research articles Journal of proteome research

Head-to-Head Comparison of Serum Fractionation Techniques

Jeffrey R. Whiteaker,[†] Heidi Zhang,[†] Jimmy K. Eng,[†,‡] Ruihua Fang,[†] Brian D. Piening,[†] Li-Chia Feng,[†] Travis D. Lorentzen,[†] Regine M. Schoenherr,[†] John F. Keane,[†] Ted Holzman,[†] Matthew Fitzgibbon,[†] ChenWei Lin,[†] Hui Zhang,[‡] Kelly Cooke,[‡] Tao Liu,[§] David G. Camp II,[§] Leigh Anderson,[||] Julian Watts,[‡] Richard D. Smith,[§] Martin W. McIntosh,[†] and Amanda G. Paulovich*[,†]

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., P.O. Box 19024, Seattle, Washington 98109-1024, Institute for Systems Biology, 1441 N. 34th Street, Seattle, Washington 98103-8904, Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P.O. Box 999, MSIN: K8-98, Richland, Washington 99352, and Plasma Proteome Institute, P.O. Box 53450, Washington DC 20009-3450

Received September 19, 2006

Multiple approaches for simplifying the serum proteome have been described. These techniques are generally developed across different laboratories, samples, mass spectrometry platforms, and analysis tools. Hence, comparing the available schemes is impossible from the existing literature because of confounding variables. We describe a head-to-head comparison of several serum fractionation schemes, including N-linked glycopeptide enrichment, cysteinyl-peptide enrichment, magnetic bead separation (C3, C8, and WCX), size fractionation, protein A/G depletion, and immunoaffinity column depletion of abundant serum proteins. Each technique was compared to results obtained from unfractionated human serum. The results show immunoaffinity subtraction is the most effective means for simplifying the serum proteome while maintaining reasonable sample throughput. The reported dataset is publicly available and provides a standard against which emergent technologies can be compared and evaluated for their contribution to serum-based biomarker discovery.

Keywords: biomarker • human serum • human plasma • fractionation • immunoaffinity depletion • LC/MS

Introduction

There is widespread consensus that informative biomarkers are needed in all areas of medicine, especially cancer, to achieve significant improvements in patient outcomes.[1,2] More effective biomarkers for disease could dramatically improve survival through early disease detection, improve treatment through more accurate diagnosis and prognosis, and enhance clinical trials by revealing and predicting therapeutic response. Ideally, biomarkers could be measured noninvasively from the bloodstream. However, the complexity and range of concentrations ($>10^{10}$) of proteins in human plasma[3] present formidable challenges to discovering highly specific tissue-derived biomarkers (e.g., prostate-specific antigen (PSA), carcinoembryonic antigen (CEA), and alpha-fetoprotein (AFP) present at ng/mL) in the presence of large quantities of interfering proteins (e.g., albumin present at 50 mg/mL, globulin present at 35 mg/mL). To address this challenge, multiple approaches for simplifying the serum proteome have been described, including biophysical fractionation,[4−6] enrichment of target sub-proteomes,[7,8] and immunodepletion of the most abundant interfering proteins.[9,10]

Recent work coupling modern LC-MS instrumentation to multidimensional fractionation of serum has produced a large number of identified proteins from human serum and/or plasma.[11,12] A recent report from a large scale collaborative study organized by the Human Proteome Organization (HUPO) describes 889 high-confidence human blood protein identifications.[13,14] Databases containing these and other data (e.g., PeptideAtlas) provide a growing catalog of known proteins from human serum and/or plasma.[15] Studies such as these engender hope for serum-based biomarker discovery; however, much of the success relies on sophisticated instrumentation and extensive fractionation, contributing to experimental variability and limiting the sample throughput and applicability to routine analysis. Experimental variability coupled with the high-dimensional nature of the data make sample size an extremely important issue when assessing the confidence of potential biomarkers. Thus, the reproducibility and performance of fractionation techniques must be determined when designing a biomarker discovery experiment.

In this study, we address the degree to which the serum proteome can be interrogated with a moderate throughput LC-MS platform (4 samples/instrument/day) coupled to con-

* To whom correspondence should be addressed. Amanda Paulovich, M.D., Ph.D., Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., LE-360, PO Box 19024, Seattle, WA 98109-1024. E-mail: apaulovi@fhcrc.org; Phone: 206-667-1912; Fax: 206-667-2277.
[†] Fred Hutchinson Cancer Research Center.
[‡] Institute for Systems Biology.
[§] Pacific Northwest National Laboratory.
[||] Plasma Proteome Institute.

Published on Web 12/22/2006
10.1021/pr0604920 CCC: $37.00

ventional biochemical fractionation technologies. Specifically, we compare two parameters critical to effective biomarker discovery efforts, the extent of proteome coverage and reproducibility, for several conventional biochemical fractionation techniques. Comparing the effectiveness of fractionation schemes is impossible from the existing literature because of many confounding variables across laboratories, including different serum samples, different chromatography and mass spectrometry platforms, and different analysis tools. In this study, the only variable is the fractionation technology; hence, this study provides the first head-to-head comparison of proteome coverage and reproducibility for eight popular serum fractionation schemes, including N-linked glycopeptide enrichment,[7,16] cysteinyl-peptide enrichment,[8] magnetic bead separation (C3, C8, and WCX),[4] size fractionation, Protein A/G depletion, and immunoaffinity subtraction of the most abundant proteins.[10] Additionally, each technique was compared to results obtained from unfractionated human serum. We discuss the strengths and weaknesses of the various approaches as well as their applicability for biomarker discovery.

Experimental

Materials. Five 100-mL bottles of frozen male human serum (catalog #H4522, lot #043K0502, Sigma-Aldrich, St. Louis, MO) were thawed on ice, combined, and divided into 0.5 mL aliquots in cryovials. The reference serum aliquots were topped with argon gas (to prevent sample oxidation), capped, and stored in liquid nitrogen until use. An aliquot of the reference serum was analyzed by SDS-PAGE and Coomassie staining to verify sample integrity.

Trypsin Digestion. Unless otherwise described (see below and Supporting Information), human serum samples were denatured and reduced with 60% methanol and 10 mM dithiothreitol (DTT) at 60 °C for 1 h and alkylated with 50 mM iodoacetamide (IAM) at room temperature in the dark for 30 min. Ammonium bicarbonate (50 mM) was added to achieve a final methanol concentration of 20% and the samples were digested with Trypsin Gold (Promega, Madison, WI) at a protein to enzyme ratio of 50:1 (w/w) at 37 °C for 6 h. The samples were dried in a SpeedVac and resuspended in 50 mM ammonium bicarbonate prior to LC-MS analysis.

