 2nd; Smith, R. D., Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol Cell Proteomics* **2006,** 5, (11), 2167-74.

20. Qian, W. J.; Kaleta, D. T.; Petritis, B. O.; Jiang, H.; Liu, T.; Zhang, X.; Mottaz, H. M.; Varnum, S. M.; Camp, D. G., 2nd; Huang, L.; Fang, X.; Zhang, W. W.; Smith, R. D., Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. *Mol Cell Proteomics* **2008,** 7, (10), 1963-73.

21. States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M., Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* **2006,** 24, (3), 333-8.

22. Wang, H.; Qian, W. J.; Chin, M. H.; Petyuk, V. A.; Barry, R. C.; Liu, T.; Gritsenko, M. A.; Mottaz, H. M.; Moore, R. J.; Camp Ii, D. G.; Khan, A. H.; Smith, D. J.; Smith, R. D., Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. *J Proteome Res* **2006,** 5, (2), 361-9.

23. Whiteaker, J. R.; Zhang, H.; Zhao, L.; Wang, P.; Kelly-Spratt, K. S.; Ivey, R. G.; Piening, B. D.; Feng, L. C.; Kasarda, E.; Gurley, K. E.; Eng, J. K.; Chodosh, L. A.; Kemp, C. J.; McIntosh, M. W.; Paulovich, A. G., Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *J Proteome Res* **2007,** 6, (10), 3962-75.

24. Slebos, R. J.; Brock, J. W.; Winters, N. F.; Stuart, S. R.; Martinez, M. A.; Li, M.; Chambers, M. C.; Zimmerman, L. J.; Ham, A. J.; Tabb, D. L.; Liebler, D. C., Evaluation of strong cation exchange versus isoelectric focusing of peptides for multidimensional liquid chromatography-tandem mass spectrometry. *J Proteome Res* **2008,** 7, (12), 5286-94.

25. Sprung, R. W., Jr.; Brock, J. W.; Tanksley, J. P.; Li, M.; Washington, M. K.; Slebos, R. J.; Liebler, D. C., Equivalence of protein inventories obtained from formalin-fixed paraffin-

embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol Cell Proteomics* **2009,** 8, (8), 1988-98.

26. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009**.

27. Cortes, H. J.; Pfeiffer, C. D.; Richter, B. E.; Stevens, T. S., Porous Ceramic Bed Supports for Fused-Silica Packed Capillary Columns Used in Liquid-Chromatography. *Journal of High Resolution Chromatography & Chromatography Communications* **1987,** 10, (8), 446-448.

28. Licklider, L. J.; Thoreen, C. C.; Peng, J.; Gygi, S. P., Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Analytical Chemistry* **2002,** 74, (13), 3076-3083.

29. Tabb, D. L.; Fernando, C. G.; Chambers, M. C., MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res* **2007,** 6, (2), 654-61.

30. Zhang, B.; Chambers, M. C.; Tabb, D. L., Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J Proteome Res* **2007,** 6, (9), 3549-57.

31. Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P., Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* **2005,** 2, (9), 667-75.

32. Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **2004,** 76, (14), 4193-201.

33. Zybailov, B.; Coleman, M. K.; Florens, L.; Washburn, M. P., Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal Chem* **2005,** 77, (19), 6218-24.

34. Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S., Global analysis of protein expression in yeast. *Nature* **2003,** 425, (6959), 737-41.

35. Malmstrom, J.; Beck, M.; Schmidt, A.; Lange, V.; Deutsch, E. W.; Aebersold, R., Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **2009,** 460, (7256), 762-5.

36. Rifai, N.; Gillette, M. A.; Carr, S. A., Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* **2006,** 24, (8), 971-83.

37. Zhou, M.; Lucas, D. A.; Chan, K. C.; Issaq, H. J.; Petricoin, E. F., 3rd; Liotta, L. A.; Veenstra, T. D.; Conrads, T. P., An investigation into the human serum "interactome". *Electrophoresis* **2004,** 25, (9), 1289-98.

38. Zolotarjova, N.; Martosella, J.; Nicol, G.; Bailey, J.; Boyes, B. E.; Barrett, W. C., Differences among techniques for high-abundant protein depletion. *Proteomics* **2005,** 5, (13), 3304-13.

39. Robert, L.; Bajic, V.; Jayle, M. F., [The effect of certain glucides and polyelectrolytes on the combination of hemoglobin with serum haptoglobin and on its catalytic activity.]. *C R Hebd Seances Acad Sci* **1956,** 242, (24), 2868-70.

40. Kennedy, M. W.; Heikema, A. P.; Cooper, A.; Bjorkman, P. J.; Sanchez, L. M., Hydrophobic ligand binding by Zn-alpha 2-glycoprotein, a soluble fat-depleting factor related to major histocompatibility complex proteins. *J Biol Chem* **2001,** 276, (37), 35008-13.

41. Philip, A.; Bostedt, L.; Stigbrand, T.; O'Connor-McCourt, M. D., Binding of transforming growth factor-beta (TGF-beta) to pregnancy zone protein (PZP). Comparison to the TGF-beta-alpha 2-macroglobulin interaction. *Eur J Biochem* **1994,** 221, (2), 687-93.

Table 1. Sample yields and recoveries for MARS-7 and MARS-14 column separations.

Method	Load volume (μL)	Load (mg)	yield (μg)	Recovery %	CV %
MARS-7	40	3.3	2487± 137	75 ± 4	6
Flowthrough			401 ± 6	12 ± 0	2
Eluate			2086 ± 132	63 ± 4	6
MARS-14	40	3.3	2481± 83	75 ± 3	3
Flowthrough			253 ± 5	8 ± 0	2
Eluate			2228 ± 81	68 ± 3	4

Table 2. Capture efficiency for MARS-7 and MARS-14 target proteins. Listed proteins 1-7 are targeted by the MARS-7 column; listed proteins 1-14 are targeted by the MARS-14 column.

