analytical chemistry

pubs.acs.org/ac Article

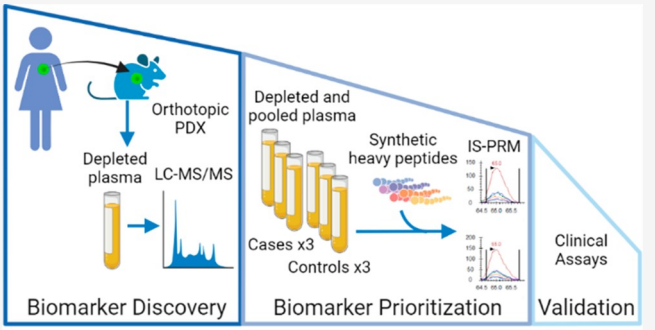
Internal Standard Triggered-Parallel Reaction Monitoring Mass Spectrometry Enables Multiplexed Quantification of Candidate Biomarkers in Plasma

Jacob J. Kennedy,[#] Jeffrey R. Whiteaker,[#] Richard G. Ivey, Aura Burian, Shrabanti Chowdhury, Chia-Feng Tsai, Tao Liu, ChenWei Lin, Oscar D. Murillo, Rachel A. Lundeen, Lisa A. Jones, Philip R. Gafken, Gary Longton, Karin D. Rodland, Steven J. Skates, John Landua, Pei Wang, Michael T. Lewis, and Amanda G. Paulovich*

Cite This: *Anal. Chem.* 2022, 94, 9540−9547 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Despite advances in proteomic technologies, clinical translation of plasma biomarkers remains low, partly due to a major bottleneck between the discovery of candidate biomarkers and costly clinical validation studies. Due to a dearth of multiplexable assays, generally only a few candidate biomarkers are tested, and the validation success rate is accordingly low. Previously, mass spectrometry-based approaches have been used to fill this gap but feature poor quantitative performance and were generally limited to hundreds of proteins. Here, we demonstrate the capability of an internal standard triggered-parallel reaction monitoring (IS-PRM) assay to greatly expand the numbers of candidates that can be tested with improved quantitative performance. The assay couples immunodepletion and fractionation with IS-PRM and was developed and implemented in human plasma to quantify 5176 peptides representing 1314 breast cancer biomarker candidates. Characterization of the IS-PRM assay demonstrated the precision (median % CV of 7.7%), linearity (median $R^2 > 0.999$ over 4 orders of magnitude), and sensitivity (median LLOQ < 1 fmol, approximately) to enable rank-ordering of candidate biomarkers for validation studies. Using three plasma pools from breast cancer patients and three control pools, 893 proteins were quantified, of which 162 candidate biomarkers were verified in at least one of the cancer pools and 22 were verified in all three cancer pools. The assay greatly expands capabilities for quantification of large numbers of proteins and is well suited for prioritization of viable candidate biomarkers.

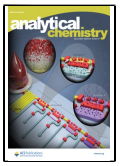


INTRODUCTION

Blood plasma is an easily accessed biofluid that reflects the physiological state of a patient; thus, it remains an attractive source of clinical biomarkers.[1,2] Despite considerable investment and advances in liquid chromatography−mass spectrometry-based (LC-MS/MS) proteomic technologies that allow for deep coverage and quantification of proteins,[3,4] the translation of biomarker discoveries to clinical use remains slow, tedious, and generally disappointing.[5,6] A large factor contributing to this state of the field is the mismatch between the large number of potential biomarkers identified and the resources required for their validation. A method to prioritize among candidate biomarkers to identify those with the greatest probability of clinical utility would allow clinical validation efforts to focus on the subset of candidates most likely to succeed.[7]

The emergence of targeted mass spectrometry-based proteomics approaches (e.g., multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM)[8−10]) enables highly sensitive, specific, and multiplexable assays that can be implemented with relatively low cost (compared to traditional immunoassays).[11] These approaches have proven useful for quantitative biomarker verification studies;[7,12−16] however, even with optimized parameters and careful attention to method details (e.g., tight retention time windows, elimination of overlapping interfering transitions, enrichment and/or fractionation for low abundance targets), it is a challenge to measure more than a few hundred proteins and maintain high analytical performance using these approaches.[17−21] To address the gap between discovery (e.g., thousands of candidates) and validation (e.g., hundreds of candidates), mass spectrometry-based approaches, like accurate inclusion mass spectrometry

Received: October 8, 2021
Accepted: June 17, 2022
Published: June 29, 2022



(AIMS),[22] were implemented to prioritize the most promising candidates for follow-up verification studies.[7,23] While beneficial for enabling a biomarker pipeline, the AIMS approach also had some limitations to the number of candidates that could be tested and suffered from relatively poor quantitative performance, requiring subsequent MRM studies to rank order candidates.

