

Multi-site assessment of the precision and reproducibility of multiple reaction monitoring–based measurements of proteins in plasma

Terri A Addona¹, Susan E Abbatiello¹, Birgit Schilling², Steven J Skates³, D R Mani¹, David M Bunk⁴, Clifford H Spiegelman⁵, Lisa J Zimmerman⁶, Amy-Joan L Ham⁶, Hasmik Keshishian¹, Steven C Hall⁷, Simon Allen⁷, Ronald K Blackman^{1,18}, Christoph H Borchers⁸, Charles Buck⁹, Helene L Cardasis¹⁰, Michael P Cusack², Nathan G Dodder⁴, Bradford W Gibson², Jason M Held², Tara Hiltke¹¹, Angela Jackson⁸, Eric B Johansen⁷, Christopher R Kinsinger¹¹, Jing Li⁶, Mehdi Mesri¹¹, Thomas A Neubert¹⁰, Richard K Niles⁷, Trenton C Pulsipher³, David Ransohoff¹², Henry Rodriguez¹¹, Paul A Rudnick⁴, Derek Smith⁸, David L Tabb⁶, Tony J Tegeler¹³, Asokan M Variyath⁵, Lorenzo J Vega-Montoto⁵, Åsa Wahlander¹⁰, Sofia Waldemarson¹⁰, Mu Wang^{13,14}, Jeffrey R Whiteaker¹⁵, Lei Zhao¹⁵, N Leigh Anderson¹⁶, Susan J Fisher⁷, Daniel C Liebler⁶, Amanda G Paulovich¹⁵, Fred E Regnier⁹, Paul Tempst¹⁷ & Steven A Carr¹

Verification of candidate biomarkers relies upon specific, quantitative assays optimized for selective detection of target proteins, and is increasingly viewed as a critical step in the discovery pipeline that bridges unbiased biomarker discovery to preclinical validation. Although individual laboratories have demonstrated that multiple reaction monitoring (MRM) coupled with isotope dilution mass spectrometry can quantify candidate protein biomarkers in plasma, reproducibility and transferability of these assays between laboratories have not been demonstrated. We describe a multilaboratory study to assess reproducibility, recovery, linear dynamic range and limits of detection and quantification of multiplexed, MRM-based assays, conducted by NCI-CPTAC. Using common materials and standardized protocols, we demonstrate that these assays can be highly reproducible within and across laboratories and instrument platforms, and are sensitive to low $\mu\text{g/ml}$ protein concentrations in unfractionated plasma. We provide data and benchmarks against which individual laboratories can compare their performance and evaluate new technologies for biomarker verification in plasma.

Proteomic technologies based on mass spectrometry (MS) have emerged as preferred components of a strategy for discovery of diagnostic, prognostic and therapeutic protein biomarkers. Because of the stochastic sampling of proteomes in unbiased analyses and the associated high false-discovery rate, tens to hundreds of potential biomarkers are often reported in discovery studies. Those few that will ultimately show sufficient sensitivity and specificity for a given medical condition must thus be culled from lengthy lists of candidates—a particularly challenging aspect of the biomarker-development pipeline and currently its main limiting step. In this context, it is highly desirable to verify, by more targeted quantitative methods,

the levels of candidate biomarkers in body fluids, cells, tissues or organs from healthy individuals and affected patients in large enough sample numbers to confirm statistically relevant differences^{1,2}. Verification of novel biomarkers has relied primarily on the use of sensitive, specific, high-throughput immunoassays, whose development depends critically on the availability of suitable well-characterized antibodies. However, antibody reagents of sufficient specificity and sensitivity to assay novel protein biomarkers in plasma are generally not available. The high cost and long development time required to generate high-quality immunoassay reagents, as well as technical limitations in multiplexing immunoassays for panels of biomarkers,

¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ²Buck Institute for Age Research, Novato, California, USA. ³Biostatistics Center, Massachusetts General Hospital, Boston, Massachusetts, USA. ⁴Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland, USA. ⁵Department of Statistics, Texas A&M University, College Station, Texas, USA. ⁶Vanderbilt University, Nashville, Tennessee, USA. ⁷Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco, San Francisco, California, USA. ⁸University of Victoria-Genome BC Proteomics Centre, Victoria, British Columbia, Canada. ⁹Purdue University, West Lafayette, Indiana, USA. ¹⁰Kimmel Center for Biology and Medicine at the Skirball Institute and Department of Pharmacology, New York University School of Medicine, New York, New York, USA. ¹¹National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. ¹²University of North Carolina, Chapel Hill, Chapel Hill, North Carolina, USA. ¹³Monarch Life Sciences, Indianapolis, Indiana, USA. ¹⁴Indiana University School of Medicine, Indianapolis, Indiana, USA. ¹⁵Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ¹⁶Plasma Proteome Institute, Washington DC, USA. ¹⁷Memorial Sloan-Kettering Cancer Center, New York, New York, USA. ¹⁸Present address: Synta Pharmaceuticals, Lexington, Massachusetts, USA. Correspondence should be addressed to S.A.C. (scarr@broad.mit.edu).

Received 5 May; accepted 31 May; published online 28 June 2009; doi:10.1038/nbt.1546

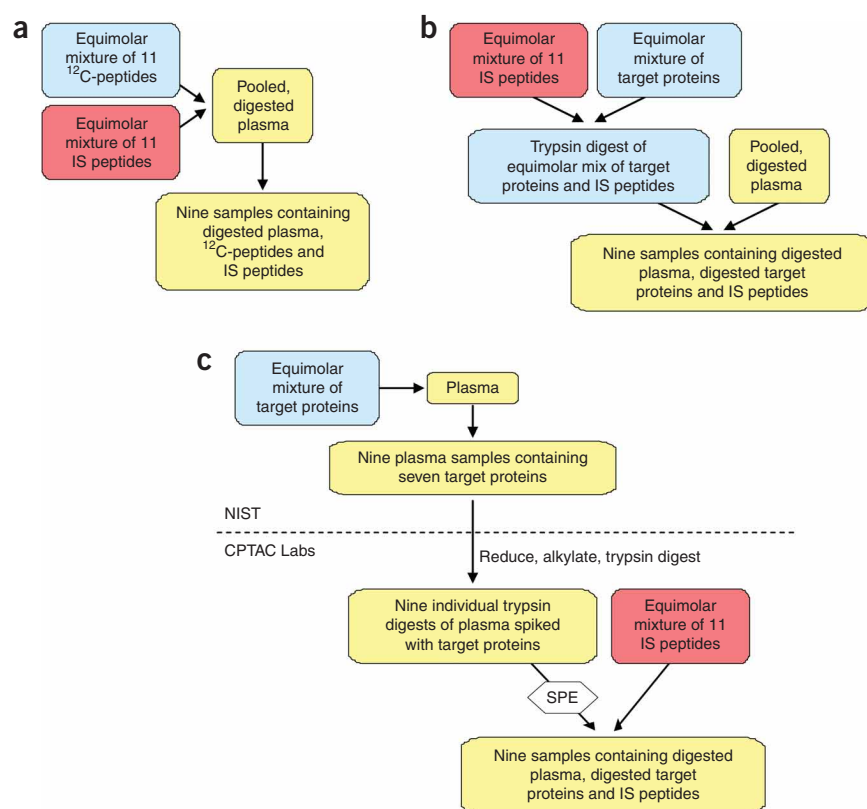


Figure 1 Sample preparation workflow for studies I, II and III. (a) Study I. Pooled, digested plasma was spiked with ^{12}C and $^{13}\text{C}/^{15}\text{N}$ peptides to generate a nine-point standard curve. (b) Study II. An equimolar mixture of the seven target proteins was digested separately and spiked with an equimolar mixture of IS peptides. The digest of target proteins plus IS peptides was added to pooled, digested plasma. A nine-point standard curve was prepared with pooled, digested plasma spiked with an equimolar mixture of IS peptides as the diluent. Study I and study II samples were prepared centrally at NIST. (c) Study III. Undiluted plasma was spiked with an equimolar mixture of the target proteins, then diluted with plasma to generate a nine-point standard curve. Three aliquots of these samples (prepared at NIST) were then shipped to the eight participating sites where reduction, alkylation, digestion and desalting were carried out before SID-MRM-MS analysis. IS, internal standard; SPE, solid phase extraction.

in the context of human plasma. The studies (I–III) sequentially introduced additional sources of variability in sample preparation and instrumental analyses, thereby enabling assessment of their impact on the quantitative measurements (Fig. 1 and Table 1). In studies I and II, samples were prepared centrally at the National Institute of Standards and

Technology (NIST) and then distributed to the laboratories for SID-MRM-MS analyses. Variability arising from digestion of the target proteins was bypassed in study I by spiking a common pool of reduced, alkylated and trypsin-digested plasma with 11 unlabeled signature peptides derived from the target proteins at nine different concentrations. In study II, seven target proteins were digested separately, mixed with a stock solution of labeled peptides and digested plasma, then diluted serially with a labeled peptide/digested plasma stock to generate the same nine concentrations. Study III, which encompassed nearly all potential sources of analytical variability normally encountered, most closely simulated an actual biomarker verification experiment. Specifically, we produced an equimolar mixture of the same seven proteins in undiluted plasma at the same nine concentrations. Then, aliquots were distributed to the eight sites where the samples were denatured, reduced, alkylated, digested and desalted according to a standard operating procedure (SOP, **Supplementary Methods**). Labeled internal standard peptides were added immediately before SID-MRM-MS analysis. In all three studies, four technical replicates were performed at each concentration; in study III, three independent process replicates (IIIa, IIIb and IIIc) assessed intralaboratory and interlaboratory variability.

is strong motivation to develop more straightforward quantitative approaches exploiting the sensitivity and molecular specificity of mass spectrometry. Recently, multiple reaction monitoring (MRM) coupled with stable isotope dilution (SID)-MS for direct quantification of proteins in cell lysates as well as human plasma and serum has been shown to have considerable promise^{3–10}. With SID-MRM-MS, up to tens of candidate proteins can be nearly simultaneously targeted and quantified in plasma by detecting ‘signature’ peptides, those that are diagnostic for each protein^{8,9}. These reports suggest that this technology may be suitable for use in preclinical studies to rapidly screen large numbers of candidate protein biomarkers in the hundreds of patient samples necessary for verification². Widespread acceptance and adoption of SID-MRM-MS methods are presently limited because the reproducibility and transferability of protein-based MRM assays across different instrument platforms and laboratories have yet to be demonstrated. To address this issue, the Clinical Proteomic Technology Assessment for Cancer network of the National Cancer Institute (NCI-CPTAC) evaluated intra- and interlaboratory analytical performance of SID-MRM-MS assays for quantifying seven target proteins added to human plasma. Our study demonstrates that targeted, quantitative and multiplexed MS-based assays can be rapidly configured and deployed in multiple laboratories to reproducibly measure proteins present at moderate to high abundance ($>2\ \mu\text{g}/\text{ml}$), with a linear dynamic range spanning three orders of magnitude, in nondepleted, nonfractionated plasma, the most complex of all biological matrices.