N-Linked Glycopeptide Enrichment. N-linked glycopeptide capture has previously been described in detail.[17] Experimental details can also be found in the Supporting Information. Briefly, 0.6 mL human serum was oxidized and coupled to 4 mL of Affi-Gel Hz Hydrazide Gel (50% slurry) (Bio-Rad, Hercules, CA). N-linked glycopeptides were released from the resin by adding 3 μL PNGase F (glycerol free) (New England Biolabs, Inc., Beverly, MA) in 2 mL of 0.1 M freshly prepared NH_4HCO_3. The cleavage reaction was performed at 37 °C with mixing overnight followed by washing with 2 × 2 mL of 80% ACN. The released glycopeptides were dried and resuspended in 0.4% acetic acid.

Cysteinyl-Peptide Enrichment. Cysteine-peptide capture has previously been described in detail.[8] Experimental details can also be found in the Supporting Information. Briefly, the reduced tryptic digest of human serum protein was incubated with Thiopropyl Sepharose 6B thiol-affinity resin (100 μL; Amersham Biosciences, Piscataway, NJ) for 1 h at room temperature and washed. The captured cysteinyl peptides were released with 100 μL of 20 mM DTT (in washing buffer) followed by 100 μL of 80% ACN. The pH was adjusted to 8.0 and peptides were alkylated with 80 mM IAM for 30 min at room temperature in the dark. The eluted cysteinyl peptides were desalted with a SPE C18 column and lyophilized.

Protein A/G Depletion. Immunoglobulins were depleted by using a 2 mL pre-packed immobilized Protein A/G column (Pierce, Rockford, IL) per the manufacturer's instructions. Briefly, 50 mL of human serum sample was diluted 1:5 with binding buffer and centrifuged at 10 000 × g for 20 min. The supernatant was applied to a Protein A/G column that was pre-equilibrated with 5 mL of binding buffer. 15 mL of binding buffer was added to wash the column, and the flow-through fraction was collected. Sample was then dialyzed against 50 mM ammonium bicarbonate and lyophilized.

Size Fractionation. Human serum samples were diluted 1:5 in 50 mM ammonium bicarbonate/20% ACN (pH 8.0) and centrifuged at 16 000 × g for 5 min at 4 °C to pellet any precipitate. A Millipore filter unit with 30 kDa molecular weight cutoff was rinsed with 50 mM ammonium bicarbonate/20% ACN and centrifuged at 2000 × g for 5 min. Diluted serum sample (400 μL) was added to the washed filter unit and centrifuged at 5000 × g at 4 °C for 30 min. The flow-through aliquot was transferred to a siliconized sterile vial. Ammonium bicarbonate/20% ACN (200 μL) was added to the remaining sample and centrifuged in the same way as in the previous step. This cycle was repeated three times. Coomassie protein assay (Pierce, Rockford, IL) was performed on the pooled flow-through aliquot to verify protein concentration. The flow through aliquot was dried in a SpeedVac and stored at −80 °C until analysis.

Bruker ClinProt C3, C8, and WCX Magnetic Beads. ClinProt purification reagent sets were obtained from Bruker Daltonics (Billerica, MA). The C3 or C8 magnetic beads were shaken 20 times to achieve a homogeneous suspension. MB−HIC binding solution (10 μL) and 5 μL of human serum were mixed in a thin wall PCR tube. Magnetic beads (5 μL) were added to the mixture. After 1 min, the tube was placed in a magnetic bead separator (MBS) for 20 s to separate the beads from the supernatant. The beads were washed with 100 μL washing solution three times. For elution, 5 μL 50% ACN was added to the beads and mixed thoroughly. After 1 min, the tube was placed in a MBS and the beads were separated from the elution solution at the wall of the tubes.

The WCX magnetic bead solution was mixed thoroughly for 1 min. MB−WCX binding solution (10 μL) and 10 μL MB−WCX beads were mixed in a thin wall PCR tube. 5 μL serum was added to the solution and incubated for 5 min. The tube was placed into a MBS and the beads were collected at the wall of the tube for 1 min. The beads were washed two more times with 100 μL MB−WCX wash solution. MB−WCX (5 μL) elution solution was added, and the beads were collected at the tube wall for 2 min. The supernatant was collected.

MARS Column Depletion. The human Multiple Affinity Removal System (MARS) was purchased from Agilent Technologies (Palo Alto, CA) to deplete albumin, transferrin, IgG, IgA, anti-trypsin, and haptoglobin from human serum. Additional experimental details can be found in the Supporting Information. Briefly, 100 μL human serum was diluted 5-fold with Buffer A, filtered (0.22 μm filter), and injected on the MARS column connected to a BioCad Vision HPLC system (Applied Biosystems, Foster City, CA). The flow-through was collected for 6 min (about 1.5 mL), desalted, and resuspended in 50 mM ammonium bicarbonate.

NanoLiquid Chromatography. Agilent 1100 nano flow systems (Agilent Technologies, Palo Alto, CA) equipped with micro

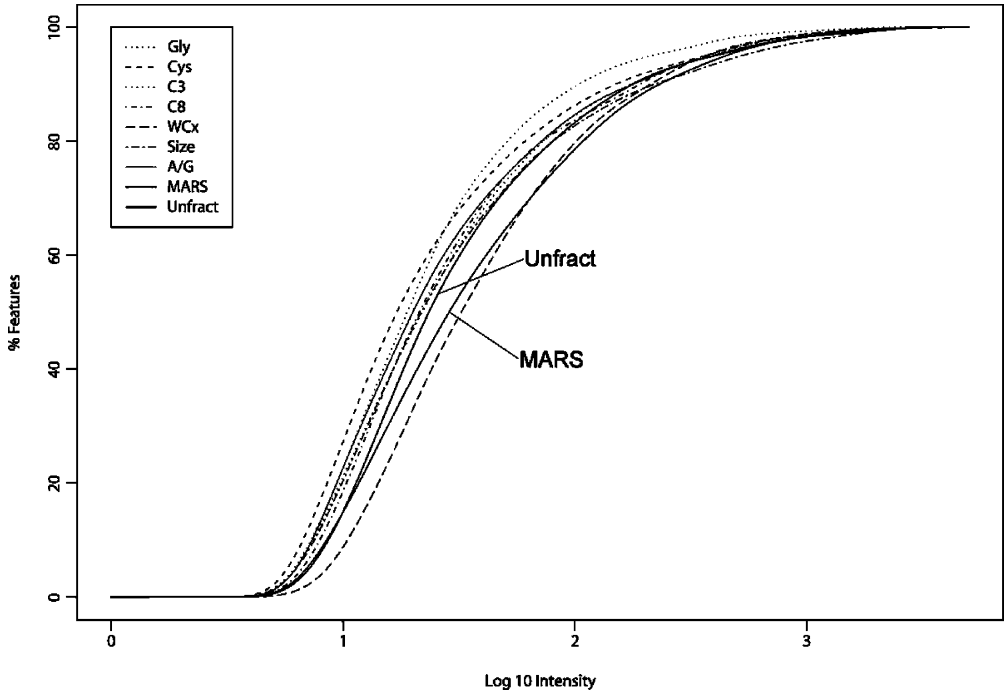

Figure 1. Percent of features with intensities less than a certain value. The log base 10 of the intensity is shown on the *x*-axis and the percent of the total number of features with intensities less than or equal to these are seen by the curves for each scheme.