			MARS-7		MARS 14		Capture efficiency (%)[c]	
	IPI no.	Protein name	B[a]	FT[b]	B	FT	MARS-7	MARS-14
#1	IPI00553177	Alpha-1-antitrypsin	115	0	130	0	>99	>99
#2	IPI00641737	Haptoglobin	148	0	159	0	>99	>99
#3	IPI00744561	IGHA1 protein	69	0	56	0	>99	>99
#4	IPI00448925	IGHG1 protein	154	0	132	0	>99	>99
#5	IPI00021885	Isoform 1 of fibrinogen alpha chain	114	46	124	39	71	76
#6	IPI00745872	Isoform 1 of serum albumin	699	0	849	3	>99	>99
#7	IPI00022463	Serotransferrin	168	0	173	0	>99	>99
#8	IPI00022429	Alpha-1-acid glycoprotein 1	21	48	48	8		86
#9	IPI00478003	Alpha-2-macroglobulin	2	401	225	47		83
#10	IPI00021841	Apolipoprotein A-I	25	154	83	92		47
#11	IPI00021854	Apolipoprotein A-II	0	70	20	44		31

#12	IPI00783987	Complement C3	0	465	252	62	80
#13	IPI00472610	IGHM protein	165	0	147	0	>99
#14	IPI00022432	Transthyretin	3	5	29	0	>99

[a]B, bound fraction; listed values are MS/MS spectral counts for proteins detected

[b]FT, flow-through fraction listed values are MS/MS spectral counts for proteins detected

[c]Capture efficiency was calculated as B/B+FT; see text for discussion.

FIGURE LEGENDS

Figure 1. Linear correlations of the spectral counts between pairs LC-MS/MS analyses of three replicate MARS-14 FT fractions. Spearman correlation coefficient (*r*) is shown for each comparison. Similar results were obtained with MARS-7 depleted flow-through samples.

Figure 2. Non-target human plasma proteins that bind to the MARS-7 and MARS-14 column in replicate LC-MS/MS analyses. A non-target capture value was calculated for each protein by dividing the spectral counts for each protein from the bound fraction (B) by the sum of spectral counts for corresponding protein from the bound fraction and from the flow-through fraction (FT), *i.e.* B/(B + FT).

Figure 3. Comparison of proteome inventories from crude plasma, MARS-7 flow-through and MARS-14 flow-through analyzed by IEF-LC-MS/MS.

Figure 4. Linear correlation between the spectral counts for each protein and previously reported concentrations of plasma proteins[1].

Figure 5. Global impact of MARS-7 (A) and MARS-14 (B) immunodepletion on detection of plasma proteins. Proteins observed with at least 10 spectral counts in the study are represented and are ranked by spectral counts in unfractionated (crude) plasma. Plotted values are $\log_2$ ratios of detected protein spectral counts in depleted vs. unfractionated plasma.

Figure 6. Shotgun proteome analyses of RKO cell lysates spiked with plasma. The total amount of protein in each sample was held constant and plasma was spiked into RKO cell lysates at the indicated percentages. Samples were analyzed by reverse phase LC-MS/MS (no IEF). The lower panel represents total MS/MS spectra. The middle panel represents confident peptide identifications (number of MS/MS spectra matching RKO or plasma peptide sequences). The top panel represents protein identifications. Two complete experiments yielded similar results; the results of one experiment are shown.

Figure 7. The distribution of spectral counts for identified proteins (A and B) and peptides (C and D) in unfractionated plasma (blue line), MARS-7 FT (red line), MARS-14 FT (green line) and RKO cell lysate (purple line). Proteins and peptides are ranked on the x-axis in order of decreasing spectral counts in unfractionated plasma. Data are plotted for the top 50 proteins and the top 100 peptides. Results shown are for a single analysis by either reverse phase LC-MS/MS (A and C) or IEF-LC-MS/MS (B and D).

Figure 7. The distribution of MS1 signal intensities for identified proteins (A and B) and peptides (C and D) in unfractionated plasma (blue line), MARS-7 FT (red line), MARS-14 FT (green line) and RKO cell lysate (purple line). Proteins and peptides are ranked on the x-axis in order of decreasing signal intensity in unfractionated plasma. Data are plotted for the top 50 proteins and the top 100 peptides. Results shown are for a single analysis by either reverse phase LC-MS/MS (A and C) or IEF-LC-MS/MS (B and D).