RESULTS

Study design

A series of interrelated studies was designed to assess the reproducibility and quantitative characteristics of MRM assays across the eight participating laboratories for measurement of peptides and proteins

The MRM assay configuration (including gradient development, selection of MRM analyte transitions for each signature peptide and general instrument settings) was performed at a single site using a nanoflow liquid chromatography (LC) (Eksigent NanoLC-2D) system coupled to a hybrid triple quadrupole/linear ion trap (AB/MDS Analytical Technologies 4000 QTRAP) mass spectrometer. These methods and parameters were transferred to all laboratories regardless of instrument platform to minimize variability arising from data acquisition (Online Methods and **Supplementary Methods**). All sites monitored three transitions per peptide, and precursor m/z values were consistent across all laboratories. Seven of the laboratories

Table 1 Target proteins and their signature peptides

Protein	Abbrev	Species	Signature peptide	MH+ (mono)	MRM transitions (m/z)			
					Q1	Q3		
Aprotinin	APR-AGL	Bovine	AGLCQT F YGG C R	1493.7	747.3	863.4	964.5	1092.5
Leptin	LEP-IND	Mouse	INDISHTQSV S A K	1407.3	469.9	590.8	647.8	728.4
Myoglobin	MYO-LFT	Horse	LFTGHPET L E K	1279.7	427.2	510.3	583.8	724.4
Myelin basic protein	MBP-HGF	Bovine	HGFLP R	732.4	366.7	391.3	538.3	595.4
Myelin basic protein	MBP-YLA	Bovine	YLASASTMDH A R	1328.6	443.5	491.2	526.8	823.4
Prostate-specific antigen	PSA-IVG	Human	IVGGWE C E K	1082.5	541.7	808.3	865.4	969.4
Prostate-specific antigen	PSA-LSE	Human	LSEPAELTDA V K	1280.7	640.8	783.4	854.5	951.2
Peroxidase	HRP-SSD	Horseradish	SSDLVALSGGHT F G K	1483.8	495.3	711.4	798.4	982.5
C-reactive protein	CRP-ESD	Human	ESDTSYV S L K	1136.6	568.8	617.4	704.4	805.4
C-reactive protein	CRP-GYS	Human	GYSIFS Y AT K	1144.6	572.8	724.4	837.5	924.5
C-reactive protein	CRP-YEV	Human	YEVQGEVFTKPQL W P	1826.9	914.0	1053.5	1181.6	1525.8

Preselected MRM transitions are listed with further details in **Supplementary Table 1**. Bold face amino acids are stable, isotopically labeled residues. Cysteines (underlined) are carbamidomethylated. Q1, Q3, first and third quadrupoles.

used 4000 QTRAP mass spectrometer instruments; the eighth site used a ThermoFisher TSQ Quantum Ultra triple quadrupole. Each laboratory tested and, if necessary, further optimized instrument parameters to maximize MS responses for the selected fragment ions on individual instruments. For the TSQ Quantum Ultra instrument, not all preselected transitions were ideal for achieving maximum sensitivity. For this subset of peptides, the site selected and optimized a substitute MRM transition for the signature peptide and its corresponding isotopically labeled analog (**Supplementary Table 1b**). Peptide YEVQGEVFTKPQLWP from C-reactive protein (CRP)-YEV did not ionize well and was detected with very low signals in the tuning mixtures or in the QC samples circulated to each site. Although MRM transitions for this peptide were included for data acquisition, subsequent data were not analyzed.

Intralaboratory reproducibility and precision of MRM assays

Intralaboratory variability and reproducibility in studies I–III were evaluated by comparing the measured concentrations to the actual concentrations across the range of spiked-in analytes and determining the coefficient of variation (CV) for these quantitative measurements. **Figure 2a** shows measured log concentration (y axis) versus theoretical (spiked-in) concentration (x axis) for the SSDLVALSGGHTFGK peptide derived from horseradish peroxidase (HRP-SSD; for all other peptides, **Supplementary Fig. 1**). Data for each site are color-coded, and organized by study and concentration. A linear trend is observed in the measured concentrations for studies I–III as spiked-in analytes increase across the concentration range. However, measured concentrations decrease as laboratories progress from study I to II to III. This trend is a result of apparent peptide loss from incomplete digestion of HRP protein and variability in sample handling at each site, as study complexity was increased (**Fig. 1**). Study I represents the optimum assay performance, as synthetic peptides (not proteins) were used as analytes. Protein digestion in study II (at a central location in the absence of plasma) and study III (at individual sites and in the presence of plasma) introduces potential sources of sample loss that decrease analyte recovery and reduce measured concentrations for studies II and III.

Intralaboratory CVs for studies I and II constitute a measure of the technical variation due to instrument and data acquisition, as all sample preparation was performed centrally. The intralaboratory CVs at each analyte concentration point are shown in **Figure 2b** for the HRP-SSD peptide with color coded markers representing individual

laboratories. Equivalent figures for all other peptides are shown in **Supplementary Figures 2** and **3**. **Table 2** summarizes the range of median intralaboratory CVs observed across studies I, II and III, and **Supplementary Table 2a–c** shows the intralaboratory CVs calculated for each analyte at each of the nine final concentrations in plasma. Intralaboratory CVs are color coded in **Supplementary Table 2a–c** to facilitate visualization of the increasing variability from studies I–III. For all ten peptides in study I, median intralaboratory CVs were $\leq 15\%$ across the concentration range (**Supplementary Fig. 1** and **Supplementary Table 2a**). The median intralaboratory CVs for study II were very similar to those found in study I, with most intralaboratory CVs $\leq 15\%$ across the concentration range (**Supplementary Fig. 1** and **Supplementary Table 2b**). Finally, the intralaboratory CVs for study III were a measure of variation of the sample processing across replicates in addition to the technical variation of data acquisition. Increased variability is observed across the laboratories as individual sites were responsible for all sample handling and preparation (**Fig. 2b**). Although the intralaboratory CVs were elevated relative to studies I and II, $>60\%$ of the median intralaboratory CVs were still $\leq 25\%$ across all concentrations, demonstrating very good reproducibility for sample processing (**Supplementary Fig. 1** and **Supplementary Table 2c**).

Interlaboratory reproducibility and precision of MRM assays

The interlaboratory reproducibility and precision of the quantitative measurements was evaluated by calculating the CV of the quadruplicate analyses at each of the nine final analyte concentrations in plasma. The median interlaboratory CVs for HRP-SSD across studies I, II and III for the entire concentration range of 1–500 fmol/ μ l were predominantly $\leq 15\%$ for this peptide in all three studies (**Fig. 2b**). As expected, interlaboratory CVs decreased as the concentration of spiked-in analyte increased to the upper range (**Fig. 2b**). However, even at lower analyte concentrations, the precision of the quantitative measurements across sites was very good. **Table 2** summarizes the interlaboratory CVs at the 2.92 fmol/ μ l concentration for all peptides. This concentration is at or near the limit of quantification (LOQ) for most analytes in diluted plasma, except the two peptides derived from CRP (see below). Box plots of median interlaboratory CVs for all other peptides are shown in **Supplementary Figure 2** (comparison of CVs across studies I, II and III) and **Supplementary Figure 3** (comparison of CVs across process replicates for studies IIIa, IIIb and IIIc).