The recent development of internal standard triggered-parallel reaction monitoring mass spectrometry[24] (IS-PRM-MS, implemented using a SureQuant method in the control software of the Thermo Scientific Orbitrap mass spectrometer) has allowed for high multiplexing with the benefits of the performance of PRM.[25,26] The IS-PRM method greatly expands the capacity of the PRM method without relying on retention time windows or coisolation of target peptides by instead relying on added internal standards to trigger the real-time measurement of endogenous peptides. Upon detection of the internal standard, quantification is performed by PRM, allowing for highly sensitive and specific measurements. Although the IS-PRM method has been demonstrated to quantify nearly 600 peptides in complex samples,[24] it has not been evaluated in the context of a biomarker development pipeline for prioritizing high numbers (i.e., thousands) of peptides for validation studies.

In this study, we developed an IS-PRM assay to quantify 5176 peptides representing 1314 candidate breast cancer plasma biomarker proteins. Candidate biomarkers were identified by leveraging preclinical patient derived xenograft (PDX) mouse models to find human proteins secreted into the plasma of the mice. We hypothesized that the IS-PRM method could quantify these candidates in human plasma with high specificity and precision to enable the rank ordering of candidate biomarkers for further investment of resources to perform validation studies in large patient cohorts. The analytical performance of the IS-PRM assay was characterized in fit-for-purpose validation experiments, and the candidates were quantified in three pools of human plasma from women diagnosed with breast cancer and three pools of human plasma from women with benign breast lesions to determine if the candidate biomarker protein signals were higher in the cancer plasma pools. The methodology developed herein presents a significant advance in reliable quantification and verification of large numbers of plasma-based biomarker candidates, and the approach is generally applicable to other diseases or translational studies requiring highly precise relative quantification of large sets of proteins.

MATERIALS AND METHODS

An Expanded Materials and Methods section is available in the Supporting Information.

PDX Plasma Sample Preparation for Biomarker Discovery. All animal experiments were approved by the Baylor College of Medicine Institutional Animal Care and Use Committee (IACUC, Protocol AN-2289) and performed in compliance with the Guide for the Care and Use of Laboratory Animals of the NIH.[27] Human tumor tissue was transplanted into epithelium-free "cleared" fat pads of four-week-old SCID/Beige (Envigo) female mice as ∼1 mm^3 fragments[28] and allowed to grow to ∼500 mm^3. Blood was collected from the mouse via the inferior vena cava using a syringe filled with 50 μL of 0.5 M EDTA and immediately centrifuged at 2000g for 10 min. Immuno-depletion columns coupled to an ÄKTA HPLC system[29] were used to deplete plasma samples, which were subsequently buffer exchanged, digested, and desalted as described,[7] with modifications noted in the Supporting Information. A portion of samples were TMT-labeled using the TMT10plex isobaric label reagent set (TMT, #90110) according to manufacturer's instructions prior to mass spectrometry analysis. Digested plasma samples were fractionated using a described basic reverse-phase liquid chromatography (bRP) workflow[30] and then analyzed by LC-MS/MS using a nanoACQUITY UPLC system (Waters) connected to a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer operated in positive mode as described[31] with modifications noted in the Supporting Information.

Cell Lysate Preparation for Response Curve for IS-PRM Method Characterization. Yeast cells (*Saccharomyces cerevisiae*) were harvested and lysed as described.[32] MCF10A cells were obtained from American Type Culture Collection (ATCC, CRL-10317) and prepared as described in the Supporting Information. Digested and desalted MCF10A lysate was serially diluted with digested and desalted yeast cell lysate to make response curve concentration points that contained 100%, 10%, 1%, 0.1%, and 0% MCF10A. A total of 50 μg of each concentration point underwent the addition of a mix of heavy stable isotope-labeled standards (SIS) and bRP fractionation, as described in the Supporting Information. A total of 96 fractions were concatenated into six fractions by column and analyzed in triplicate.