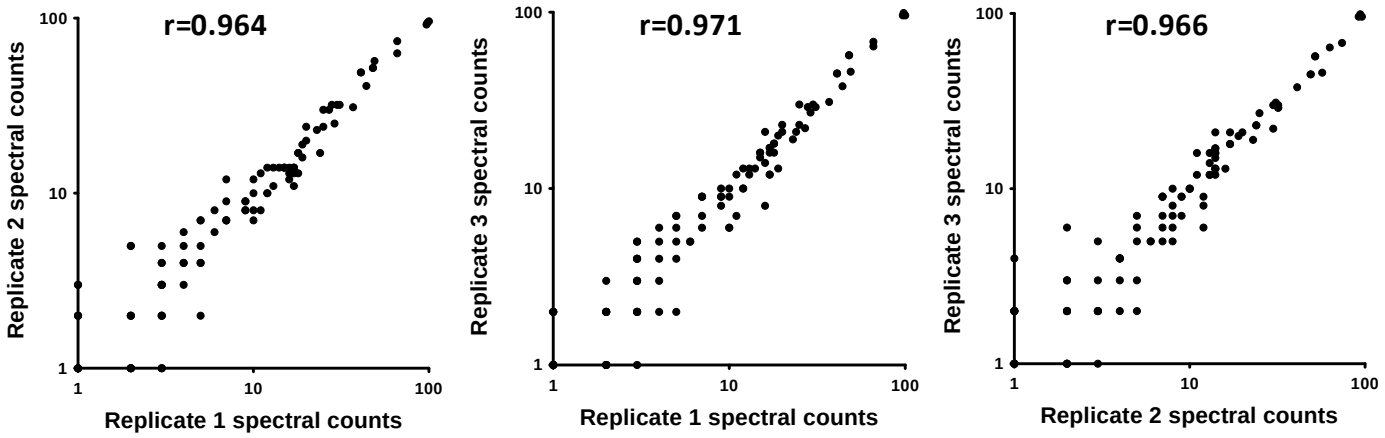

r=0.964
Replicate 2 spectral counts
Replicate 1 spectral counts
100
10
1
1
10
100
r=0.971
Replicate 3 spectral counts
Replicate 1 spectral counts
r=0.966
Replicate 3 spectral counts
Replicate 2 spectral counts

Figure 1

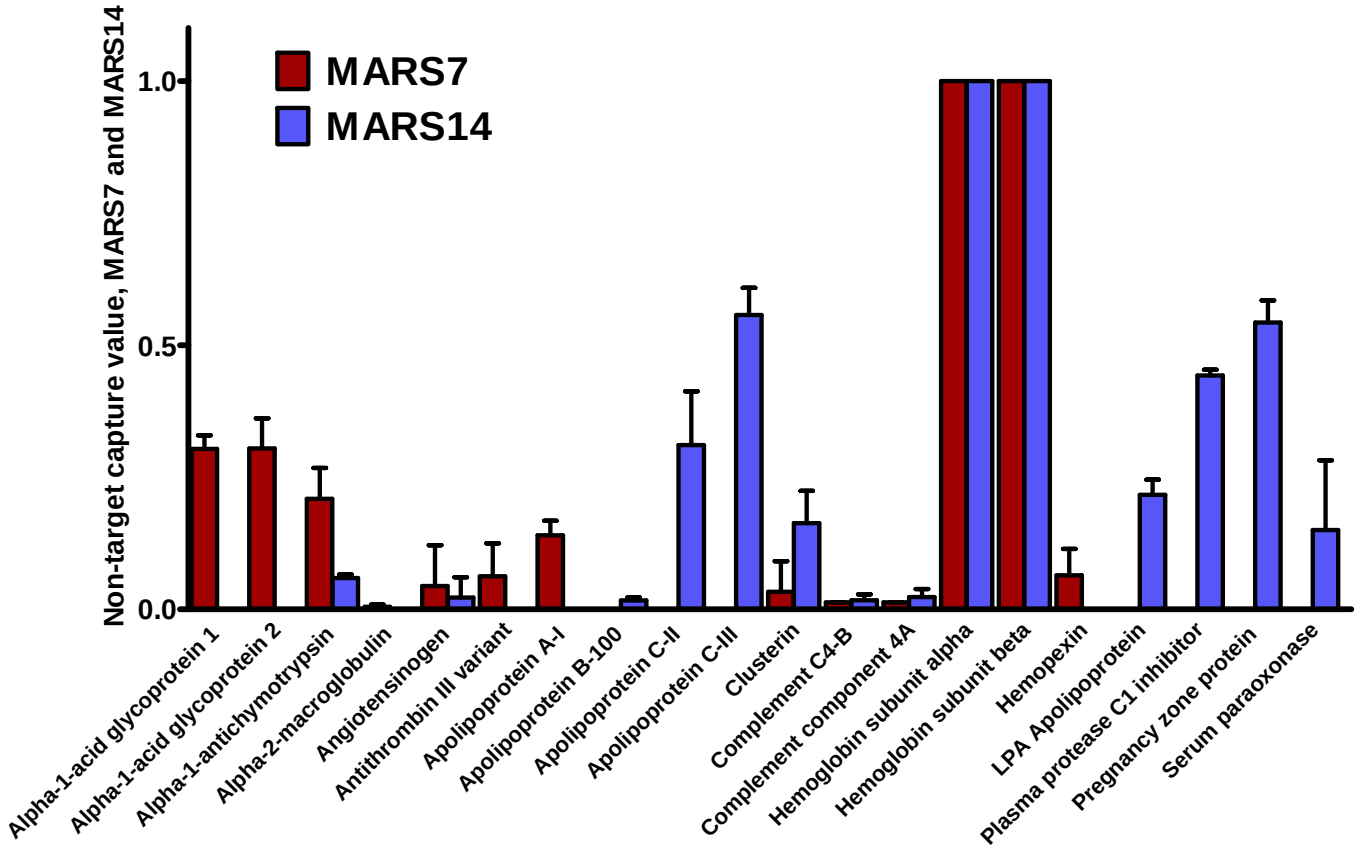

MARS7
MARS14
Non-target capture value, MARS7 and MARS14
1.0
0.5
0.0
Alpha-1-acid glycoprotein 1
Alpha-1-acid glycoprotein 2
Alpha-1-antichymotrypsin
Alpha-2-macroglobulin
Angiotensinogen
Antithrombin III variant
Apolipoprotein A-I
Apolipoprotein B-100
Apolipoprotein C-II
Apolipoprotein C-III
Clusterin
Complement C4-B
Complement component 4A
Hemoglobin subunit alpha
Hemoglobin subunit beta
Hemopexin
LPA Apolipoprotein
Plasma protease C1 inhibitor
Pregnancy zone protein
Serum paraoxonase