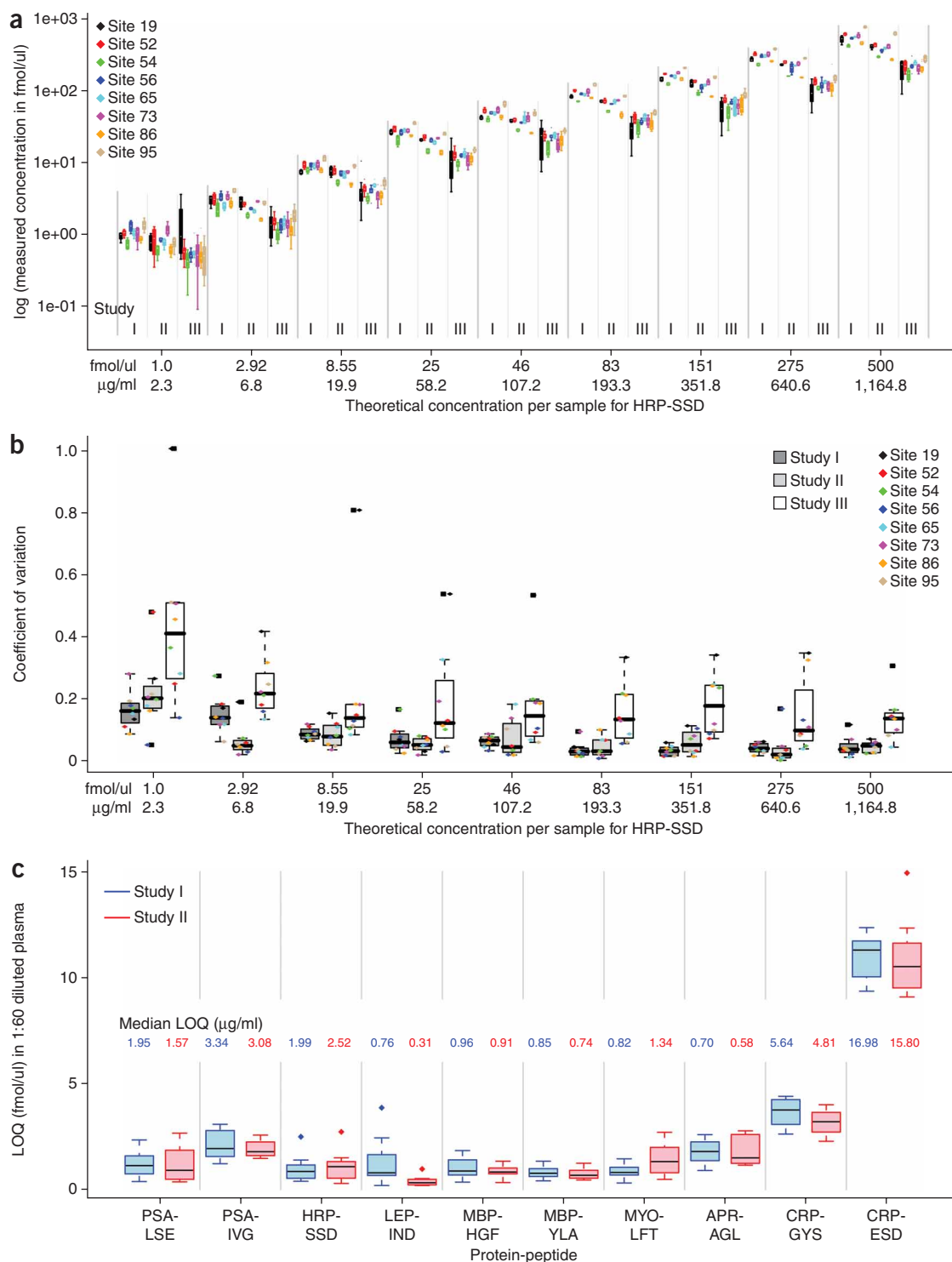


Figure 2 Box plots of variation in MRM quantitative measurements, interlaboratory CV and LOQ. **(a)** Intralaboratory assay CV. Box plots showing measured log concentration (y axis) versus theoretical (spiked-in) concentration (x axis) for HRP-SSD across the entire concentration range in diluted plasma. Protein concentration in $\mu\text{g/ml}$ is μg protein equivalent in 1 ml undiluted plasma. The box plots for studies I and II are based on four replicate measurements, whereas those for study III summarize 12 measurements (four each from III a, b and c). Each of the eight sites was assigned a random numerical code (19, 52, 54, 56, 65, 73, 86, 95) for anonymization. **(b)** Interlaboratory assay CV. Values are shown for studies I–III for the entire range of HRP-SSD final analyte concentrations in plasma. Within each box plot, actual intralaboratory CV values for individual laboratories are shown with color-coded markers. The CV values are calculated based on the single best performing transition (lowest combined CV) across studies I and II. This same transition is also used for study III. **(c)** Interlaboratory assay LOQ. Values determined in studies I and II for the peptides indicated (see **Table 1** for protein-peptide pair abbreviations). The inset values display the conversion of median LOQ to $\mu\text{g/ml}$ (μg protein equivalent per 1 ml undiluted plasma) for each peptide. All measurements were made in 60-fold diluted plasma. Median is shown as a heavy horizontal line in all box plots. The box spans the interquartile range (IQR), with the whiskers extending to $1.5 \times \text{IQR}$. Values $> 1.5 \times \text{IQR}$ are deemed outliers, and shown as separate points.

Table 2 Summary of results for studies I, II, and III

Signature Peptide	Study I				Study II				Study III ^a			
	Interlaboratory CV ^b	Intralaboratory CV ^c	Linear slope	Recov. (%) ^d	Interlaboratory CV ^b	Intralaboratory CV ^c	Linear slope	% Recov. ^d	Interlaboratory CV ^b	Intralaboratory CV ^c	Linear slope	Recov. (%) ^d
APR-AGL	9.2%	3.9–11.2%	1.157	114.5	13.1%	2.0–7.8%	0.575	57.5	13.7%	7.3–45.2%	0.738	79.4
CRP-ESD	5.9%	2.2–5.9%	1.124	118.4	10.5%	3.1–8.4%	0.573	61.4	16.7%	8.5–18.1%	0.439	48.9
CRP-GYS	5.4%	1.4–10.2%	1.324	140.5	5.6%	1.2–6.4%	0.546	56.0	18.5%	6.6–35.0%	0.159	18.5
HRP-SSD	14.1%	4.0–8.9%	1.198	120.4	5.5%	4.6–7.3%	0.794	82.3	21.9%	8.4–21.4%	0.430	45.7
LEP-IND	12.5%	2.9–10.3%	1.163	119.1	29.5%	2.6–15.3%	0.152	14.9	50.4%	11.7–54.9%	0.242	25.6
MBP-HGF	4.3%	1.7–6.3%	1.161	118.6	9.3%	1.5–7.8%	0.758	77.3	21.8%	7.4–32.8%	0.238	23.8
MBP-YLA	5.1%	2.1–9.3%	1.275	130.3	4.1%	1.5–14.1%	0.806	83.8	N.M.	N.M.	N.M.	<1.0
MYO-LFT	4.9%	1.6–5.7%	1.518	154.4	3.8%	2.0–6.3%	1.012	101.3	23.1%	8.9–21.6%	0.504	60.4
PSA-IVG	6.9%	1.3–14.7%	1.658	165.4	5.5%	2.0–11.2%	0.848	81.9	17.2%	7.9–20.3%	0.587	58.0
PSA-LSE	8.9%	1.2–6.9%	1.098	111.4	5.3%	2.0–4.6%	1.524	151.3	10.3%	7.6–13.7%	0.918	92.7

^aCombined results for process replicates a, b, c for each peptide across sites for interlaboratory CV, intralaboratory CV, linear slope and percent recovery. ^bInterlaboratory CV was calculated from all replicates for each peptide using a single transition. The interlaboratory CV represented here is the median value across all sites for each peptide by study at the 2.92 fmol/μl concentration point. This concentration is at or near the LOQ for all peptides except those derived from CRP. ^cIntralaboratory CV was calculated from all replicates for each peptide using a single transition. The range of the median intralaboratory CV (over all concentrations) is reported here. Outlier laboratories (with CVs > 1.5 times the interquartile range) have been excluded; in all three studies, the majority of the sites (seven or greater) are included in the intralaboratory CV range. ^dPercent recovery was determined from the mid-concentration point, 46 fmol/μl. The value shown is the average percent recovery across the eight sites using the same single transition as in CV and LOQ calculations. Recov., recovered; N.M., not measured.

For study I, the interlaboratory CVs ranged from 4.3 to 14.1% at 2.92 fmol/μl, with eight of ten peptides in excellent agreement with values ≤10%. Because the interlaboratory CVs decreased at higher analyte concentrations, the median interlaboratory CVs across the entire concentration range was ≤5% (**Supplementary Fig. 2** and **Supplementary Table 2a**). These results demonstrate excellent precision and reproducibility of the MRM assays for the signature peptides between laboratories when the major analytical variable is limited to the LC-MS system. Study II introduced new sources of variability attributable to sample loss during reduction, alkylation and trypsin digestion of the target proteins and desalting of the resulting peptide mixtures (**Fig. 1** and Online Methods). The median interlaboratory CVs at 2.92 fmol/μl for study II ranged from 3.8% to 30% for all peptides, with nine of ten peptides having interlaboratory CVs ≤15%. Median interlaboratory CVs were predominantly ≤10% over the entire concentration range for study II (**Supplementary Fig. 2** and **Supplementary Table 2b**), indicating that reproducibility of the assay across sites was not hampered by decreased recovery of target peptides. Finally, study III introduced the potential for the largest variability as each of the laboratories reduced, alkylated and trypsin digested the target proteins in plasma and desalted the subsequent peptide mixtures in three process replicates. Despite these additional sources of variability, average interlaboratory CVs for study III across process replicates IIIa, IIIb and IIIc ranged from 10.3–50% at 2.92 fmol/μl for nine of ten peptides (**Table 2**). Eight peptides had interlaboratory CVs ≤25%. Across the concentration range, the median interlaboratory CV was predominantly ≤20% (**Supplementary Fig. 2** and **Supplementary Table 2c**).

Limits of detection and quantification

For studies I and II, inter- and intralaboratory measurement reproducibility of the ten signature peptides was determined at their estimated limits of detection (LOD) (**Supplementary Fig. 4**) and LOQ (**Fig. 2c**). The LOQ values represented in the box plot are based on the amount of peptide (in fmol) detected in plasma that was diluted 60-fold to a final protein concentration of 1 μg/μl for SID-MRM-MS analysis. The corresponding LOQ values for measurement of the proteins in undiluted plasma (in μg/ml) were also calculated (**Fig. 2c**). LOD and LOQ values calculated for each peptide at each site are shown in **Supplementary Table 3**.

The reproducibility of the LOQ estimations across sites was very good. For example, in study I, eight of ten peptides had median LOQ values between 0.66 and 2.0 fmol/μl when peptides were added into 1:60 diluted plasma (equivalent to a range of 0.70–3.34 μg/ml protein in plasma; **Fig. 2c**). The remaining two CRP peptides were detected at endogenous levels in the blank and 0 fmol/μl spiked plasma samples. A commercial enzyme-linked immunosorbent assay (ELISA) was performed on the plasma stock and yielded a concentration of 6 μg/ml of this protein (data not shown), which is equivalent to 4 fmol/μl of CRP in the diluted plasma. The LOQ values obtained in study II, which were similar to those obtained in study I, ranged between 0.31 and 1.8 fmol/μl for the same eight of ten peptides. The LOD/LOQ values for studies I and II were similar in magnitude for a majority of the signature peptides and showed acceptable variation across all eight laboratories.