Human Plasma Sample Depletion, Pooling, And Processing for IS-PRM Evaluation. A total of 138 human plasma samples were obtained from the National Cancer Institute's Early Detection Research Network (EDRN) biorepository.[33,34] The plasma samples were assigned to one of six pools: nonproliferative control (20 samples), proliferative control (20), atypia control (20), Her2+ (19), Triple Negative (19), and two pools of ER+Her2− (20 samples per pool). Each sample pool was further divided into four subpools and these subpools were randomized across the depletion process to reduce the chance of introducing batch effects. Immuno-depletion columns coupled to an ÄKTA HPLC system[29] were used to deplete plasma samples of high-abundant proteins (human IgY14 LC10 (Sigma S5074)) and midabundant proteins (human Supermix LC5 (Sigma S5324)). Independent depleted plasma samples were collected in a single 20 mL flowthrough fraction, pooled by subpool, concentrated with an Amicon Ultra Centrifugal Filter Units (3 kDa cutoff, Millipore UFC900324) and buffer exchanged with 50 mM Ammonium bicarbonate (Sigma A6141). Depleted human plasma samples were denatured with 0.5% RapiGest (Waters 186002123), then digested and desalted as above. A mix of all SIS peptides was added to 200 μg of each of the six digested and desalted human plasma pools and fractionated by bRP fractionation, as described in the Supporting Information. A total of 96 fractions were concatenated into 24 fractions by an alternating column and analyzed by IS-PRM.

LC-MS/MS Analysis of Samples for IS-PRM Method Development, Characterization, and Evaluation. IS-PRM and directed DDA methods were implemented by LC-MS/MS on an Easy-nLC 1000 (Thermo Scientific) coupled to an Orbitrap Eclipse mass spectrometer (Thermo Scientific) operated in positive ion mode, as described in the Supporting Information. Directed DDA MS/MS analysis included targeted mass lists consisted of 7775 entries based on +2 and +3 charge states for each SIS peptide with m/z in the full scan MS range. Raw MS/MS spectra from the analysis were searched as described[35] with modifications noted in the Supporting

https://doi.org/10.1021/acs.analchem.1c04382
Anal. Chem. 2022, 94, 9540−9547

Information. A spectral library was built from the search results using SpectraST.[36]

IS-PRM Mass Spectrometry. IS-PRM was adapted from the SureQuant native implementation in the instrument control software of the Orbitrap-Eclipse as described[24] with modifications noted in the Supporting Information. PRM peak integration was performed by Skyline, and the sum of all six target transitions was used for quantification. Peptide concentrations are reported as the peak area ratio (PAR) of the light and heavy peptides. IS-PRM parameter optimization is described in the Supporting Information. The precursor m/z and intensity thresholds are listed in Table S2, and fragment ions used for identification and quantification are listed in Table S3.

Verification of Candidate Biomarkers. Peptide PAR from the individual plasma pools were filtered to include only those that were greater than 2-fold of the maximum PAR reported in the three yeast blank samples. For the three breast cancer subtypes, PAR were compared to that in the proliferative and nonproliferative control pools. A weighted z-score for each protein was derived based on joint evidence from multiple peptides of the protein to obtain the regression coefficient and p-value of the trend for all the candidate biomarkers as described in the Supporting Information.

Public Access to Data. The LC-MS/MS data associated with biomarker discovery in PDX samples have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[37] with the data set identifier PXD028306. All PRM and directed DDA data associated with the development, characterization and application of the IS-PRM have been deposited in Panorama Public[38] at https://panoramaweb.org/OpKF60.url.

RESULTS AND DISCUSSION

Identification of Biomarker Candidates. Biomarker candidates of breast cancer were identified using plasma from patient-derived xenograft (PDX)-bearing mouse breast cancer models, where proteins leaked, secreted, or shed from the transplanted human breast tumors were the exclusive source of human proteins in the plasma. Plasma samples from 23 PDX-bearing mice were depleted, pooled, proteolytically digested, fractionated, and analyzed by LC-MS/MS (Figure 1). In total, 1314 human proteins were identified as breast cancer biomarker candidates. Of note, 1179 (90%) of the candidate biomarkers were previously observed in proteomic profiles of human breast cancers.[39]

We next sought to prioritize the list of 1314 candidate breast cancer biomarkers. Since these candidates were identified as human proteins present in murine PDX models of cancer, we sought to prioritize those proteins found at differential levels in the plasma of human cancer patients versus controls. Specifically, we deployed targeted proteomics to quantify as many candidates as possible, with high specificity, precision, and sensitivity, while controlling the costs and timeline. Candidates showing differential expression in pooled plasma from cancer vs control patients have the highest priority for downstream investment in quantitative assays that can be run with higher throughput in case-control validation studies using individual patient plasma samples (i.e., without pooling).

Targeted mass spectrometry methods for alleviating the bottleneck in biomarker verification and validation have been presented and used in a variety of scenarios.[7,23] Generally speaking, the targeted methods of MRM and PRM are well suited for highly quantitative assays but are limited in their multiplexing capability to several hundred target peptides/proteins.[40] Quantifying several thousand peptides/proteins for prioritizing candidates for further assay development, such as our intent in this study, has been performed using directed DDA,[7,22] an approach with limited quantitative performance. The recent development of IS-PRM methodology provided improved quantitative performance versus directed DDA, but had not been deployed at the scale of thousands of peptides. Thus, we developed an IS-PRM method, targeting peptides to as many of the candidate proteins as possible, characterized the quantitative performance, and employed the method for successful prioritization of the biomarker candidates in plasma.