Figure 2

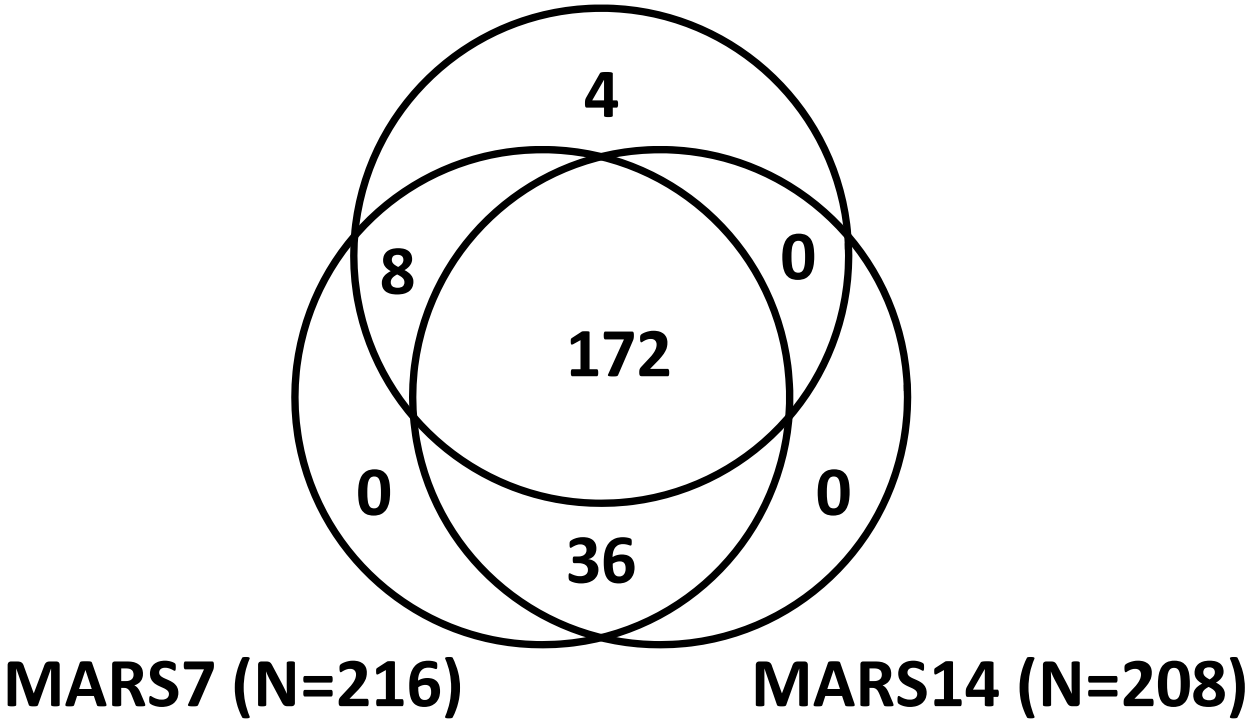


Figure 3

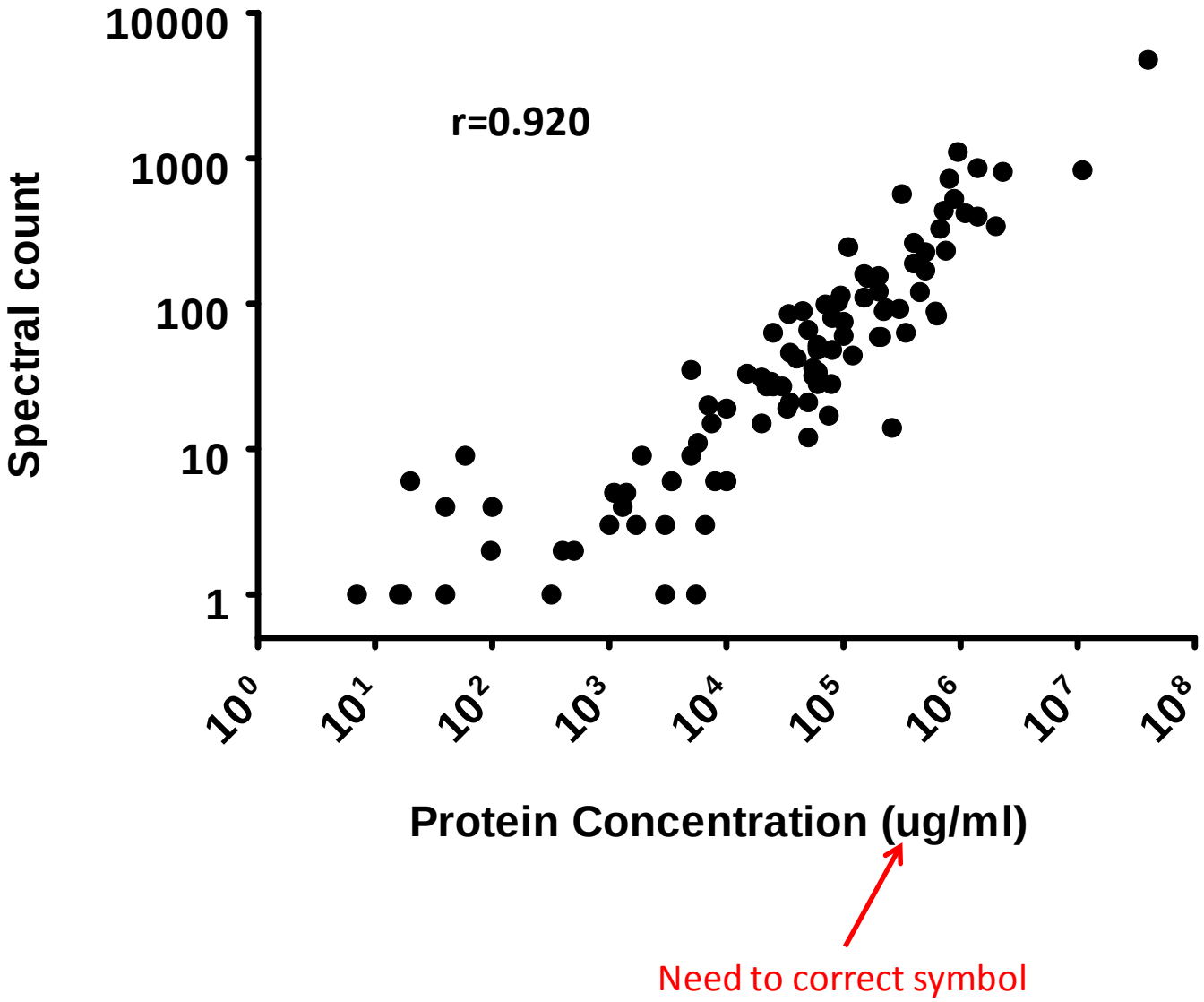
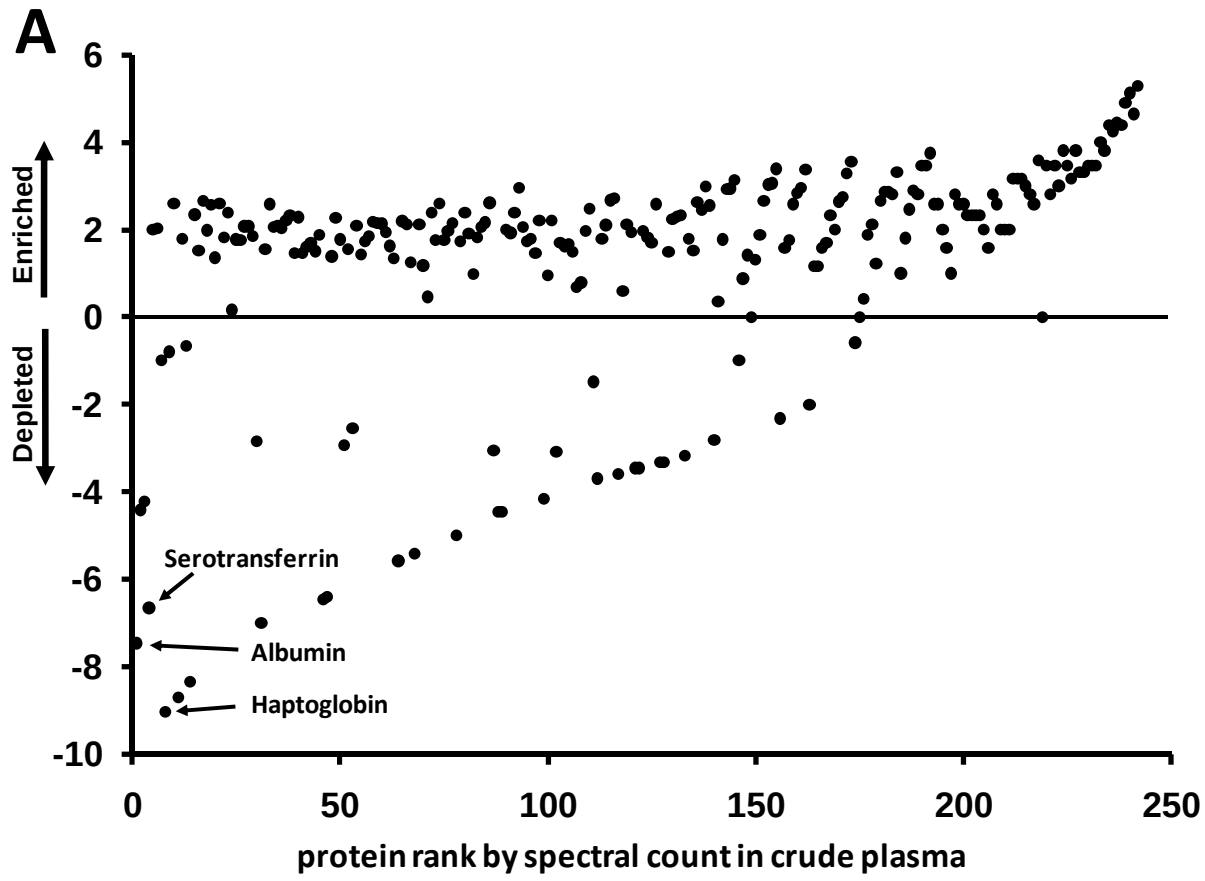

10000
1000
100
10
1
r=0.920
Spectral count
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
Protein Concentration (ug/ml)
Need to correct symbol

Figure 4


A
6
4
2
0
-2
-4
-6
-8
-10
Enriched
Depleted
Serotransferrin
Albumin
Haptoglobin
0
50
100
150
200
250
protein rank by spectral count in crude plasma