Reproducibility of linear response and peptide recovery

Figure 3 shows a compilation of response curves (study II) obtained at the eight sites and plotted on a linear-linear scale for the HRP-SSD peptide. Response curves are plots of experimentally determined concentrations versus theoretical concentrations of the target analyte, and provide useful visual representations of reproducibility and linearity. Quadruplicate replicates are shown at every concentration for all three MRM transitions. Interlaboratory reproducibility of linear responses and quantitative measurements across all laboratories and all three studies was, in general, very good (**Table 2**). The fitted slopes presented in **Table 2** demonstrate the consistency in the linear response with a change in actual peptide (study I) or protein (studies II–III) concentration across the measurements made in each laboratory, and are also an estimation of peptide recovery. A slope of 1.0 is equal to the theoretical slope in which measured concentration is proportional to analyte concentration and recovery is equal to 100%. Slopes <1 indicate <100% recovery, whereas slopes >1 indicate >100% recovery (the latter likely a result of errors in the initial concentrations of the peptide or protein stock solutions). For the representative peptide, HRP-SSD, the average slope in study I was 1.2 with an interlaboratory CV of 15.6% (**Table 2** and **Supplementary Table 4a**), showing excellent reproducibility between sites and highly consistent linear responses across laboratories and instrument platforms as indicated by the slopes being close to the

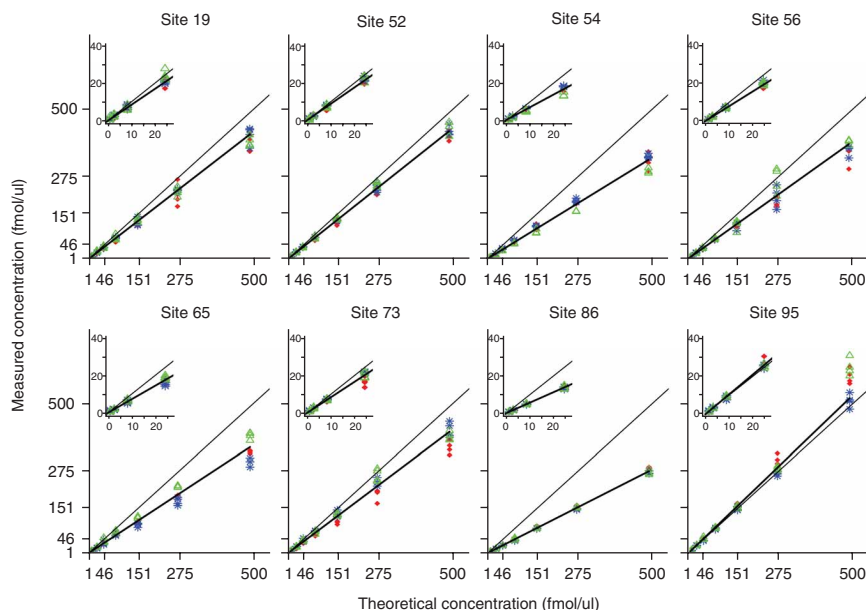


Figure 3 Interlaboratory reproducibility of linear calibration curve slopes for study II. The eight plots display the concentration curves for the detection of HRP-SSD in study II across all laboratories. Each of the eight sites was assigned a random numerical code (19, 52, 54, 56, 65, 73, 86, 95) for anonymization. Comparison of the plots demonstrates good linearity, with the slopes falling close to the diagonal, black line (theoretical slope = 1), and good agreement between the three transitions at each concentration point. Four replicate measurements are represented at each concentration point. Analyte transitions: red diamond, transition 1, (m/z 492.6 $\rightarrow$ 703.4); blue asterisk, transition 2, (m/z 492.6 $\rightarrow$ 790.4); green triangle, transition 3, (m/z 492.6 $\rightarrow$ 974.5). In some cases, the data points overlay such that transition 1 is not visible. Inset plots show more detail of lower end of the concentration range. The mean slope calculation across all laboratories in this example is 0.794 with an interlaboratory CV of 18.7%. Final concentrations of heavy and light peptides and added proteins were adjusted according to the gravimetric measurements described in **Supplementary Table 6a–f**.

theoretical line. As an estimation of the average percent recovery across the concentration range, the average slope for the HRP-SSD peptide agrees well with the calculation of percent recovery determined at the mid-concentration point of the response curve (46 fmol/ul; **Table 2**).

Response curves for all other peptides and proteins generated by each laboratory in all three studies are plotted on the linear-linear scale with scale-expansion insets to facilitate visualization of the lower concentration range (**Supplementary Fig. 5**). A weighted robust linear regression on the linear-linear scale was used to determine slope and percent recovery. In addition, the response curves are plotted on the log-log scale (**Supplementary Appendix**) without regression lines to facilitate data visualization. Individual parameters for slope, y intercept and their associated standard errors for each peptide across all sites are shown in **Supplementary Tables 4a–e**. Altogether, peptide responses in study I had an average slope ranging from 1.1 to 1.6 with an interlaboratory CV $\leq 10\%$ for most of the peptides (**Table 2** and **Supplementary Tables 4a**). The average slope value was more variable in study II, with a range of 0.15 to 1.5 across all peptides. Interlaboratory CV for slope in study II was $\leq 15\%$ for nine of ten peptides (**Supplementary Table 4b**). Study III exhibited the lowest average slope values, which ranged from 0.16 to 0.92 for nine of ten peptides, and interlaboratory CVs for slope were $\leq 25\%$ for the majority of peptides across the process replicates (**Supplementary Table 4c–e**). One peptide, MBP-YLA, was not detected by any site in any process replicate of study III. Overall, the responses were reproducible as indicated by the low interlaboratory CVs, and the measurements of the three transitions were highly uniform such that the replicates often overlaid at each concentration (**Fig. 3** and **Supplementary Fig. 5**).

Because the slope is an estimation of percentage recovery, the decrease and variability in the slopes of the response curves observed across these studies (**Supplementary Fig. 6**) correlate with the increasing level of sequential experimental complexity, from the introduction of protein digestion in study II and protein digestion in the presence of plasma in study III (**Fig. 1**). Again, the average slopes for all peptides agree well with the calculation of percent recovery at the mid-point of the concentration range (**Table 2**). For

study I and two of the ten peptides in study II, recovery $\geq 100\%$ was observed for many peptides. This could most likely be attributed to the effect of errors in quantification of the protein or peptide stock concentrations by amino acid analysis, and inaccuracies associated with sample preparation, such as pipetting and freeze-thawing. In study III, six of the nine peptides detected had percent recoveries $\geq 40\%$, which is within an acceptable range for verification assays^{2,9}. Four peptides (CRP-GYS, LEP-IND, MBP-HGF and MBP-YLA) had recoveries $\leq 25\%$, and would not be considered useable for verification or clinical validation assay purposes. No significant differences in peptide recovery were observed across the concentration range or between studies II and III (**Supplementary Table 5** for two representative examples). Although $<100\%$ recovery of the target peptides limits the sensitivity of the assays, these results show very good reproducibility for recovery of most peptides and demonstrate the large role sample handling has in the variability of peptide recovery.

Common sources of variance and their detection

Although most of the signature peptides exhibited excellent reproducibility within and between laboratories (**Supplementary Fig. 5**), deviations from the trend lines were observed for some peptides at one or more sites. Typical problems that can arise in developing and applying MRM assays to quantify proteins in plasma are illustrated in **Figure 4**. The most common problem related to the appearance of 'outliers' was interference in one or more of the fragment-ion transitions monitored for either the light ($^{12}\text{C}/^{14}\text{N}$) peptides or heavy ($^{13}\text{C}/^{15}\text{N}$)-labeled internal standard peptides. **Figure 4a,b** illustrates interferences in transition 1 and 2 of the light peptides for MBP-HGF and MYO-LFT, respectively, at two analysis laboratories. In both cases, the relative ratios of the transitions were altered from those observed in the absence of plasma during assay configuration, resulting in considerable deviation from linearity for the respective product ions. Monitoring multiple transitions for each peptide, as done in our study, enables reliable quantification, which is accomplished by using the other unaffected transitions. In the case of CRP-ESD (**Fig. 4c**), obvious and highly consistent deviation from linearity was observed for all three transitions monitored at the lower end of the response curves. This flattening of the curves was due to the presence of

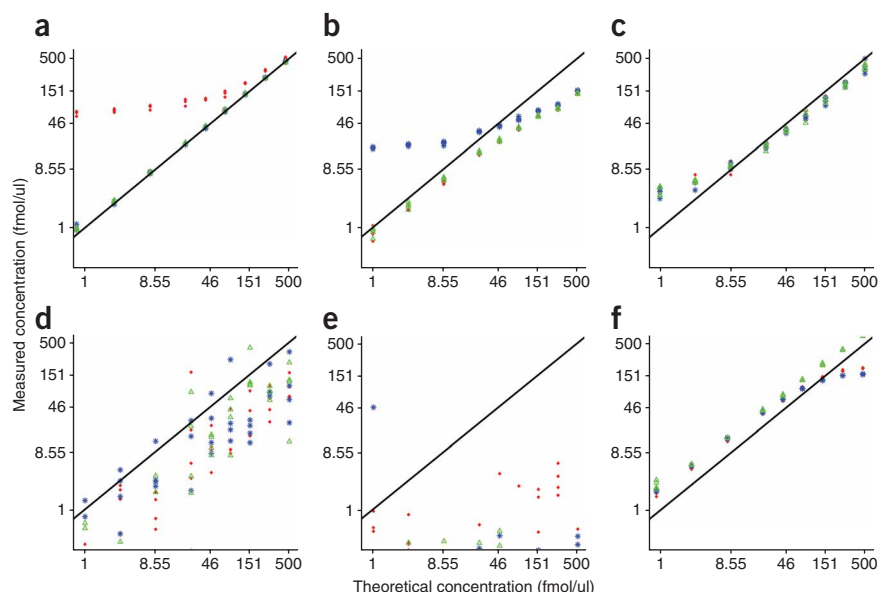


Figure 4 Response curves representing deviations from the trend line. Red diamond, transition 1; blue asterisk, transition 2; green triangle, transition 3. **(a)** Study I, site 52, MBP-HGF: interference in transition 1 of the analyte. **(b)** Study IIIb, site 95, MYO-LFT: interference in transition 2 of the analyte, which was also observed in study I, II and IIIa for this laboratory. **(c)** Study II, site 86, CRP-ESD: endogenous protein level increased the estimated protein concentration at the low end of the concentration range of spiked-in proteins, resulting in flattening of slope. **(d)** Study IIIa, site 56, LEP-IND: unstable electrospray conditions resulted in a substantial increase in interlaboratory CV to 99%. **(e)** Study IIIa, site 19, MBP-YLA: no detection of MBP-YLA peptide at any site. **(f)** Study I, site 86, PSA-IVG: saturation at highest two concentrations. Site codes are identical to those given in **Figures 2 and 3**.