well-plate autosamplers and isocratic capillary pumps were used for liquid chromatography. Solvents were water/0.1% formic acid (mobile phase A) and acetonitrile/0.1% formic acid (mobile phase B). An IntegraFrit trap column (2 cm × 100 μm, New Objective, Woburn, MA) was connected to a Pico-Frit nanocolumn (15 cm × 75 μm, New Objective, Woburn, MA) via a micro-cross connector (Upchurch Scientific, Oak Harbor, WA). The columns were packed in house at a pressure of 1500 psi using Atlantis C_{18} material (3 μm particle, 100 Å pore size, Waters Corporation, Milford, MA). Samples were loaded on the trapping column by the isocratic pump at a flow rate of 10 μL/min and desalted by washing with 2% B for 5 min. Samples were eluted by a linear gradient of mobile phase B, developed from 10−40% B for 120 min at a flow rate of 200 nL/min.

Electrospray Ionization Time-of-Flight Mass Spectrometry. A time-of-flight mass spectrometer (LCT Premier, Waters Corporation) was interfaced with the nanoLC system (Agilent 1100) for LC-MS analysis. A capillary voltage of 2000 V was applied to a platinum wire in the micro-cross connector; the cone voltage was 60 V. The source temperature was 120 °C. Mass spectra were acquired over the range m/z 400−1600 every 1.0 s with a 0.05 s interscan delay time. The instrument was mass calibrated with a sodium formate solution prior to analysis. Glu-fibrinopeptide (m/z 785.8426) was used as a reference compound for lock-mass calibration.

Electrospray Ionization Linear Ion Trap Mass Spectrometry. The nanoLC system was connected to a linear ion trap mass spectrometer (LTQ, Thermo Electron, San Jose, CA) equipped with a nano electrospray interface operated in the positive ion mode. Typical instrument settings included a spray voltage of 1.5 kV and an ion transfer tube temperature of 200 °C. Voltages across the capillary and the quadrupole lenses were tuned for optimal signal intensity using the +2 ion of angiotensin I (m/z 649). The scan sequence consisted of 1 full MS scan followed by 5 MS/MS scans of the five most abundant ions. Ions were dynamically excluded over a period of 3 min, with a maximum of 500 ions excluded.

Normalization of Peptide Loading. Because the yield of peptides varied among the samples, especially between different fractionation schemes, sample loading was normalized based on total ion currents of LC-TOF-MS pre-runs of each sample. Each sample was run at a series of dilutions and the intensity distribution of features was compared across the different fractionation schemes. The sample loading was adjusted accordingly to normalize the intensity distributions. The maximum amount of peptides that could be loaded without compromising chromatography was empirically determined to be about 10 picomoles (data not shown). Figure 1 shows the intensity distribution for all schemes.

Analysis of LC-MS Profiles. Raw data were converted to mzXML[18] and peptide features were extracted by msInspect[19] by identifying their monoisotopic mass (de-isotoping) and charge state (de-convoluting); thus, feature counts were de-isotoped and deconvoluted. Elution time was denoted by the scan at which maximal signal intensity was reached and intensity was denoted by the maximum signal intensity. MsInspect was also used for alignment of features, briefly, one run was chosen as a reference, and a global nonlinear transformation was computed to register each run against the reference. Features from multiple runs were matched based on a tolerance window for mass (0.2 Da) and elution time (50 scans). Features that did not meet the following criteria were excluded from the analysis: $1 \leq$ charge ≤ 6; $0 < m/z \leq 4000$; intensity ≥ 2; and isotope peaks ≥ 2.

Analysis of MS/MS Data. The MS/MS data were searched with the X! Tandem database search engine[20] with a previously described score plugin,[21] PeptideProphet analysis,[22] and results were uploaded to CPAS (Computational Proteomics Analysis System).[23] Peptide identifications were required to have

Table 1. Number of LC-MS Peptide Features by Run and Scheme[a]

scheme	median features	aligned all	% aligned all	aligned pairwise	% aligned pairwise	median % CV	25% quant. % CV	75% quant. % CV
Gly	6935	390	5.2	2459	35.5	14	10	19
Cys	4615	796	18.1	2085	45.2	19	13	25
C3	2452	320	12.3	988	40.3	16	12	21
C8	1942	12	0.4	521	26.8	22	16	29
WCX	4934	60	1.2	861	17.5	13	8	19
Size	4970	469	9.3	1701	34.2	19	14	25
A/G	2127	155	7.5	631	29.7	11	7	17
MARS	8390	1979	18.3	4407	52.5	11	7	16
Unfract	5798	746	12.8	2736	47.2	15	10	21

[a] For each scheme, the median number of features found, the number of those features that match across all 10 runs, and the median number of features that align among any pairwise alignment of runs are shown. The variation in feature intensity is presented as percent coefficient of variation for the median, 25th, and 75th quantiles. Features are deisotoped and deconvoluted.

Table 2. Cumulative Number of Peptide and Protein Identifications for Each Fractionation Scheme[a]

	cumulative peptide identifications			cumulative protein identifications[b]		cumulative protein identifications[c]		cumulative protein identifications[d]	
scheme	total	unique	unique to scheme	total	unique to scheme	total	unique to scheme	total	unique to scheme
Gly	10420	505	307	153	35	83	6	61	0
Cys	14807	478	57	135	26	64	1	45	2
C3	8439	968	25	157	12	88	0	76	0
C8	7800	716	17	140	13	72	0	60	0
WCX	15724	1328	145	203	29	110	2	93	2
Size	10269	625	90	130	23	73	4	52	2
A/G	4379	291	24	76	10	41	1	30	0
MARS	24283	1958	868	252	92	138	23	117	24
Unfract	8882	761	50	147	9	78	4	57	1

[a] High-confidence peptide identifications were determined by PeptideProphet score ≥ 0.95. [b] One high confidence peptide identification required for protein assignment [c] Two high confidence peptide identifications required for protein assignment [d] ≥3 high confidence peptide identifications required for protein assignment

PeptideProphet score ≥0.95. A peptide was considered unique if it had a unique raw peptide sequence; peptides identified with unmodified and oxidized forms of methionine were only counted once. All data were searched against version 3.02 of the Human International Protein Index (IPI) sequence database released on January 6, 2005. All searches were performed with tryptic enzyme constraint allowing for up to two missed cleavages. Peptide MH+ mass tolerances were set at ±3.0 Da. Oxidized methionine was set as a variable modification for all fraction methods. Cysteine residues were considered alkylated for all methods except for size fractionation. Additionally, X! Tandem by default considers N-terminal glutamine and glutamic acid modifications to pyrolidone carboxylic acid ($-NH_3$ and $-OH$, respectively).