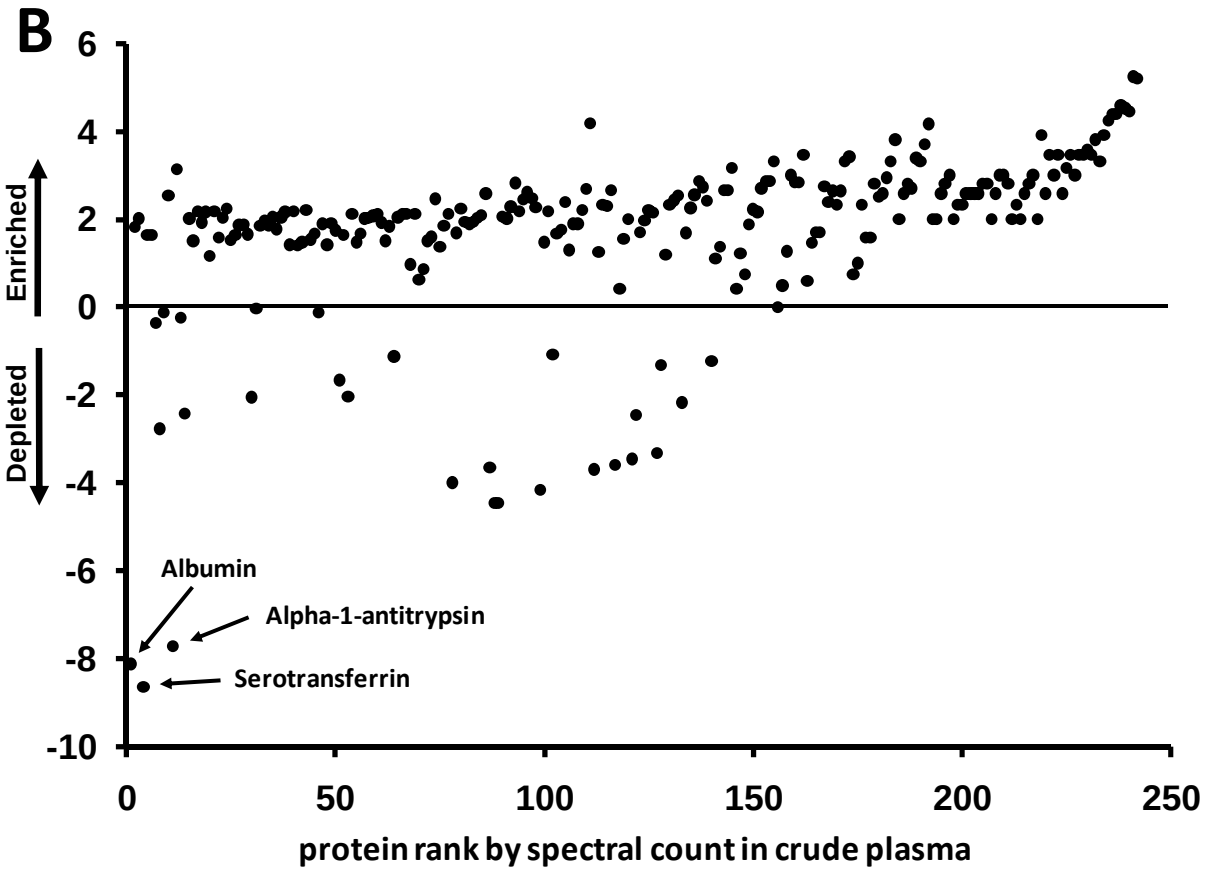

B
6
4
2
0
-2
-4
-6
-8
-10
Enriched
Depleted
Albumin
Alpha-1-antitrypsin
Serotransferrin
0
50
100
150
200
250
protein rank by spectral count in crude plasma