4–14%, 4–13% and 10–23% interlaboratory CVs at or near the estimated LOQ for study I, II and III, respectively. Intralaboratory CVs

were predominantly <15% and <25% at the identical concentration for studies I/II and III, respectively (**Supplementary Table 2**). Although the current assay performance under real biomarker conditions (study III) is below that generally stated for clinical assays (typically <10–15%), the performance achieved is sufficient for the verification of candidate biomarkers² present at more than ~2–6 µg/ml in plasma, with a linear dynamic range spanning three orders of magnitude. In all cases, interlaboratory and intralaboratory CVs improved with increasing analyte concentration. Such modest differences between interlaboratory and intralaboratory CVs underscore the excellent agreement between the eight participating laboratories. Likewise, the progressive increases in CVs from studies I to III indicate convincingly that sample preparation contributes more to assay variability than instrumental variability, further highlighting the data quality obtainable from SID-MRM-MS. Although most important parameters were governed by detailed SOPs, the transfer of MRM assays across LC-MS platforms did require optimization of the transitions being monitored to compensate for differing instrument-specific ion source and collision-induced dissociation parameters, and to ensure that each platform achieved optimum sensitivity (**Supplementary Tables 1a–e**). Despite these variations concerning a small number of analyte peptides, interlaboratory variability and specificity of the assay were not affected (**Table 2**).

Differences emerged in assay performance for different peptides. Most peptides performed well at all eight sites, whereas a few exhibited variable or poor behavior. This result highlights the dependence of MRM assay performance in plasma on specific properties of the peptides selected as surrogates for the target proteins. Ideally the final selection of signature peptides for SID-MRM-MS biomarker assays should be based on multisite studies so as to ensure the most robust performance.

The most frequent cause of poor peptide performance was the presence of interference from the background plasma digest matrix, in either the analyte or internal-standard channels, which altered the ratios of these transitions. Monitoring a minimum of three transitions per analyte is critical in maintaining assay selectivity and recognizing such interferences when they occur. Most participating sites observed interferences in one or more peptides over the course of the three studies. In the case of CRP, we were able to establish that the flattening

endogenous levels of the protein within the measurable range of the MRM assays. We confirmed the level of CRP present in the plasma by ELISA. Other issues, such as unstable electrospray conditions, lack of recovery during sample processing and saturation of the MS detector were also observed and gave rise to recognizable patterns of misbehavior (**Fig. 4d–f**). Instability of the LC system and deterioration of the LC column are also common problems that are readily recognized. If not corrected, they can cause large shifts in peptide retention time and chromatographic peak broadening or tailing, particularly for early-eluting hydrophilic species, resulting in decreased reproducibility for peptide detection and quantification.

DISCUSSION

Targeted MRM assays have been used very successfully for quantifying small molecules (e.g., hormones, drugs and their metabolites) in pharmaceutical research and in clinical laboratories in applications such as screening newborns for disease¹¹. More recently, the merits of SID-MRM-MS for quantifying peptides derived from proteins in plasma have been demonstrated in several laboratories^{4–9,12}. These studies have, however, only addressed assay performance at a single laboratory, and thus were not able to demonstrate the multisite robustness needed in large-scale biomarker research and ultimately in preclinical and clinical applications. The main purpose of this study was to provide such a demonstration by performing an assessment of the analytical characteristics of a multiplexed, SID-MRM-MS assay across eight laboratories using seven target proteins with which to spike human plasma. A three-tiered experimental protocol was used that progressively introduced sample preparation variables likely to affect inter- and intralaboratory reproducibility, transferability, precision and sensitivity. Our results demonstrate that reproducible, quantitative measurements of proteins in plasma can be made by SID-MRM-MS in multiple laboratories using different instrument platforms through use of standardized protocols for sample preparation, data acquisition and data analysis. The robustness of such a targeted assay approach compensates for the greater variability in protein measurements inherent in shotgun ('discovery' proteomics) methods^{13,14}, enabling the development of an effective biomarker pipeline¹.

Reproducibility and precision of the quantitative measurements for nine of ten peptides tested across eight laboratories ranged from

of the response curves was due to the presence of endogenous levels of CRP as all three transitions monitored were affected equally and the expected ratios of the transition-ion abundances to one another were maintained. Other interferences arose from problems with chromatography (e.g., large peak widths, shifting retention times, or early elution and consequent sensitivity to intermittent or unstable electrospray conditions), which can be addressed by further refinement of protocols, particularly in LC operation and data acquisition.

Recovery of signature peptides generally decreased from study I to III, as proteolytic digestion and subsequent sample handling, such as desalting, were introduced into the experimental workflow. Digestion efficiency of proteins in the plasma matrix has only recently begun to be studied¹⁵. If a signature peptide is not detected in an MRM assay, it is often unclear if this is because of (i) losses from sample handling, such as fractionation or desalting, (ii) poor enzymatic digestion, (iii) concentration below LOD, (iv) post-translational modification such as glycosylation and phosphorylation, (v) artifactual modifications to reactive amino acids, such as oxidation or carbamylation, or (vi) some combination thereof. The effect of decreasing control of sample preparation was reflected in the increased variability and lower peptide recoveries for a majority of peptides as sites progressed from study II to III (Table 2 and Supplementary Fig. 6). In study III, one peptide was not recovered in any process replicate performed at all participating laboratories, and four peptides had <25% recovery (Table 2 and Supplementary Fig. 6). Addition of labeled internal standard (IS) peptides at an early stage in sample processing (e.g., during enzymatic digestion) could help to account for peptide loss. However, lower recovery of signature peptides does not impede the use of these assays for verification where the goal is to precisely define the relative difference in abundance for candidate proteins between cases and controls rather than to determine the absolute concentration of each protein. Absent a general method ensuring stoichiometric digestion, absolute concentration measurements would likely require addition of isotopically labeled, recombinant protein standards at the start of sample processing.

The purpose of the present study was not to define the ultimate sensitivity possible for proteins by SID-MRM-MS, but rather to evaluate the transferability and robustness of the technology within and between laboratories. For this first study, we made no attempt to reduce the complexity of the plasma matrix by either depletion of abundant proteins or fractionation. The sensitivity of protein quantification by SID-MRM-MS in plasma is severely limited by the complexity and 10¹¹ dynamic range of protein abundances in blood, and the susceptibility to interference from other peptides and their fragment ions is greatest in this matrix¹⁶. Typical LODs and LOQs observed in prior studies of unfractionated plasma are in the high 100s of ng/ml to low µg/ml range of target protein^{6,8,17}. Results described here are consistent with these reports across sites and instrument platforms (Fig. 2c and Supplementary Fig. 4). Although emphasis is often placed on discovery and verification of low-abundance candidate biomarkers (≤ng/ml levels in serum), high-abundance serum proteins, such as CRP, transferrin, complement components, immunoglobulin classes and lipoproteins, are clinically relevant markers of disease and their levels in blood make them directly accessible by SID-MRM-MS using the approaches described here. The LODs and LOQs of MS-based assays have been extended into the low ng/ml range in plasma by using immunoaffinity depletion of high-abundance proteins, limited protein or peptide fractionation, or immunoaffinity enrichment at the protein or peptide level before SID-MRM-MS^{9,17–24}. The additional processing steps used are likely to introduce new sources of experimental variation that will

have to be assessed in interlaboratory studies similar to those described here. Nevertheless, the assay performance reported in the present studies, measured at maximum levels of interfering high-abundance peptides in unfractionated plasma digests, suggests that similar or better intra- and interlaboratory assay performance may be achievable for quantitative, multiplexed measurement of proteins in the low ng/ml range in plasma by MS.

Our study demonstrates that targeted, quantitative and multiplexed MS-based assays can be rapidly configured and deployed in multiple laboratories to yield robust and reproducible assays for proteins down to low µg/ml levels in the context of unfractionated plasma. This is a critical first step toward potential widespread implementation of SID-MRM-MS assays for verification of novel protein biomarker candidates. The SID-MRM-MS technology has the potential to become the critical filter used to assess candidate biomarker performance in a sufficient number of patient samples before committing the very substantial time and resources required to create clinical-grade immunoassays. The performance required of such assays² is not as stringent as that currently required for US Food and Drug Administration–approved clinical assays²⁵. Beyond candidate verification, SID-MRM-MS assays may eventually have potential to replace certain clinical immunoassays, especially in cases where interferences are known to exist²³ or multiplex measurements are needed. By detecting a structural component of the protein, the signature peptide, with near-absolute structural specificity, SID-MRM-MS should avoid inter-assay differences that occur when different immunoassays for the same protein detect distinct, potentially labile epitopes. Furthermore, the simplicity of producing and characterizing peptide-based reference materials for SID-MRM-MS could help overcome well-known problems with ELISA assay standardization, which lead to varying results across multiple clinical laboratories^{26,27}.

The methods, reagents and multilaboratory data sets presented here should facilitate testing and implementation of MRM-based multiplex assays for quantifying target proteins in plasma by the proteomics community. Our results should foster greater acceptance by the clinical community of SID-MRM-MS technology as a generally applicable approach to verify candidate biomarkers in large clinical sample sets, and thus provide a critical component for a systematic biomarker-development pipeline.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturebiotechnology/>.