Bioinformatic Assessment of Performance of Fractionation Schemes. Only unique peptides with a PeptideProphet score ≥0.95 were included in the bioinformatic analysis (Table 4). A single protein identifier was mapped to each peptide. If a peptide is encapsulated in multiple entries in the sequence database, the first protein in the database which contains the peptide sequence is chosen and associated with the peptide identification. The *N*-glycosylation motif used in the first column (Table 4) was "Asn-X-[Ser/Thr]" where X can be any amino acid except proline. Similarly, the cysteinyl peptides were determined by counting the peptides that contained "C". The molecular weights (MWs) were determined by summing the average MW for each amino acid in the parent sequence of the peptides. The parent sequences were taken from the FASTA file provided with IPI v3.02, released on 4-Jan-2005. The protein pIs and hydrophobicities were derived from the same sequences. A peptide hydrophobicity model was applied to whole proteins.[24] The peptides that derived from IgG were determined as follows: a list of accession numbers for human IgG was downloaded from GeneCards. These accession numbers were mapped to the set of all human IPI sequences which were constructed using them, or which referred to the same sequences. This generated a list of 220 IPI identifiers. The IPI numbers of *all* peptides (PeptideProphet ≥0.95) matching this list were tallied.

Quality Control. Additional details can be found in the Supporting Information. Briefly, repeats for each fractionation scheme were run in pairs separated by extensive column washing and blank runs. The overall run order for digests of seven of the eight fractionation schemes plus the unfractionated serum was randomized. Thus, repeats from each fractionation scheme were analyzed on several different columns and under various conditions to remove biases. Because the MARS column was not available when the other schemes were run, it was completed at a later date. To confirm the results, a second dataset of 10 independent preparations of the MARS column depletion were completed at a later date. To ensure comparability of the data, aliquots of the reference serum were run throughout the generation of both datasets to serve as an internal reference. Specifically, repeat runs of the unfractionated reference serum were alternated with the second MARS dataset to most approximate identical conditions to the other fractionation schemes (see Supporting Information for more details). Additionally, a Quality Control (QC) peptide mix was profiled on the LC-MS platform after every pair of serum runs.

Public Access to the Dataset. The data are available on the public CPAS website at http://proteomics.fhcrc.org/CPAS in the "Serum Fractionation" folder under the "Published Experiments."

Table 3. Sequence Coverage of Abundant Serum Proteins

IPI	protein name	MW	% sequence coverage ALL	Gly	Cys	C3	C8	WCX	size	A/G	MRS	Unf
IPI00022434	serum albumin	69367	91.3	46.0	77.7	81.8	83.3	81.4	48.3	51.1	53.7	91.1
IPI00164623	complement C3	187164	80.0	15.7	18.3	53.2	28.1	60.9	26.9	11.4	74.4	46.5
IPI00021841	apolipoprotein A−I	30778	78.7	57.3	0.0	65.2	63.3	70.0	71.5	41.9	71.2	61.4
IPI00022463	serotransferrin	77050	80.7	34.1	38.8	67.8	70.5	73.6	36.0	43.4	12.2	76.9
IPI00465313	α−2-macroglobulin	166127	63.1	17.7	15.5	8.8	14.8	18.7	14.1	12.2	51.9	28.8
IPI00305457	α−1-antitrypsin	46737	83.7	69.6	13.4	51.0	59.1	55.7	59.3	52.4	0.0	61.5
IPI00022229	apolipoprot. B−100	515563	52.6	5.8	2.1	6.4	0.0	17.4	13.6	0.6	49.4	7.6
IPI00032258	complement C4	192771	60.4	15.4	7.2	19.4	21.8	39.4	9.9	2.8	55.3	24.9
IPI00431645	HP protein	31382	50.5	32.4	33.5	31.0	36.7	36.7	0.0	35.9	36.7	36.7
IPI00022488	hemopexin	51676	78.6	41.1	34.4	48.3	47.0	58.0	16.0	13.0	70.3	39.8
IPI00304273	apolipoprotein A-IV	45371	74.2	0.0	0.0	59.8	34.8	63.6	72.0	0.0	64.1	19.9
IPI00019580	plasminogen	90569	73.7	0.0	37.2	54.6	30.9	69.6	2.7	0.0	65.3	22.8
IPI00298853	vitamin D-binding	52964	79.1	0.0	28.7	61.2	59.7	57.4	13.3	13.3	75.9	54.0
IPI00029739	complement factor H	139125	58.2	10.4	26.6	31.0	3.9	35.2	0.0	1.1	50.4	15.4
IPI00022431	α−2-HS-glycoprot.	39325	63.5	17.4	38.1	41.7	26.4	39.5	8.2	0.0	47.1	37.1
IPI00017601	ceruloplasmin	122205	61.8	13.5	19.6	19.5	22.8	29.5	18.5	8.9	54.3	35.4
IPI00026195	hypothetical protein	26235	40.6	23.0	22.2	36.4	37.2	38.5	29.7	15.9	7.1	40.6
IPI00004618	IGHG4 protein	51986	39.3	19.2	33.2	28.5	28.5	28.5	11.2	4.7	0.0	28.5
IPI00019591	complement factor B	85533	66.2	6.9	16.9	36.4	33.2	44.5	20.0	11.5	53.1	22.8
IPI00061977	MGC27165 protein	54154	40.8	21.6	20.4	19.0	25.0	25.0	5.4	13.6	0.0	28.8
IPI00022432	transthyretin	15887	73.5	72.8	28.6	68.7	69.4	69.4	69.4	65.3	69.4	69.4
IPI00019568	prothrombin	70037	62.9	13.3	31.8	43.2	26.5	47.9	21.4	9.2	50.3	19.1
IPI00021854	apolipoprotein A-II	11175	72.0	51.0	43.0	70.0	61.0	61.0	70.0	33.0	59.0	61.0
IPI00032328	kininogen	71945	40.4	11.3	14.0	26.4	13.4	24.4	12.1	3.7	30.0	18.0
IPI00218192	inter-α-trypsin inhibitor heavy H4	101242	56.0	6.3	2.2	24.8	17.0	35.8	19.5	7.8	47.8	11.2
IPI00332161	hypothetical protein	50927	38.4	21.0	18.5	22.7	22.7	21.0	3.6	0.0	0.0	23.6
IPI00296170	haptoglobin-related	43078	57.1	13.0	23.9	23.1	19.5	23.6	7.3	18.7	34.3	22.6
IPI00026314	gelsolin	85698	50.6	0.0	1.4	39.3	28.4	49.5	7.7	0.0	42.6	14.7
IPI00298828	β−2-glycoprotein I	38298	78.6	22.9	54.5	42.0	35.9	54.2	0.0	0.0	61.7	46.4
IPI00022429	α−1-acid glycoprot. 1	23512	43.8	18.4	21.9	19.4	19.4	25.4	12.4	12.4	32.8	19.4

[a] The 30 most abundant proteins detected in this dataset were determined by total peptide identifications. "MW" stands for molecular weight, "ALL" reports percent sequence coverage for abundant proteins detected from all fractionation schemes together.

