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 Blackman5,18, 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 Variyath2, Lorenzo J Vega-Montoto5, Åsa Wahlander10, Sofia Waldemarson10, Mu Wang13,14, Jeffrey R Whiteaker15, Lei Zhao13, N Leigh Anderson16, Susan J Fisher7, Daniel C Liebler6, Amanda G Paulovich15, Fred E Regnier9, Paul Tempest17 & 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.

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, 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 differences1,2. 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,

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 differences1,2.

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,
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. 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. 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 verification. 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 (CPTAC) evaluated intra- and interlaboratory analytical performance of SID-MRM-MS assays for quantifying seven target proteins added to human plasma. Our studies demonstrate 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 µg/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 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.

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
used 4000 QTRAP mass spectrometer; 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 YEVOQEVFTKQPQLWP 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 SSDLVALSGHGTFGK 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 ≤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 ≤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 ≤25% across all concentrations, demonstrating very good reproducibility for sample processing (Supplementary Fig. 1 and Supplementary Table 2c).

**Interlabalatory 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/µl were predominantly ≤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/µl 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).

---

**Table 1 Target proteins and their signature peptides**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Abbrev</th>
<th>Species</th>
<th>Signature peptide</th>
<th>MH+ (mono)</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aprotinin</td>
<td>APR-AGL</td>
<td>Bovine</td>
<td>AGLCQGTVYGGGR</td>
<td>1493.7</td>
<td>747.3</td>
<td>863.4</td>
</tr>
<tr>
<td>Leptin</td>
<td>LEP-IND</td>
<td>Mouse</td>
<td>INDI8HTQVESAK</td>
<td>1407.3</td>
<td>469.9</td>
<td>590.8</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>MYO-LFT</td>
<td>Horse</td>
<td>LFTGHPELTK</td>
<td>1279.7</td>
<td>427.2</td>
<td>510.3</td>
</tr>
<tr>
<td>Myelins basic protein</td>
<td>MBP-HGF</td>
<td>Bovine</td>
<td>HGFLFR</td>
<td>732.4</td>
<td>366.7</td>
<td>391.3</td>
</tr>
<tr>
<td>Myelins basic protein</td>
<td>MBP-YLA</td>
<td>Bovine</td>
<td>YLASASTMDHAR</td>
<td>1328.6</td>
<td>443.5</td>
<td>491.2</td>
</tr>
<tr>
<td>Prostate-specific antigen</td>
<td>PSA-IVG</td>
<td>Human</td>
<td>IWGVWCEK</td>
<td>1082.5</td>
<td>541.7</td>
<td>808.3</td>
</tr>
<tr>
<td>Prostate-specific antigen</td>
<td>PSA-LSE</td>
<td>Human</td>
<td>LSEPAELTDRIK</td>
<td>1280.7</td>
<td>640.8</td>
<td>783.4</td>
</tr>
<tr>
<td>Peroxidase</td>
<td>HRP-SSD</td>
<td>Horseradish</td>
<td>SSDLVALSGHGTFGK</td>
<td>1483.8</td>
<td>495.3</td>
<td>711.4</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>CRP-ESD</td>
<td>Human</td>
<td>ESDDSYSVSLK</td>
<td>1136.6</td>
<td>568.8</td>
<td>617.4</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>CRP-GYS</td>
<td>Human</td>
<td>GYSIFSATIK</td>
<td>1144.6</td>
<td>572.8</td>
<td>724.4</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>CRP-YEV</td>
<td>Human</td>
<td>YEVOQEVFTKQPQLWP</td>
<td>1826.9</td>
<td>914.0</td>
<td>1053.5</td>
</tr>
</tbody>
</table>

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.
Figure 2 Box plots of variation in MRM quantitative measurements, interlaboratory CV, intralaboratory 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 µg/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 µg/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 x IQR. Values >1.5 x IQR are deemed outliers, and shown as separate points.
Table 2 Summary of results for studies I, II, and III