Data accession. A password-protected website was developed to manage the large number of data files generated for the described interlaboratory studies. This website, hosted at NIST, was designed as a portal used by the teams for initiating uploads and downloads of large data files. The data transfers were performed using Tranche (<http://trancheproject.org/>) an open source, secure peer-to-peer file-sharing tool. A customized user interface employed by the participating laboratories was developed and added to the Tranche code base. This tool allowed the website and database to communicate tracking information with Tranche by employing custom URLs. The Tranche hash (a unique data identifier) and pass-phrase, for each website, was automatically recorded into the website's database when file uploading was complete. These stored links allow subsequent retrieval of data files using the Tranche download tool. The Tranche hashes and pass-phrases provide a simple and portable mechanism to access data sets and can be easily associated with supporting annotation. The data associated with this manuscript may be downloaded from the ProteomeCommons.org Tranche network using the following hash:

CKpfN0bl2ULLwCaIovXn/spuw4rYfjF6H/L+/6sHAKGzCsj4fzTD0RauJjAwf9baB8tI36HQ0izji2tupYAPM29P2cAAAAAAT0iw==. The hash may be used to show exactly what files were published as part of this manuscript's data set, and the hash may also be used to check that the data have not changed since publication. Accessible information includes all raw data files, all processed data export files, 4000 QTRAP MultiQuant results files, as well as detailed data submission sheets and file annotation legends for studies I–III from the eight participating laboratories.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This work was supported by grants from the National Cancer Institute (NCI) (U24 CA126476, U24 126477, U24 126480, U24 CA126485, and U24 126479), part of NCI Clinical Proteomic Technologies for Cancer initiative. A component of this initiative is the Clinical Proteomic Technology Assessment for Cancer (CPTAC) Network and teams, which include the Broad Institute of MIT and Harvard (with the Fred Hutchinson Cancer Research Center, Massachusetts General Hospital, the University of North Carolina at Chapel Hill, the University of Victoria and the Plasma Proteome Institute), Memorial Sloan-Kettering Cancer Center (with the Skirball Institute at New York University), Purdue University (with Monarch Life Sciences, Indiana University, Indiana University-Purdue University Indianapolis and the Hoosier Oncology Group), University of California, San Francisco (with the Buck Institute for Age Research, Lawrence Berkeley National Laboratory, the University of British Columbia and the University of Texas M.D. Anderson Cancer Center) and Vanderbilt University School of Medicine (with the University of Texas M.D. Anderson Cancer Center, the University of Washington and the University of Arizona). A full listing of the CPTAC Team Network can be found at <http://proteomics.cancer.gov/programs/CPTAC/networkmembership>.

AUTHOR CONTRIBUTIONS

The CPTAC Network contributed collectively to this study. The following CPTAC Network investigators contributed significant intellectual contributions to work described in this paper.

S.E.A., T.A., N.L.A., D.M.B., S.C.H., A.-J.L.H., H.K., D.R., B.S., S.J.S., L.J.Z. and S.A.C. contributed to study design and SOP development. D.M.B. and N.G.D. prepared and shipped samples. S.E.A., T.A., S.A., H.L.C., J.M.H., A.J., E.B.J., H.K., D.S., T.J.T., J.R.W., A.W., S.W., L.Z., and L.J.Z. contributed to generation of data. M.P.C., J.L., D.R.M., R.K.N., S.J.S., T.C.P., P.A.R., C.H.S., D.L.T., A.M.V., and L.J.V.-M. contributed to bioinformatics and statistical analysis. S.E.A., T.A., H.K., D.R.M., S.J.S. and L.J.Z. centrally reviewed data. S.E.A., T.A., N.L.A., S.A.C., S.J.F., S.C.H., A.-J.L.H., H.K., D.R.M., B.S., S.J.S., and L.J.Z. wrote and prepared the manuscript. R.K.B., C.B., C.H.B., S.A.C., S.J.F., B.W.G., T.H., C.R.K., D.C.L., M.M., T.A.N., A.G.P., H.R., F.E.R., P.T., and M.W. contributed to experimental design. S.C.H. chaired the CPTAC Experimental Design and Statistics Verification Studies Working Group that designed interlaboratory studies and generated data.

Published online at <http://www.nature.com/naturebiotechnology/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Rifai, N., Gillette, M.A. & Carr, S.A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24**, 971–983 (2006).
- Paulovich, A.G., Whiteaker, J.R., Hoofnagle, A.N. & Wang, P. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteomics Clin. Appl.* **2**, 1386–1402 (2008).
- Barr, J.R. *et al.* Isotope-dilution mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin. Chem.* **42**, 1676–1682 (1996).

- Barnidge, D.R. *et al.* Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Anal. Chem.* **75**, 445–451 (2003).
- Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem mass spectrometry. *Proc. Natl. Acad. Sci. USA* **100**, 6940–6945 (2003).
- Barnidge, D.R., Goodmanson, M.K., Klee, G.G. & Muddiman, D.C. Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-MS/MS using protein cleavage and isotope dilution MS. *J. Proteome Res.* **3**, 644–652 (2004).
- Kuhn, E. *et al.* Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and ¹³C-labeled peptide standards. *Proteomics* **4**, 1175–1186 (2004).
- Anderson, L. & Hunter, C.L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **5**, 573–588 (2006).
- 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* **6**, 2212–2229 (2007).
- Aguiar, M., Masse, R. & Gibbs, B.F. Mass spectrometric quantitation of C-reactive protein using labeled tryptic peptides. *Anal. Biochem.* **354**, 175–181 (2006).
- Wilcken, B., Wiley, V., Hammond, J. & Carpenter, K. Screening newborns for inborn errors of metabolism by tandem mass spectrometry. *N. Engl. J. Med.* **348**, 2304–2312 (2003).
- Whiteaker, J.R. *et al.* Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *J. Proteome Res.* **6**, 3875–3876 (2007).
- Anderson, N.L. *et al.* The human plasma proteome: a non-redundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3**, 311–326 (2004).
- Omenn, G.S. *et al.* Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5**, 3226–3245 (2005).
- Kuzyk, M.A. *et al.* MRM-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell. Proteomics* published online, doi:10.1074/mcp.M800540-MCP200 (1 May 2009).
- Anderson, N.L. & Anderson, N.G. The human plasma proteome: History, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867 (2002).
- Whiteaker, J.R. *et al.* Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *J. Proteome Res.* **6**, 3962–3975 (2007).
- Berna, M. *et al.* Quantification of NTproBNP in rat serum using immunoprecipitation and LC/MS/MS: a biomarker of drug-induced cardiac hypertrophy. *Anal. Chem.* **80**, 561–566 (2008).
- Berna, M.J., Zhen, Y., Watson, D.E., Hale, J.E. & Ackermann, B.L. Strategic use of immunoprecipitation and LC/MS/MS for trace-level protein quantification: Myosin light chain 1, a biomarker of cardiac necrosis. *Anal. Chem.* **79**, 4199–4205 (2007).
- Labugger, R. *et al.* Strategy for analysis of cardiac troponins in biological samples with a combination of affinity chromatography and mass spectrometry. *Clin. Chem.* **49**, 873–879 (2003).
- Nicol, G.R. *et al.* Use of an immunoaffinity-mass spectrometry-based approach for the quantification of protein biomarkers from serum samples of lung cancer patients. *Mol. Cell. Proteomics* **7**, 1974–1982 (2008).
- Anderson, N.L. *et al.* Mass spectrometric quantitation of peptides and proteins using stable isotope standards and capture by anti-peptide antibodies (SISCAPA). *J. Prot. Res.* **3**, 235–244 (2004).
- Hoofnagle, A.N., Becker, J.O., Wener, M.H. & Heinecke, J.W. Quantification of thyroglobulin, a low-abundance serum protein, by immunoaffinity peptide enrichment and tandem mass spectrometry. *Clin. Chem.* **54**, 1796–1804 (2008).
- Kuhn, E. *et al.* Developing multiplexed assays for troponin I and interleukin-33 in plasma by peptide immunoaffinity enrichment and targeted mass spectrometry. *Clin. Chem.* published online, doi:10.1373/clinchem.2009.123935 (16 April 2009).
- Biopharmaceutics Coordinating Committee, Center for Drug Evaluation and Research, Center for Veterinary Medicine, US Food and Drug Administration (FDA). *Guidance for Industry: Bioanalytical Method Validation* (FDA, Rockville, MD; May 2001) <<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070107.pdf>>.
- Slev, P.R., Rawlins, M.L. & Roberts, W.L. Performance characteristics of seven automated CA 15–3 assays. *Am. J. Clin. Pathol.* **125**, 752–757 (2006).
- Sapin, R. Insulin immunoassays: fast approaching 50 years of existence and still calling for standardization. *Clin. Chem.* **53**, 810–812 (2007).

ONLINE METHODS

Commercial instrumentation and materials are identified in this work to adequately describe the experimental procedure. Such identification does not imply recommendation or endorsement by the authors and the National Institute of Standards and Technology nor does it imply that the equipment, instruments or materials are necessarily the best available for the purpose.

Materials. The light ($^{12}\text{C}/^{14}\text{N}$) and heavy ($^{13}\text{C}/^{15}\text{N}$) forms of 11 unique signature peptides corresponding to the target proteins were synthesized and purified by Anaspec. Signature tryptic peptides containing C-terminal arginine and lysine residues were synthesized as $^{13}\text{C}_6$ and $^{13}\text{C}_6^{15}\text{N}_2$ analogs, respectively. Two tryptic peptides contained a $^{13}\text{C}_5$ -valine residue and one N-terminal partial tryptic peptide was prepared as the $^{13}\text{C}_6$ -leucine analog (Table 1). Target proteins were purchased either from Sigma (equine myoglobin, bovine myelin basic protein, bovine aprotinin, murine leptin and horseradish peroxidase) or from Scripps Laboratories (human C-reactive protein and human prostate-specific antigen). Pooled and filtered (0.2 μm) human $\text{K}_2\text{-EDTA}$ plasma was purchased from Bioreclamation. ReproSil-Pur C18-AQ resin (3 μm particle size) was purchased from Dr. Maisch. Mass spectrometry grade Trypsin Gold was obtained from Promega. Iodoacetamide, dithiothreitol and urea were purchased from Sigma Chemical or from ThermoFisher Scientific.

Peptide purity of synthetic peptides and amino acid analysis. Peptide and isotopic purity of the synthetic peptides was estimated to be >98% as determined by LC-UV and matrix-assisted laser desorption ionization (MALDI)-MS at Anaspec. In addition, isotopic purity of heavy peptides was assessed at the National Institute of Standards and Technology (NIST) by MALDI-MS on a 4700 tandem time of flight (TOF)/TOF mass spectrometer (Applied Biosystems/MDS Analytical Technologies) in reflector mode. The laser intensity was adjusted to keep the maximum ion count below 25,000, preventing detector saturation and distortion of the observed peptide isotope distribution. For all heavy peptides, comparison between the observed isotope distributions obtained in the acquired spectra and simulated distributions, with varying percent incorporation of the heavy label, indicated that the isotopic peptide purity was ~99%. The exact concentrations of synthetic peptides and target proteins were determined by amino acid analysis after gas-phase acid hydrolysis followed by isotope dilution-LC-tandem MS (MS/MS)²⁸ at NIST.