Figure 5

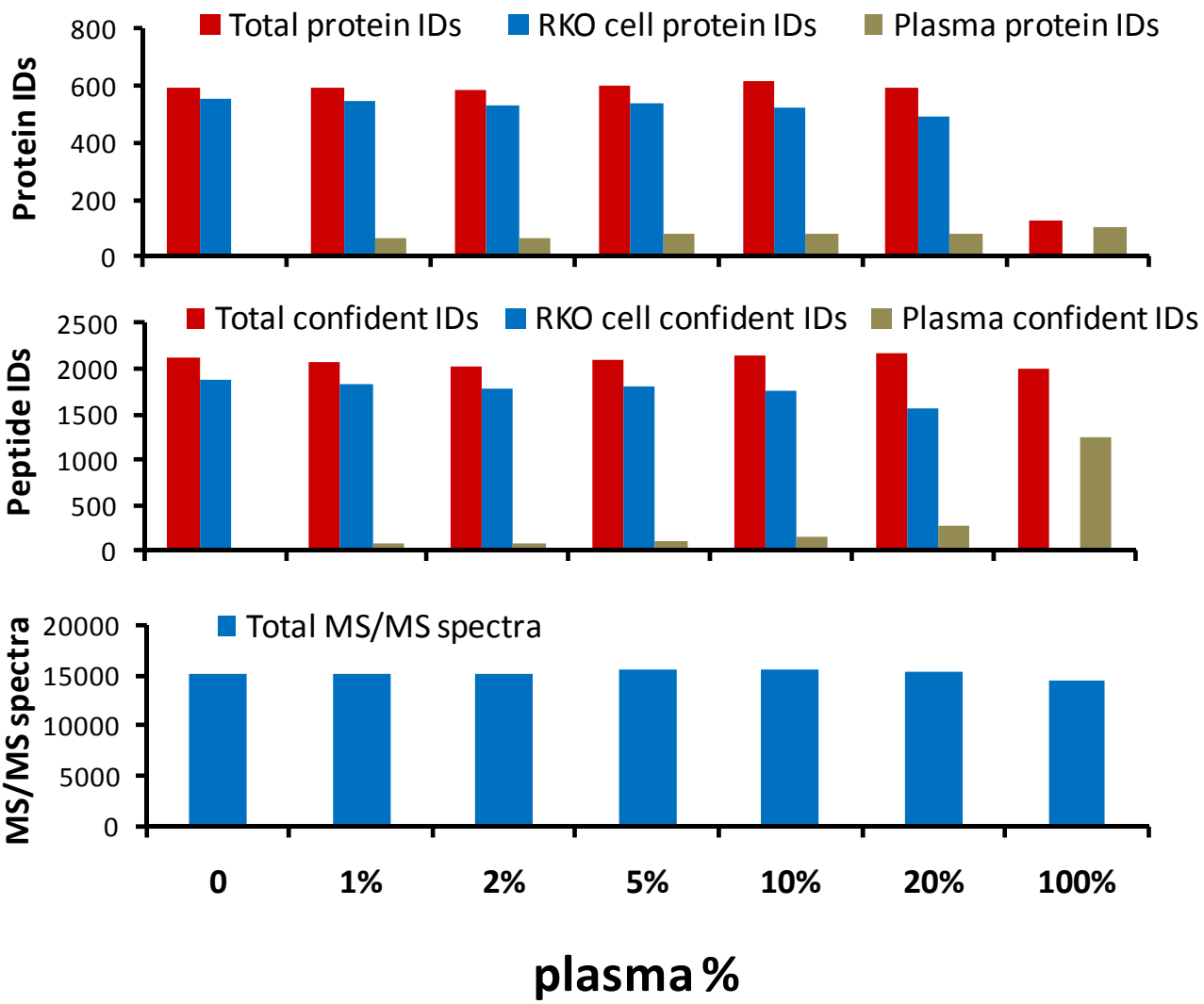

Total protein IDs
RKO cell protein IDs
Plasma protein IDs
Protein IDs
800
600
400
200
0
Total confident IDs
RKO cell confident IDs
Plasma confident IDs
Peptide IDs
2500
2000
1500
1000
500
0
Total MS/MS spectra
MS/MS spectra
20000
15000
10000
5000
0
0
1%
2%
5%
10%
20%
100%
plasma %