Table 4. Bioinformatic Assessment of Performance of Fractionation Schemes

scheme	% *N*-glyco motif in peptides	mean MW of proteins	% cysteinyl peptides	mean p*I* of proteins	IgG peptides	mean hydrophobicity of proteins
Gly	55.5	76568 ± 80647	37.0	6.78 ± 1.36	1054 (10%)	96.21 ± 17.59
Cys	1.5	75582 ± 86760	86.6	6.55 ± 1.29	8442 (57%)	92.57 ± 16.02
C3	2.2	67436 ± 77380	38.1	6.78 ± 1.39	3104 (37%)	91.05 ± 16.11
C8	2.9	50295 ± 39531	4.3	6.94 ± 1.60	3468 (45%)	88.56 ± 15.90
WCX	2.4	82769 ± 272354	30.3	6.95 ± 1.27	5513 (35%)	93.06 ± 18.37
SIZE	2.4	65266 ± 66499	4.3	6.53 ± 1.27	1878 (18%)	95.63 ± 16.12
A/G	1.7	70191 ± 81783	8.6	6.62 ± 1.43	2489 (57%)	94.53 ± 16.34
MARS	2.4	72457 ± 81783	31.2	6.67 ± 1.42	4515 (19%)	94.73 ± 18.15
Unfract	2.1	59860 ± 73731	41.7	6.90 ± 1.53	4673 (53%)	89.71 ± 15.59
Entire IPI	7.4	51518 ± 61398	28.0	7.56 ± 1.90	—	90.42 ± 24.91

[a] See experimental section for details on how calculations were performed. Mean molecular weight (MW) of proteins, mean isoelectric point (p*I*) of proteins, and mean hydrophobicity of proteins are presented ±1 standard deviation.

Results and Discussion

Overview of Experimental Design. In this study, we set out to ask to what extent and with what reproducibility the serum proteome could be interrogated with moderate throughput (4 samples/instrument/day) using current biochemical and LC-MS technologies. We began with a large volume of reference human serum, from which identical aliquots were subjected to eight state-of-the-art biochemical fractionation schemes: N-linked glycopeptide enrichment,[7,16] cysteinyl-peptide enrichment,[8] magnetic bead separation (Bruker ClinProt C3, C8, and WCX),[4,25] size fractionation, Protein A/G depletion, and immunoaffinity subtraction of the most abundant proteins (MARS column immunodepletion).[9,10] Each scheme was repeated *10 independent times* to assess reproducibility. Ten independent tryptic digests of *unfractionated* serum were also included as a control. Peptides from each of the 10 repeats from the 8 fractionation schemes (as well as 10 aliquots of unfractionated serum) were subjected to LC-MS analysis on a single time-of-flight (TOF) mass spectrometer and LC-MS/MS analysis on a single linear ion trap instrument (a total of 180 LC-MS runs), generating the largest dataset reported for assessing the performance of serum fractionation.

Assessment of Reproducibility. Reproducibility of LC-MS measurements is critical for biomarker discovery experiments. We assessed the reproducibility of fractionation protocols using two parameters of LC-MS measured on an electrospray time-of-flight mass spectrometer: (1) the alignment of features across multiple runs, and (2) the variation of the intensity of features across the 10 runs, in addition to one parameter of

LC-MS/MS on a linear ion trap: (3) the variability of peptide identifications.

We define a "feature" as the signature of a unique peptide sequence observed in the LC-MS data; multiple isotopic peaks from a single peptide were combined (deisotoped) and multiple charge states from a peptide were combined (deconvoluted) into a single feature. Data showing the total number of features, alignable number of features, and reproducibility of feature intensities are summarized for each scheme in Table 1. The median number of features seen in a run for each fractionation scheme varied 4-fold from a median of 1942 features for C8 beads to 8390 for MARS column depletion. The median number of features that aligned between any two pairs of runs varied 8-fold between these schemes (Table 1), from a median of 521 features aligned for C8 bead fractionation to 4407 features for MARS depletion. The number of features that aligned across all 10 runs, the most strict measure of reproducibility, varied even more dramatically (160×). Alignment across all 10 runs is a more stringent measure as only one poor performing replicate among all 10 schemes will be sufficient to affect the number of aligned features. The median pairwise alignment measure tolerates a small number of poor replicates, and so may be interpreted as an upper-bound of reproducibility performance. Compared to unfractionated serum, MARS depletion had the highest number of features and median pairwise alignment.

Feature alignment across runs is challenging due to multiple factors, including the imperfect reproducibility of the fractionation protocols, the overall low intensity of features, and current limitations of LC-MS instrumentation and alignment tools. For each scheme, the majority of features were of relatively low intensity. For example, approximately 85% of the features detected for each scheme have intensity <100 (Figure 1). Some improvements to alignment can be expected by using higher quality ultra-HPLC separations,[26,27] more powerful MS instrumentation with accurate mass, and improved feature alignment algorithms.[28]

Another aspect of reproducibility from Table 1 is the variation in feature intensity. This is reported in Table 1 as the percent coefficient of variation for log-intensity values for all features. The median percent CVs, as well as 25th and 75th percentile are reported. The median CV for feature intensity ranged from 11 to 22%. Features were measured over 10 repeats with median CV less than 25%. In some cases, manipulation of the serum (i.e., fractionation) lead to an increase in the CVs of the intensities compared to those for unfractionated serum (see Table 1). However, some schemes (e.g., MARS depletion, Protein A/G depletion) showed an improvement in CVs over the unfractionated serum. This phenomenon has been ascribed to improved chromatography following depletion of interfering abundant serum proteins.[29]

Figure 2 summarizes the cumulative number of new peptide and protein identifications from each LC-MS/MS repeat of a given fractionation protocol. The number of *new* peptide and protein identifications for most schemes was very low approaching the tenth run. Focusing on the number of peptide identifications (Figure 2a), the schemes generally fall into three classes: N-linked glycopeptide enrichment, cysteinyl-peptide enrichment, size fractionation, Protein A/G depletion, MARS depletion, and unfractionated serum all demonstrated relative rapid leveling of the identification curves and tight box plots, indicating low run-to-run variability. WCX, C3, and C8 magnetic bead-based fractionation showed more diffuse box plots and slower leveling curves, suggesting greater variation among the repeats.