ELISA. To determine the endogenous CRP and PSA concentrations of the human plasma used to prepare the study samples, the plasma was analyzed using a Quantikine Human C-Reactive Protein ELISA kit and a Quantikine Human Kallikrein 3/Prostate Specific Antigen Protein ELISA kit (R&D Systems). The stock CRP and PSA solutions used to prepare the spiked plasma study samples were used to prepare calibrants for the respective ELISAs. For the CRP assay, the plasma sample was diluted 200-fold before analysis, whereas for the PSA assay, the plasma was assayed undiluted. Both assays were performed according to the manufacturer's instructions. The assay response was measured using a BioTek Synergy HT microplate reader. The plasma CRP concentration of the unspiked plasma was ~6.0 mg/l. The PSA level in the unspiked plasma was below the detection limit of the ELISA, which is about 1 ng/ml.

Sample preparation for study I: digested plasma spiked with signature peptides. The samples for studies I and II were prepared (including tryptic digestion) at NIST, and shipped to the eight participating laboratories. Stock solutions (100 pmol/ μl) of the individual heavy and light peptides were dissolved in an aqueous solution of 30% acetonitrile (vol/vol) and 0.1% formic acid (vol/vol). Equimolar mixtures containing either the light or the heavy peptides were prepared at 1 pmol/ μl .

As the background matrix, 1 ml of pooled human K_2EDTA plasma was diluted with 2 ml of 150 mM Tris, pH 8.0, containing 9 M urea and 30 mM dithiothreitol. The final protein concentration of human plasma before dilution was ~63.7 g/l by a bicinchoninic acid colorimetric assay (Pierce Biotechnology) and a 7% solution of BSA as a standard (NIST Standard Reference Material). Plasma proteins were reduced and denatured by heating for 30 min at 37 °C. The sample was cooled to ~20–23 °C before a 260 μl aliquot of 500 mM aqueous iodoacetamide was added to achieve a final concentration of 40 mM. The sample was incubated at ~20–23 °C for 30 min in the dark.

Next, the plasma sample was diluted approximately tenfold with 100 mM Tris, pH 8.0, and digested with Promega Trypsin Gold (1 mg) at 37 °C. After 18 h, proteolysis was stopped by acidifying the solution to pH 2 with 1% (vol/vol) formic acid. The digested plasma was desalted using a 35 ml Oasis HLB LP solid-phase extraction cartridge (Waters) and the peptide eluate was lyophilized. Finally, the mixture of plasma peptides was reconstituted with 60 ml of an aqueous solution containing 0.6% (vol/vol) acetonitrile and 1% (vol/vol) formic acid, achieving a 60-fold dilution of the plasma that resulted in a final concentration of ~1 $\mu\text{g}/\mu\text{l}$ (total protein).

A multistep process was used to prepare trypsin-digested plasma samples that contained varying amounts of the light peptide mixture (500, 275, 151, 83, 46, 25, 8.6, 2.9, 1.0 fmol/ μl) and 50 fmol/ μl of the heavy peptide mixture. First, aqueous 1,000 fmol/ μl stock solutions of the light (solution A) and the heavy (solution B) peptides were prepared. Second, a 50 fmol/ μl solution of the heavy signature peptides (solution C) was prepared by lyophilizing an aliquot (1 ml) of solution B and reconstituting with 20 ml of digested diluted plasma. Third, an aliquot (0.9 ml) of solution A was lyophilized and reconstituted with an appropriate volume (1.8 ml) of solution C to produce solution D (study I, sample J) that contained a mixture of light and heavy signature peptides at concentrations of 500 and 50 fmol/ μl , respectively. Fourth, the remaining standards were prepared by serial dilution of solution D with solution C. Aliquots (25 μl) of the working standards were dispensed into polypropylene sample tubes, stored at –80 °C, and shipped frozen on dry ice to the participating laboratories.

Sample preparation for study II: digested plasma spiked with digested proteins. Individual solutions of the seven target proteins were prepared in water (ranging between 62 pmol/ μl and 145 pmol/ μl). Aliquots of these stock solutions were co-lyophilized and reconstituted in 100 mM Tris, pH 8.0, containing 6 M urea and 5 mM dithiothreitol to produce an equimolar mixture (100 pmol/ μl). Reduction, denaturation and alkylation of the proteins were carried out as described for study I. Next, the protein mixture was diluted tenfold with 100 mM Tris, pH 8.0, and Promega Trypsin Gold was added at an enzyme/substrate ratio of 1:50 (wt/wt). Tryptic digestion was carried out as described above and the resulting peptide mixture was desalted using a 1 ml Oasis HLB solid phase extraction cartridge. The eluted peptides were lyophilized to dryness and reconstituted with an aqueous solution of 6% (vol/vol) acetonitrile and 1% (vol/vol) formic acid. Finally, study II samples containing 500, 275, 151, 83, 46, 25, 8.6, 2.9 or 1.0 fmol/ μl of the trypsin-digested protein mixture and 50 fmol/ μl of isotopically labeled signature peptides were prepared as described for study I above. Aliquots (25 μl) were dispensed into polypropylene sample tubes, stored at –80 °C, and shipped frozen on dry ice to the participating laboratories.

Sample preparation for study III: digestion of plasma spiked with target proteins. Stock solutions of human pooled K_2EDTA plasma, human pooled K_2EDTA plasma spiked with the seven target proteins spanning a concentration range of 0.06–30 pmol/ μl , and a 500 fmol/ μl mixture of the 11 isotopically labeled signature peptides were prepared and aliquoted at NIST (SOP, **Supplementary Methods**). A working solution was prepared by lyophilizing an aliquot of the 50 pmol/ μl mixture of the seven target proteins and reconstituting the sample with whole plasma to a final concentration of 30 pmol/ μl . Solutions with lower concentrations of spiked-in proteins were prepared by serial dilution of the 30 pmol/ μl solution with plasma (SOP, **Supplementary Methods**). Study samples were aliquoted (35 μl), stored at –80 °C and shipped to the eight participating sites.

The remaining sample preparation steps were performed in triplicate (study IIIa, IIIb and IIIc) at each site. Plasma and spiked plasma samples were digested with trypsin using a scaled-down version of the protocol described for study I. Briefly, a 25 μl aliquot of each plasma sample was combined with 50 μl of buffer (300 mM Tris, pH 8/9M urea/20 mM DTT), reduced (30 min at 37 °C), and alkylated by adding 500 mM aqueous iodoacetamide (40 mM final concentration) and incubating at ~20–23 °C for 30 min in the dark. Samples were diluted tenfold with 100 mM Tris (pH 8) and digested with Promega Trypsin Gold (enzyme/substrate ratio of 1:50 (wt/wt), 37 °C for 18 h). The trypsin was provided to each participating site by NIST and was from the same lot as that used to prepare samples for studies I and II. Proteolysis was

stopped by lowering the pH to 2 with 1% formic acid and the resulting peptide mixtures were desalted off-line by using Waters Oasis HLB 1 cc, 30 mg solid phase extraction cartridges (**Supplementary Methods**). Eluted tryptic peptides were lyophilized to dryness and resuspended in 25 μ l of aqueous solution containing 3% acetonitrile and 5% formic acid. A mixture of the labeled signature peptides was added to aliquots of each reconstituted plasma digestion solution to yield standards that contained 50 fmol/ μ l $^{13}\text{C}/^{15}\text{N}$ -signature peptides and tryptic ^{12}C -peptides (derived from the digested added-in target proteins) that spanned a range of concentrations (500, 275, 151, 83, 46, 25, 8.6, 2.9, 1.0 fmol/ μ l).

Reversed phase nanoflow liquid chromatography (nanoLC). Peptide mixtures were separated by online reversed phase nano high-performance liquid chromatography using dual pumping systems equipped with autosamplers: specifically six nanoLC-2D and one nanoLC-1D Plus System from Eksigent Technologies and one model 1100 Nanosystem from Agilent Technologies. PicoFrit (New Objective) columns, 75 μ m internal diameter (i.d.) $\times$ 120 mm long, 10 μ m i.d. tip, were self-packed with ReproSil-Pur C18-AQ (3 μ m particle size and 120 Å pore size). Separations were performed at mobile phase flow rates of either 200 nl/min (Agilent) or 300 nl/min (Eksigent) on the binary pump systems using 0.1% (vol/vol) formic acid in water (mobile phase A) and 90% (vol/vol) acetonitrile with 0.1% (vol/vol) formic acid (mobile phase B). One microliter injections of the peptide digestion mixtures were separated using a binary gradient of 3–20% B in 3 min, 20–60% B in 35 min, 60–90% B in 2 min and at 90% B for 4 min (**Supplementary Methods** and **Supplementary Table 1a**).

4000 QTRAP instruments. Seven 4000 QTRAP hybrid triple quadrupole/linear ion trap mass spectrometers (Applied Biosystems/MDS Analytical Technologies) located at different sites were used to acquire MRM data for studies I, II and IIIa–c. General instrument operating parameters for the 4000 QTRAP instruments were kept uniform across the seven sites (see SOP, **Supplementary Methods** and **Supplementary Table 1a**). Typically, these mass spectrometers were required to operate with ion spray voltages of $2,200 \pm 200$ V, curtain gas 20, nebulizer gas (GS1) 5 ± 2 , and interface heater temperature (IHT) 150 °C. MRM transitions were optimized for maximum transmission efficiency and sensitivity for individual instruments by infusion of unlabeled signature peptides. Optimized declustering potential, collision energy and collision cell exit potential are reported in **Supplementary Tables 1b–e** for each MRM transition along with the corresponding instrument used at each site. A total of 66 MRM transitions (3 per peptide) were monitored during an individual sample analysis. Identical instrument parameters were used for each unlabeled/labeled peptide pair. Due to the complexity of the matrix and to achieve maximum specificity, MRM transitions were acquired at unit resolution in the first and third quadrupoles (Q1 and Q3). Dwell times of 10 ms were used for all transitions and cycle times were set to 0.99 s.

TSQ Quantum Ultra instrument. A TSQ Quantum Ultra triple quadrupole mass spectrometer (ThermoFisher Scientific) was used to acquire MRM data for studies I, II and IIIa–c. Instrument operating parameters, based on precursor ion charge states and m/z values, were optimized for all MRM transitions by direct infusion of each unlabeled signature peptide (**Supplementary Table 1b**). Representative instrument operating parameters for the TSQ Quantum Ultra included a spray voltage of $1,200 \pm 200$ V, a capillary offset voltage of 35 V, a skimmer offset voltage of –5 V and a capillary temperature of 210 °C. Tube lens voltages used for all unlabeled and labeled peptides, which were based on values generated during the automatic tuning and calibration process, were not individually optimized. A single scan event was used to monitor a total of 66 MRM transitions, 3 MRM transitions per peptide, using the following parameters: Q1 and Q3 unit resolution of 0.7 FWHM, Q2 gas pressure of 1.5 mTorr, scan width of 0.004 m/z and a scan time of 15 ms (**Supplementary Table 1a**).