Figure 6

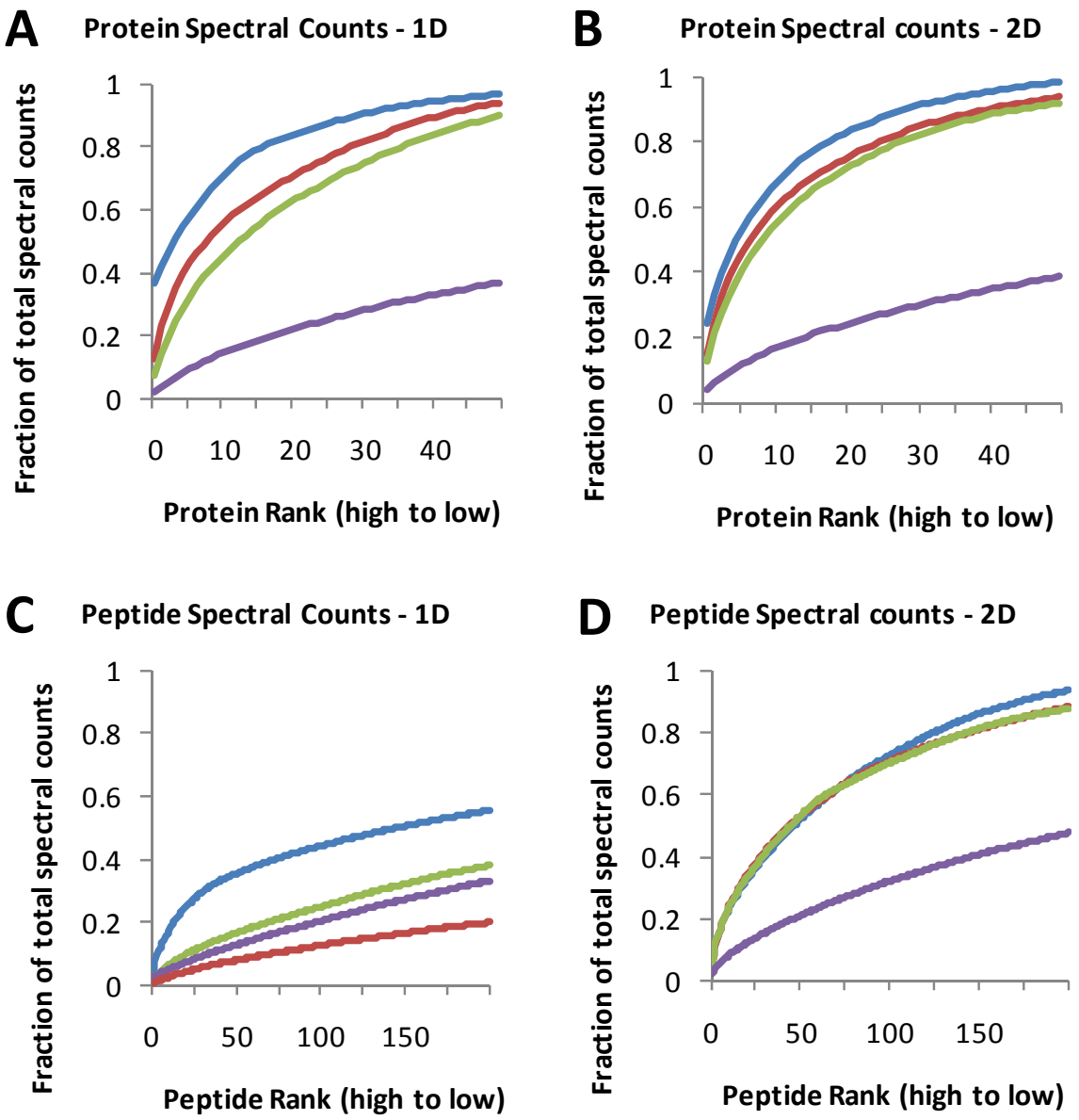

A
Protein Spectral Counts - 1D
Fraction of total spectral counts
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
Protein Rank (high to low)
B
Protein Spectral counts - 2D
Fraction of total spectral counts
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
Protein Rank (high to low)
C
Peptide Spectral Counts - 1D
Fraction of total spectral counts
1
0.8
0.6
0.4
0.2
0
0
50
100
150
Peptide Rank (high to low)
D
Peptide Spectral counts - 2D
Fraction of total spectral counts
1
0.8
0.6
0.4
0.2
0
0
50
100
150
Peptide Rank (high to low)

Figure 7

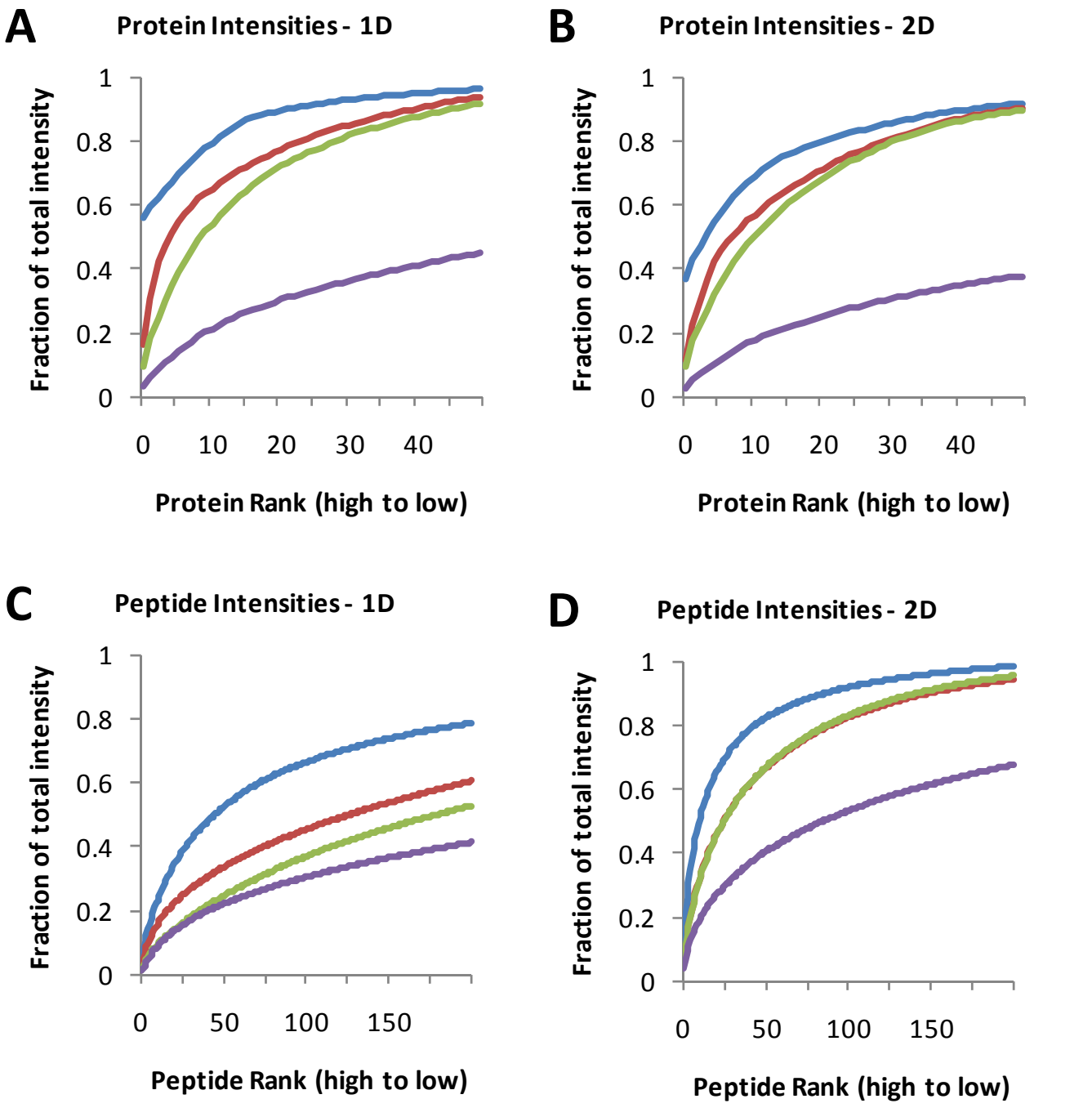

A
Protein Intensities - 1D
Fraction of total intensity
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
Protein Rank (high to low)
B
Protein Intensities - 2D
Fraction of total intensity
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
Protein Rank (high to low)
C
Peptide Intensities - 1D
Fraction of total intensity
1
0.8
0.6
0.4
0.2
0
0
50
100
150
Peptide Rank (high to low)
D
Peptide Intensities - 2D
Fraction of total intensity
1
0.8
0.6
0.4
0.2
0
0
50
100
150
Peptide Rank (high to low)

Figure 8