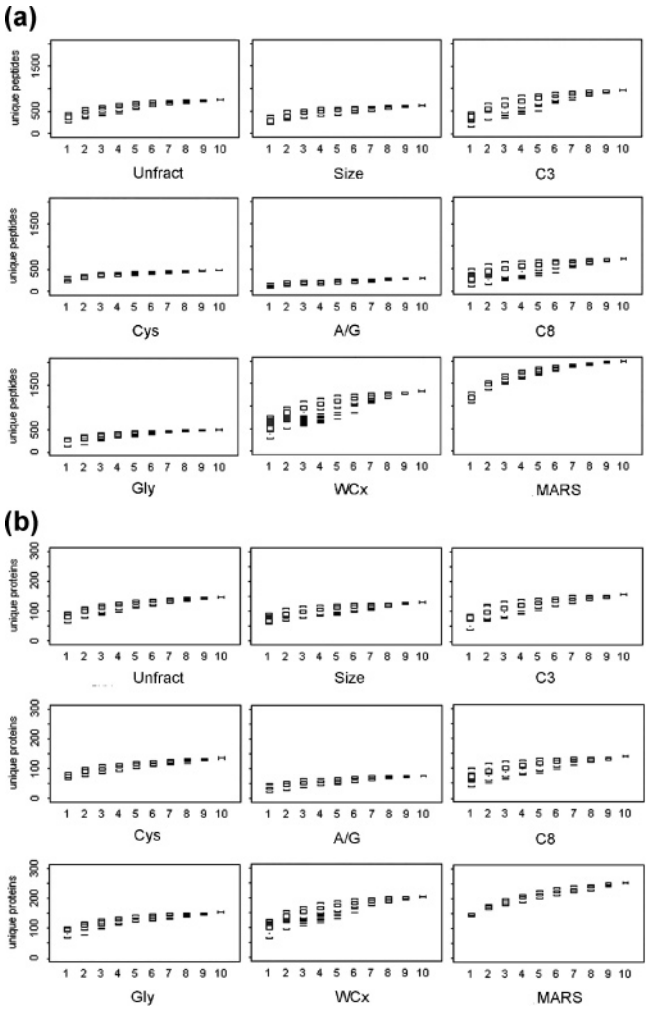

Figure 2. Cumulative number of peptides (a) or proteins (b) identified over multiple runs of each fractionation scheme. Each box plot indicates the number of unique peptides or proteins that are identified on any combination of *N* number of runs where *N* is the *x*-axis value ranging from 1 to 10. Lines on the box plot indicate the 0, 1, 5, 10, 25, 50, 75, 90, 95, 99, and 100% range of the data. Methods that level off faster in these graphs indicate that new peptides/proteins are not being discovered in subsequent runs. Tighter box plots indicate less variability as a function of LC-MS/MS runs at the peptide identification level.

Assessment of Extent of Proteome Coverage. In addition to reproducibility, proteome coverage is an important parameter in biomarker discovery because it is desirable to sample as large of a space of the proteome as possible to maximize the chances of detecting a low abundance biomarker. We estimated proteome coverage for each fractionation scheme by three parameters of LC-MS/MS data obtained on an electrospray linear ion trap mass spectrometer: (1) the total number of different peptides/proteins identified across 10 repeats, (2) the number of identified peptides/proteins unique to a given scheme, and (3) diminution of sequence coverage of the most abundant serum proteins in the dataset. The term "proteome coverage" refers to these criteria as the actual depth is not known for all of the proteins that are reported.

One metric to evaluate breadth and extent of coverage is the cumulative number of different peptide and protein identifications from all 10 runs for each fractionation method. As can be seen in Table 2, the number of high-confidence unique peptide identifications varied from 291 for Protein A/G depletion to 1958 for MARS depletion. The number of proteins identified varied from 76 for Protein A/G depletion to 252 for MARS column immunodepletion. Including the stringent requirement of ≥3 peptides present for protein identification, the range of total number of proteins decreased to 30 for Protein A/G and 117 for MARS depletion. For N-linked glycopeptide and cysteinyl-peptide enrichment (which seek to deliberately capture only a subset of peptides per protein), one might expect a lower number of proteins to have multiple peptide identifications. Overall, MARS depletion led to the highest number of cumulative peptide and protein identifications compared to the other methods.

Another metric to evaluate proteome coverage is the number of identified peptides/proteins *unique* to a given scheme (i.e., not observed following any of the other fractionation schemes). This is important because some fractionation schemes may be complementary in surveying different subsets of the proteome. As can be seen in Table 2, MARS depletion allowed detection of 92 proteins not observed using any other scheme. This was, on average, 4.6× the number of proteins unique to any other scheme. This difference was even more pronounced when requiring 2 or more peptides per protein identification (Table 2). To further explore the potential of fractionation schemes to survey complementary subsets of the proteome, we examined the percent overlap between proteins identified for each pairwise combination of schemes as well as for each scheme relative to unfractionated serum (Supplemental Table 1, Supporting Information). There was 28−55% overlap in the proteins identified in the fractionated samples with those identified in the unfractionated control. The overall high degree of overlap among the schemes demonstrated that each scheme predominantly sampled a subset of the same proteins.

A third metric to evaluate proteome coverage is the degree of diminution of sequence coverage of the most abundant serum proteins. This is because, although there are predicted to be 1−10 million unique peptides in serum (including the repertoire of circulating antibodies), the 30 most abundant serum proteins constitute 99% of total serum protein mass and sit atop a greater than 10 order of magnitude range of concentration of serum proteins.[3] Hence, because it is the very high concentration of abundant interfering proteins that prohibit deeper sampling of the proteome, it is reasonable to assume that any fractionation technology allowing significant sampling of the low abundance proteome must be associated with decreased sequence coverage of the abundant proteins. Therefore, we evaluated each fractionation scheme for the sequence coverage of the top 30 proteins from the dataset.