<table>
<thead>
<tr>
<th>Signature Peptide</th>
<th>Study I</th>
<th>Study II</th>
<th>Study III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intralaboratory CV</td>
<td>Intralaboratory CV</td>
<td>Intralaboratory CV</td>
</tr>
<tr>
<td></td>
<td>Linear slope</td>
<td>Recov. (%)</td>
<td>Linear slope</td>
</tr>
<tr>
<td>APR-AGL</td>
<td>9.2%</td>
<td>1.157</td>
<td>11.4%</td>
</tr>
<tr>
<td>CRP-ESD</td>
<td>5.9%</td>
<td>1.124</td>
<td>118.4</td>
</tr>
<tr>
<td>CRP-GYS</td>
<td>5.4%</td>
<td>1.324</td>
<td>140.5</td>
</tr>
<tr>
<td>HRP-SSD</td>
<td>14.1%</td>
<td>1.198</td>
<td>120.4</td>
</tr>
<tr>
<td>LEP-IND</td>
<td>12.5%</td>
<td>1.163</td>
<td>119.1</td>
</tr>
<tr>
<td>MBP-HGF</td>
<td>4.3%</td>
<td>1.161</td>
<td>118.6</td>
</tr>
<tr>
<td>MBP-YLA</td>
<td>5.1%</td>
<td>1.275</td>
<td>130.3</td>
</tr>
<tr>
<td>MYO-LFT</td>
<td>4.9%</td>
<td>1.518</td>
<td>154.4</td>
</tr>
<tr>
<td>PSA-IVG</td>
<td>6.9%</td>
<td>1.658</td>
<td>165.4</td>
</tr>
<tr>
<td>PSA-LSE</td>
<td>8.9%</td>
<td>1.098</td>
<td>111.4</td>
</tr>
</tbody>
</table>

*Combined results for process replicates a, b, c for each peptide across sites for interlaboratory CV, intralaboratory CV, linear slope and percent recovery. †Intralaboratory 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. ‡Percent recovery was determined from the mid-concentration point, 46 fmol/mg/mL, 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 theoretical slope 1.
ANALYSIS

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/µl; 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-c. Altogether, peptide responses in study I had an average slope ranging from 1.1 to 1.6 with an interlaboratory CV ≤10% for most of the peptides (Table 2 and Supplementary Tables 4a-d). 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 ≤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 ≤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 ≥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 ≥40%, which is within an acceptable range for verification assays2,9. Four peptides (CRP-GYS, LEP-IND, MBP-HGF and MBP-YLA) had recoveries ≤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 (13C/14N)-labeled internal standard peptides or heavy (13C/15N)-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-EKD (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
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 electrospay 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 disease\textsuperscript{11}. More recently, the merits of SID-MRM-MS for quantifying peptides derived from proteins in plasma have been demonstrated in several laboratories\textsuperscript{4–8,12}. 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\textsuperscript{13,14}, enabling the development of an effective biomarker pipeline\textsuperscript{1}.

Reproducibility and precision of the quantitative measurements for nine of ten peptides tested across eight laboratories ranged from 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 biomarkers\textsuperscript{2} present at more than \(\sim 2–6 \mu g/ml\) 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–c). 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...
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 retentions 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 studied\(^\text{15}\). 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 Supplemental 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 Supplemental 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 10\(^{11}\) dynamic range of protein abundances in blood, and the susceptibility to interference from other peptides and their fragment ions is greatest in this matrix\(^\text{16}\). Typical LODs and LOQs observed in prior studies of unFractionated plasma are in the high 100s of ng/ml to low \(\mu\)g/ml range of target protein\(^\text{6,17,17}\). Results described here are consistent with these reports across sites and instrument platforms (Fig. 2c and Supplemental 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\(^\text{9,17-24}\). 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 \(\mu\)g/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 assays\(^\text{2}\) is not as stringent as that currently required for US Food and Drug Administration–approved clinical assays\(^\text{25}\). Beyond candidate verification, SID-MRM-MS assays may eventually have potential to replace certain clinical immunoassays, especially in cases where interferences are known to exist\(^\text{23}\) 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\(^\text{26,27}\).

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:
CKpN0b12ULLwCaIoxXn/spuw4rYIF6FH/L+/6sHAKGZCsj4f7TD0Rau
JjAwv9b8At8136H0iqj2itYPAPm29P2caAAAAAAT6w==. 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 informa-
tion 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

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.A.C. contributed to study design and SOP development. D.M.B and N.G.D.
D.S., T.J.T., J.R.W., A.W., S.W., L.Z., and L.J.Z. contributed to generation of data.
L.I.-Y.-M. contributed to bioinformatics and statistical analysis. S.E.A., T.A., H.K.,
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/