MRM data acquisition. Mass spectrometric data were acquired for the three separate studies that used the sample sets described above. Additional samples were analyzed for quality control (QC) purposes. These included equimolar mixtures of ^{12}C - and $^{13}\text{C}/^{15}\text{N}$ -signature peptides (no plasma background matrix). Study samples were analyzed in a specified order, from

lowest to highest concentration of added-in peptides, with four technical replicates for each sample as described in detail in the accompanying SOP (**Supplementary Methods**). A total of 57 LC-MRM runs were recorded per study.

Data analysis platforms. Instrument-specific data analysis software was employed for quantitative analyses: MultiQuant (Applied Biosystems/MDS Analytical Technologies) was used to process 4000 QTRAP data and SRM Workflow (prototype, ThermoFisher Scientific) was used to process TSQ Quantum Ultra data. Briefly, the MRM transitions for each peptide were individually integrated to generate ion current peak areas representing each of the ^{12}C and $^{13}\text{C}/^{15}\text{N}$ peptide fragment ion signals. A peak area ratio characteristic for each MRM transition was calculated by dividing the ^{12}C peak area by that of its corresponding $^{13}\text{C}/^{15}\text{N}$ counterpart. All data analysis and peak area integrations were initially performed at each of the eight sites, however, for the seven 4000 QTRAP instruments data sets were further ‘re-integrated’ at one central site to guarantee uniform data analysis and uniform determination of outlier peaks. MultiQuant data were directly exported as a text file for further statistical analyses. SRM Workflow data were exported as .csv files and reformatted using an in-house Perl script, which was written to enable cross-site comparisons. Software versions and other details for data analysis for each site are listed in **Supplementary Table 1a**.

Statistical methods. For all statistical calculations, final concentrations of heavy and light peptides and added proteins were adjusted according to the gravimetric measurements described in **Supplementary Table 6a–f**.

Graphical methods. Data from MRM experiments were exported from MultiQuant (MQ) or SRM Workflow and imported into the R statistical computing environment (<http://www.R-project.org/>) for graphical review and statistical analyses²⁹. Comprehensive plots were made of all experiments (studies I, II, IIIa–c) for all peptides (ten) and sites (eight) with estimated concentration on the vertical axis and theoretical concentrations on the horizontal axis. Estimated concentrations were based upon the following equation: [calculated concentration (fmol/ μ l)] = peak area ratio of analyte to internal standard $\times$ 50 fmol/ μ l of internal standard. Plots were made on the linear scale (**Supplementary Fig. 5**), with additional plots used for data visualization shown in **Supplementary Fig. 6** and **Supplementary Appendix**. Plots were sent to the sites to identify and adjust integration errors by inspecting integrations in MQ or SRM Workflow from visually identified outliers in the plots.

Statistical models for linear calibration curves. When added into solution (buffer or plasma) the observed concentration y of a signature peptide should be identical to x the concentration at time of addition. However, full recovery of the peptide by the assay does not always occur, and percent recovery r is often $<100\%$. Thus, the relation between the observed and the expected concentration is $y = rx$. Variation between replicates as measured by the s.d. usually increases in proportion to the concentration x , although CV = s.d./mean generally decreases across the range of concentrations. Thus, the statistical model for the linear regression lines is:

$$E(y) = r \times x + c, \text{ s.d.}(y) = k \times x$$

where $E(y)$ is the average observed concentration (based on the 12 observations arising from the three transitions and four replicates) at the concentration x at time of addition. $\text{s.d.}(y)$ is the standard deviation of the observations at x , and increases proportionally with x , k being the proportionality constant. The slope of the line is the percent recovery r and the line has an expected intercept c of 0. A statistically significant nonzero intercept can be interpreted as the endogenous level of the peptide existing in the solution with no added-in signature peptide. The s.d. increasing with the spiked-in concentration x requires the linear regression to be weighted proportionally to the inverse of the variance ($(\text{s.d.})^2$), and so the weight is $1/x^2$ (refs. 30,31).

Robust linear regression. In some cases, data points were observed that did not fall near the linear trend line. These points, assigned as outliers, had plausible explanations in the majority of cases, including interference in the

heavy or light channels, saturation, presence of endogenous protein or were unexplained. A robust linear regression method was applied to down-weight the influence of outliers based on the estimation of the parameters in a linear regression: slope, intercept and s.d. about the regression line. In addition, this model also reports the standard error for the slope and the intercept so that 95% confidence intervals can be calculated as the estimate $\pm 1.96 \times$ standard error (Supplementary Table 4). Data points from all three MRM transitions and the four replicate injections were included in estimating the linear trend. The robust linear model function in R was applied^{32,33} using Tukey's biweight function to minimize the influence of outliers. For example, where there was an interference in one transition, the robust method was little influenced by the aberrant transition, and the parameter estimates resulted mainly from the data contributed by the other two transitions, which were usually coincident on the linear trend. This approach allowed all data points to be included in the estimation process and did not require subjective elimination of outliers.

Assessment metrics for quantitative MRM assays. The metrics used for assessing reproducibility of the MRM assays for the seven target proteins (ten peptides) were: (i) intralaboratory precision, represented by the median CV calculated from all concentration points for a particular peptide (based on quadruplicate measurements for a single transition used to calculate LOQ/LOD, see below) for each site, and for each study, and (ii) interlaboratory precision, represented by the median CV calculated at each concentration point for a particular peptide (based on quadruplicate measurements for a single transition used to calculate LOQ/LOD, see below) across all sites and for each study.

The CV is calculated as the ratio of the s.d. to the mean of a set of measurements. The CV calculations at each concentration point for a peptide at a given laboratory is based on four replicates for studies I and II and on 12 data points (four technical replicates for each of the three process replicates) for study III.

Determining LOD and LOQ. The following methods can be used to calculate the lower LOD for an analyte (defined as the concentration level at which the analyte can be reliably detected in the sample under consideration) and the lower LOQ, defined as the level at which the analyte can be detected and measured with sufficient precision. Methods range from straightforward modeling of blank sample variance using normal distributions³⁴, modeling variance as a function of concentration³⁵ fitting the relative s.d. along the concentration curve³⁶ and empirical methods³⁷. For this study, a simple method was chosen for calculating LOD³⁴. Once the LOD was determined separately for each transition of each peptide, the LOQ was calculated using the customary relation: $LOQ = 3 \times LOD$ ³². The LOD was based on the variance of the blank sample (sample A1, with no analyte added in) and the variance of the lowest level added-in sample (sample B, with analyte at 1 fmol/ μ l). Assuming a type I error rate $\alpha = 0.05$ for deciding that the analyte is present when it is not,

and a type II error rate $\beta = 0.05$ for not detecting the analyte when it is present, the LOD was derived as:

$$LOD = LOB + c_{\beta} \times s.d._s$$

LOB (limit of blank) was defined as the 95th percentile of the blank A1 samples³⁸. This was estimated as the mean plus $t_{1-\beta} \times s.d._b$, where $s.d._b$ was the standard deviation of the blank samples, and $s.d._s$ was the standard deviation of the lowest analyte concentration point, sample B. For a relatively small number of repeated measurements for sample B, c_{β} was approximated as $t_{1-\beta}$, where $t_{1-\beta}$ is the $(1-\beta)$ percentile of the standard t distribution on f degrees of freedom. It is important to relate the LOD calculations to the measurement process. In this study, the final result of measuring the sample is obtained from the four replications as measured by the best transition. The LOD calculation when four values are averaged to obtain the final measurement requires the s.d. estimates to be halved, so the LOD equation becomes:

$$LOD = \text{mean}_b + t_{1-\beta} \times (s.d._b + s.d._s)/2.$$

LOD values were initially calculated for all three transitions monitored for each peptide. The transition with the smallest root mean square deviation from the minimum LODs for both studies I and II was chosen as the best transition. This transition is used to report LOD and LOQ for both studies I and II, and for interlaboratory and intralaboratory CV calculations for all studies.

28. Burkitt, W. *et al.* Toward Système International d'Unité-traceable protein quantification: from amino acids to proteins. *Anal. Biochem.* **376**, 242–251 (2008).
29. R Development Core Team. *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2008).
30. Lavagnini, I. & Magno, F. A statistical overview on univariate calibration, inverse regression, and detection limits: Application to gas chromatography/mass spectrometry technique. *Mass Spectrom. Rev.* **26**, 1–18 (2007).
31. Xu, X., Keefer, L.K., Ziegler, R.G. & Veenstra, T.D. A liquid chromatography-mass spectrometry method for the quantitative analysis of urinary endogenous estrogen metabolites. *Nat. Protocols* **2**, 1350–1355 (2007).
32. Marazzi, A.. *Algorithms, Routines and S Functions for Robust Statistics* (CRC Press, Boca Raton, Florida, USA, 1993).
33. Venables, W.N. & Ripley, B.D. *Modern Applied Statistics with S* edn. 4 (Springer, New York, 2002).
34. Currie, L.A. Limits for qualitative detection and quantitative determination. *Anal. Chem.* **40**, 586–593 (1968).
35. Zorn, M.E., Gibbons, R.D. & Sonzogni, W.C. Weighted least-squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration. *Anal. Chem.* **69**, 3069–3075 (1997).
36. Vial, J., Mapihan, K.L. & Jardy, A. What is the best means of estimating the detection and quantification limits of a chromatographic method? *Chromatographia* **57** Supplement, S303–S306 (2003).
37. Armbruster, D.A., Tillman, M.D. & Hubbs, L.M. Limit of detection (LOD)/limit of quantitation (LOQ): Comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused Drugs. *Clin. Chem.* **40**, 1233–1238 (1994).
38. Linnet, K. & Kondratovich, M. Partly nonparametric approach for determining the limit of detection. *Clin. Chem.* **50**, 732–740 (2004).