Figure 3 shows the numbers of peptides identified (for each fractionation scheme) that correspond to the 30 most abundant proteins from this dataset (by peptide counts) vs the numbers of identified peptides that map to all other serum proteins. As expected, using unfractionated serum, more high abundance peptides than lower abundance peptides were observed. With the exception of N-linked glycopeptide enrichment and MARS depletion, each fractionation method resulted in a higher number of abundant peptide identifications. MARS depletion resulted in the greatest ratio of low:high abundance peptides

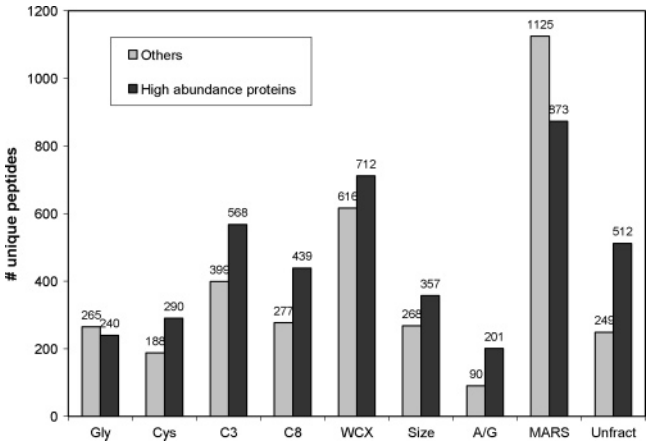

Figure 3. Number of detected peptides derived from the 30 most abundant serum proteins for each fractionation scheme. For each fractionation scheme, the bar graphs display the number of peptides identified that map to high abundance serum proteins vs all other serum proteins.

(∼1.3) and identified significantly more lower abundance peptides overall (1125 compared to 616 for WCX beads).

Finally, Table 3 summarizes the sequence coverage of the 30 most abundant proteins found in the LC-MS/MS data following fractionation. For the majority of the fractionation schemes, coverage of the abundant proteins did not appear to be significantly diminished compared to unfractionated serum. But as reported previously, N-linked glycopeptide enrichment was associated with an approximate 2-fold reduction in albumin sequence coverage.[7,16] Also, as expected, MARS depletion significantly reduced the coverage of albumin, transferrin, and antitrypsin, all of which are targeted by the immunoaffinity column. MARS did not show a decrease in sequence coverage for the haptoglobin-related protein (IPI00296170), a protein with 71% sequence homology to haptoglobin (IPI00478493). However, for peptides found only in haptoglobin, sequence coverage decreases from 12.6% in unfractionated serum to 0% following MARS depletion. The MARS column likely does not have high affinity for the haptoglobin-related protein. (Note also that MARS depletion of this subset of abundant proteins was also associated with an increase in the sequence coverage of the remaining abundant proteins, due to the "zoom-in" effect of depletion.)

Assessment of Performance of Fractionation Schemes. We can gauge the performance of many of the fractionation protocols via bioinformatic analysis of our dataset. For example, if N-linked glycopeptide enrichment was effective, a greater proportion of peptides identified following glycopeptide capture should contain the N-linked glycosylation motif (N−X−S/T, where X denotes any amino acid except proline)[30] compared to unfractionated serum. Indeed, 55% of unique peptides identified after glycopeptide capture contained the *N*-glycosylation motif whereas only 2% of unique peptides identified from the unfractionated serum contained the motif (Table 4). Similarly, 87% of unique peptides identified after cysteinyl-peptide enrichment contained a Cys residue, whereas only 42% of unique peptides identified from the unfractionated serum contained a cysteine. Evaluating the total number of peptides (using spectral counts) provided further evidence of the enrichment of these techniques. For example, the mean percent of total identified peptides containing the *N*-glycosylation was 83.8%, whereas the mean for unfractionated was

0.75% (measured over 10 independent repeats). Similarly, the mean percent of total identified peptides containing Cys was 97.1%, whereas the mean for unfractionated was 49.6% (measured over 10 independent repeats). As discussed above, MARS depletion was successful as evidenced by the significant reduction in sequence coverage of peptides emanating from proteins that were targeted for depletion (Table 3). MARS also significantly depleted peptides from IgG (Table 4). The data were not so compelling for the remaining schemes. For example, size fractionation did not significantly change the molecular weight distribution of the proteins identified (Table 4). This has been reported previously and may be due to the presence of degradation products of the most abundant serum proteins.[31,32] Protein A/G depletion did not impact sequence coverage of IgG, likely due to saturation of the column by the abundant immunoglobulins. Similarly, C3 and C8 bead-based fractionation did not significantly change the hydrophobicity distribution of the identified proteins relative to unfractionated serum, nor did fractionation by WCX beads impact the distribution of isoelectric points (Table 4).

Conclusions

MARS column depletion was unique among the methods tested in demonstrating comparatively high performance both in terms of proteome coverage and reproducibility. This is not surprising because the MARS column specifically depletes six of the most abundant interfering proteins and because the depletion process utilizes fully automated LC separation and fraction collection, and thus is more likely to provide better reproducibility. We note that our results apply specifically to analysis of the human serum proteome, which is uniquely challenging due to the extraordinary range of protein concentrations. For example, the cysteinyl-peptide enrichment method has proved to be effective in reducing sample complexity and detecting low abundance proteins for samples prepared from mammalian cells, although it failed to increase proteome coverage in the current study. This is because ~95% of the cysteinyl peptides from human plasma are contributed by only two highly abundant proteins: albumin and transferrin. Thus, combining MARS depletion and cysteinyl-peptide enrichment should be able to provide broader proteome coverage, especially given the superior reproducibility and relatively low overlap in peptide identifications (Supplementary Table 1, Supporting Information) for both methods. Other fractionation techniques may also be useful when combined with depletion; however, only the most reproducible strategies should be used in combination because further sample manipulation will lead to increased variability.

Although reproducibility and proteome coverage are important parameters for determining the success of a biomarker discovery platform, other parameters that are more difficult to measure are also important. For example, a fractionation scheme yielding lower reproducibility and proteome coverage might still be advantageous if it were capable of selecting a subset of the proteome most likely to contain biomarkers. For example, one might argue that secreted or cell surface molecules have the greatest possibility of reaching the plasma. Hence, N-linked glycopeptide enrichment, which enriches the secreted and cell surface subpopulation of proteins, could arguably be useful despite the overall lower number of total proteins interrogated (Figure 2).

Failure of the present fractionation technologies to extend sampling of the serum proteome more than 1.7-fold compared to unfractionated serum (Table 2) highlight the need for great improvements in fractionation technologies and better fractionation schemes. New technologies should be aimed at true reduction in the dynamic concentration range of serum proteins (such as that achieved by MARS depletion) rather than simple partitioning based on biophysical properties, because the abundance range of proteins is the major obstacle to deeper sampling. It is desirable to remove as many high-abundance proteins as possible. For example, although the data indicate that the MARS column was effective in depleting most of the targeted abundant proteins (albumin, IgG, IgA, haptoglobin, transferrin, and antitrypsin), sequence coverages of the remaining most abundant serum proteins were universally increased, and now predominate the data (Table 3). Our publicly available dataset provides a valuable standard against which new fractionation technologies could be compared and evaluated for their potential contribution to serum-based biomarker discovery before ever being deployed in time- and resource-intensive biomarker discovery experiments.