1. Rifai, N., Gillette, M.A. & Carr, S.A. Protein biomarker discovery and validation: the
biomarker discovery and clinical validation: The tar pit of the protein biomarker
3. Barr, J.R. et al. Isotopic-dilution mass spectrometric quantification of specific proteins:
rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide
5. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantifica-
tion of proteins and phosphoproteins from cell lysates by tandem mass spectrometry.
tion of the model biomarker prostate-specific antigen in serum by LC-MS/MS using
7. Kuhn, E. et al. Quantification of C-reactive protein in the serum of patients with
rheumatoid arthritis using multiple reaction monitoring mass spectrometry and
8. Anderson, L. & Hunter, C.L. Quantitative mass spectrometric multiple reaction
9. Keshishian, H., Addona, T., Burgess, M., Kuhn, E. & Carr, S.A. Quantitative, multi-
plied assays for low abundance proteins in plasma by targeted mass spectrometry and
10. Agrawal, G., Masse, R. & Gibbs, B.F. Mass spectrometric quantitation of C-reactive
(2008).
12. 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).
14. Oommen, G.S. et al. Overview of the HUPO plasma proteome project: results from the
pilot phase with 35 collaborating laboratories and multiple analytical groups, generat-
ing a core dataset of 3020 proteins and a publicly-available database. Proteomics 5,
15. Kuzyk, M.A. et al. MRMP-based, multiplexed, absolute quantitation of 45 proteins in
MCP200 (1 May 2009).
17. 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).
18. Berna, M. et al. Quantification of NTproBNP in rat serum using immunonanoprecipit-
immunoprecipitation and LC/MS/MS for trace-level protein quantification: Myosin light
20. 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,
21. Nicol, G.R. et al. Use of an immunoaффinity-mass spectrometry-based approach for the
quantification of protein biomarkers from serum samples of lung cancer patients. Mol.
stable isotope standards and capture by anti-peptide antibodies (SISCAPA). J. Proteome
24. Kuhn, E. et al. Developing multiplexed assays for troponin I and interleukin-33 in
25. 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).
26. Slev, P.R., Rawlins, M.L. & Roberts, W.L. Performance characteristics of seven
methods for absolute quantification of N-terminal pro-B-type natriuretic peptide in
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 ($^{13}$C$_2$N$^4$H) and heavy ($^{13}$C$_6$N$^4$H) forms of 11 unique signature peptides corresponding to the target proteins were synthesized and purified at AnaSpec. Signature tryp tic peptides containing C-terminal arginine and lysine residues were synthesized as $^{13}$C$_2$N and $^{13}$C$_6$N$^4$H$^2$ analogs, respectively. Two tryp tic peptides contained a $^{13}$C$_{2}$-valine residue and one N-terminal partial tryp tic peptide was prepared as the $^{13}$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 Scirpis Laboratories (human C-reactive protein and human prostate-specific antigen). Pooled and filtered (0.2 μm) human K$_2$-EDTA plasma was purchased from Bioréclamation. Reprozil-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/Prostrate 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 calibrators 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 tryp tic 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$_2$-EDTA 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 bicinchonic 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 at 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 SP 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 μg/ml (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 1) 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$_2$-EDTA plasma, human pooled K$_2$-EDTA 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, IIb 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 8.9/3M 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. The 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. Protein was degraded...
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 μl 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/μl 12C/15N-signature peptides and tryptic 12C-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/μl).

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 μm internal diameter (i.d.) × 120 mm long, 10 μm i.d. tip, were self-packed with ReproSil-Pur C18-AQ (3 μm particle size and 120 A 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 90% (vol/vol) acetonitrile with 0.1% (vol/vol) formic acid (mobile phase B).

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 ± 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–f 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 ± 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 12C- and 13C/15N-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 12C and 13C/15N peptide fragment ions. A peak area ratio characteristic for each MRM transition was calculated by dividing the 12C peak area by that of its corresponding 13C/15N 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 analyses30. 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/μl) = peak area ratio of analyte to internal standard × 50 fmol/μl 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 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. 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 line regression to be weighted proportionally to the inverse of the variance ((s.d.)/2), and so the weight is 1/x2 (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
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 distributions\[^34\], modeling variance as a function of concentration\[^35\] fitting the relative s.d. along the concentration curve\[^36\] and empirical methods\[^37\]. For this study, a simple method was chosen for calculating LOD\[^34\]. Once the LOD was determined separately for each transition of each peptide, the LOQ was calculated using the customary relation: LOQ = 3 \times \text{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/µl). 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:

\[
\text{LOD} = \text{LOD}_0 + c_1 \times s.d_b
\]

LOD (limit of blank) was defined as the 95\(^{th}\) percentile of the blank A1 values\[^38\]. This was estimated as the mean plus \(t_{1,\beta} \times s.d_x\), where s.d.x was the standard deviation of the blank samples, and s.d.b was the standard deviation of the lowest analyte concentration point, sample B. For a relatively small number of repeated measurements for sample B, \(c_1\) was approximated as \(t_{1,\beta}\), where \(t_{1,\beta}\) is the (1–\(\beta\)) percentile of the standard \(t\) distribution on \(df\) 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:

\[
\text{LOD} = \text{mean}_b + t_{1,\beta} \times (s.d_{b1} + s.d_{b2})/2
\]

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.