Acknowledgment. This work was funded by NCI subcontract 23XS144A as well as by generous gifts from the Canary Foundation, the Keck Foundation, and the Paul G. Allen Family Foundation. We thank Andrea E. Detter for technical assistance with figures and tables. We also thank the Environmental Molecular Sciences Laboratory at PNNL for use of some of the instrumentation applied in this research. The Environmental Molecular Sciences Laboratory is a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC06-76RLO 1830.

Supporting Information Available: Supplemental data include detailed experimental protocols for N-linked glycopeptide enrichment, cysteinyl-peptide enrichment, MARS column depletion, quality control information, a table and figure characterizing repeated unfractionated and MARS depleted serum runs, and a table showing the percent overlap between proteins identified for each pairwise combination of schemes as well as for each scheme relative to unfractionated serum. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Etzioni, R.; Urban, N.; Ramsey, S.; McIntosh, M.; Schwartz, S.; Reid, B.; Radich, J.; Anderson, G.; Hartwell, L. *Nat. Rev. Cancer* **2003**, *3*, 243−252.
(2) Hartwell, L.; Lander, E. Report to National Cancer Advisory Board, NCAB Working Group on Biomedical Technology, 2005.
(3) Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1*, 845−867.
(4) Baumann, S.; Ceglarek, U.; Fiedler, G. M.; Lembcke, J.; Leichtle, A.; Thiery, J. *Clin. Chem.* **2005**, *51*, 973−980.
(5) Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. *Mol. Cell. Proteomics* **2003**, *2*, 1096−1103.
(6) Marshall, J.; Jankowski, A.; Furesz, S.; Kireeva, I.; Barker, L.; Dombrovsky, M.; Zhu, W.; Jacks, K.; Ingratta, L.; Bruin, J.; Kristensen, E.; Zhang, R.; Stanton, E.; Takahashi, M.; Jackowski, G. *J. Proteome Res.* **2004**, *3*, 364−382.
(7) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21*, 660−666.
(8) Liu, T.; Qian, W. J.; Strittmatter, E. F.; Camp, D. G., 2nd; Anderson, G. A.; Thrall, B. D.; Smith, R. D. *Anal. Chem.* **2004**, *76*, 5345−5353.

(9) Pieper, R.; Su, Q.; Gatlin, C. L.; Huang, S. T.; Anderson, N. L.; Steiner, S. *Proteomics* **2003**, *3*, 422−432.
(10) Brand, J.; Haslberger, T.; Zolg, W.; Pestlin, G.; Palme, S. *Proteomics* **2006**, *6*, 3236−3242.
(11) Chan, K. C.; Lucas, D. A.; Hise, D.; Schaefer, C. F.; Xiao, Z.; Janini, G. M.; Buetow, K. H.; Issaq, H. J.; Veenstra, T. D.; Conrads, T. P. *Clin Proteomics* **2004**, *1*, 101−226.
(12) Shen, Y.; Kim, J.; Strittmatter, E. F.; Jacobs, J. M.; Camp, D. G., 2nd; Fang, R.; Tolie, N.; Moore, R. J.; Smith, R. D. *Proteomics* **2005**, *5*, 4034−4045.
(13) States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M. *Nat. Biotechnol.* **2006**, *24*, 333−338.
(14) 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. *Proteomics* **2005**, *5*, 3226−3245.
(15) Deutsch, E. W.; Eng, J. K.; Zhang, H.; King, N. L.; Nesvizhskii, A. I.; Lin, B.; Lee, H.; Yi, E. C.; Ossola, R.; Aebersold, R. *Proteomics* **2005**, *5*, 3497−3500.
(16) Zhang, H.; Yi, E. C.; Li, X. J.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G., 2nd; Smith, R. D.; Kemp, C. J.; Aebersold, R. *Mol. Cell. Proteomics* **2005**, *4*, 144−155.
(17) Zhang, H.; Aebersold, R. *Methods Mol. Biol.* **2006**, *328*, 177−185.
(18) Pedrioli, P. G.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R. *Nat. Biotechnol.* **2004**, *22*, 1459−1466.
(19) Bellew, M.; Coram, M.; Fitzgibbon, M.; Igra, M.; Randolph, T.; Wang, P.; May, D.; Eng, J.; Fang, R.; Lin, C.; Chen, J.; Goodlett, D.; Whiteaker, J.; Paulovich, A.; McIntosh, M. *Bioinformatics* **2006**, *22*, 1902−1909.
(20) Craig, R.; Beavis, R. C. *Bioinformatics* **2004**, *20*, 1466−1467.
(21) Maclean, B.; Eng, J. K.; Beavis, R. C.; McIntosh, M. *Bioinformatics* **2006**.
(22) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2002**, *74*, 5383−5392.
(23) Rauch, A.; Bellew, M.; Eng, J.; Fitzgibbon, M.; Holzman, T.; Hussey, P.; Igra, M.; Maclean, B.; Lin, C. W.; Detter, A.; Fang, R.; Faca, V.; Gafken, P.; Zhang, H.; Whiteaker, J.; States, D.; Hanash, S.; Paulovich, A.; McIntosh, M. W. *J. Proteome Res.* **2006**, *5*, 112−121.
(24) Krokhin, O. V.; Craig, R.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. *Mol. Cell. Proteomics* **2004**, *3*, 908−919.
(25) Ketterlinus, R.; Hsieh, S. Y.; Teng, S. H.; Lee, H.; Pusch, W. *Biotechniques* **2005**, *Suppl*, 37−40.
(26) Shen, Y.; Tolic, N.; Masselon, C.; Pasa-Tolic, L.; Camp, D. G., 2nd; Hixson, K. K.; Zhao, R.; Anderson, G. A.; Smith, R. D. *Anal. Chem.* **2004**, *76*, 144−154.
(27) Shen, Y.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **2005**, *77*, 3090−3100.
(28) Wang, P.; Coram, M.; Tang, H.; Fitzgibbon, M. P.; Zhang, H.; Yi, E.; Aebersold, R.; McIntosh, M. *Biostatistics,* **2006**, in press.
(29) Anderson, L.; Hunter, C. L. *Mol. Cell. Proteomics* **2006**, *5*, 573−588.
(30) Bause, E. *Biochem. J.* **1983**, *209*, 331−336.
(31) Koomen, J. M.; Li, D.; Xiao, L. C.; Liu, T. C.; Coombes, K. R.; Abbruzzese, J.; Kobayashi, R. *J. Proteome Res.* **2005**, *4*, 972−981.
(32) Villanueva, J.; Shaffer, D. R.; Philip, J.; Chaparro, C. A.; Erdjument-Bromage, H.; Olshen, A. B.; Fleisher, M.; Lilja, H.; Brogi, E.; Boyd, J.; Sanchez-Carbayo, M.; Holland, E. C.; Cordon-Cardo, C.; Scher, H. I.; Tempst, P. *J. Clin. Invest.* **2006**, *116*, 271−284.

PR0604920