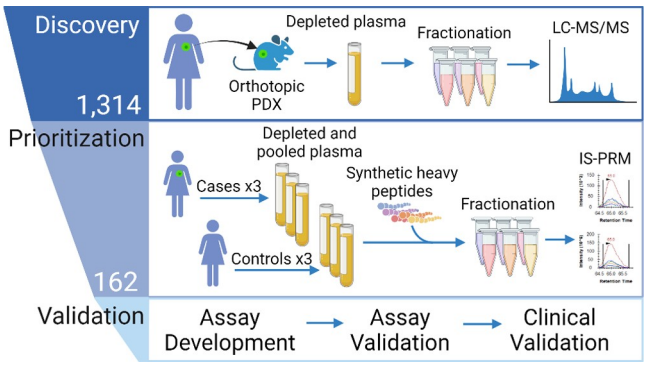


Figure 1. Targeted IS-PRM assay for prioritization of breast cancer biomarkers for validation studies. A candidate list of protein biomarkers was derived from profiling depleted plasma from mice harboring patient-derived xenografts (PDX) of human breast cancer or normal breast tissue to identify human proteins secreted or shed from tumors. Plasma samples from 23 PDX-bearing mice were depleted, pooled, proteolytically digested, fractionated, and profiled by LC-MS/MS, which identified 1314 unique human proteins across the three independent profiles. Because validation of the candidate biomarkers is resource intensive, we sought to use quantitative IS-PRM to prioritize candidate biomarkers showing differential expression in pooled plasma samples from women diagnosed with breast cancer vs women diagnosed with benign breast lesions.

Targeted IS-PRM Method Development. The first step in targeted proteomics method development is identification of proteotypic peptides. Proteotypic peptides representing each of the 1314 candidate biomarker proteins were identified from among peptides empirically observed in the PDX plasma biomarker discovery data. In order to use at least three proteotypic peptides per protein (with the exception of keratins and IgG), selections from the empirical data set were supplemented by additional peptides from Peptide Atlas[37] (http://www.peptideatlas.org/) and/or SRMAtlas[41] (http://www.srmatlas.org/) (n = 2122) and peptides identified from previous assay development efforts (n = 208).[42] In total, we synthesized heavy stable isotope-labeled standards (SIS) for 5176 target peptides (Table S1), with ≥3 peptides identified for 1303 (99%) candidate biomarker proteins (Figure 2).

The IS-PRM (Figure S1) assay quantifies endogenous ("light") peptide after first observing and identifying its cognate spiked-in isotope-labeled ("heavy") internal standard peptide. After a positive identification is confirmed, quantification is achieved by performing targeted PRM on the endogenous peptide. This method accomplishes high sensitivity and specificity with improved multiplexing (required to prioritize large numbers of biomarker candidates) by using fast MS scans for identification and maximizing the time devoted to quantitative scans, improving the efficiency of the acquisition

https://doi.org/10.1021/acs.analchem.1c04382
Anal. Chem. 2022, 94, 9540−9547

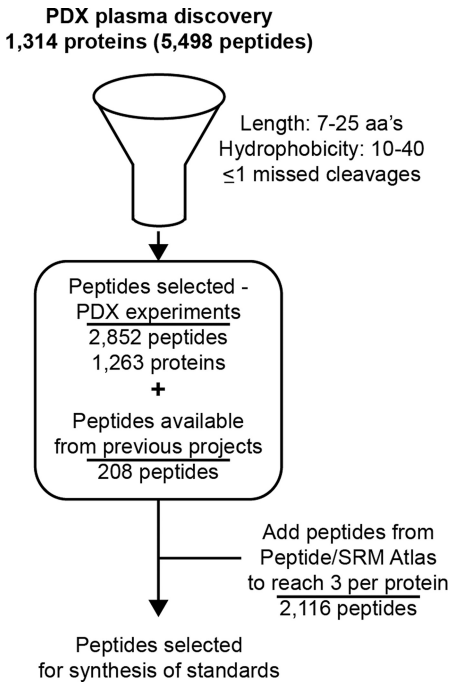


Figure 2. Summary of peptide selection for targeted proteomics. Peptides were selected from three sources to obtain at least three peptides per protein: (i) those directly observed in the PDX discovery experiments, (ii) peptides available in-house from previous projects, and (iii) peptides from the online databases Peptide Atlas (http://www.peptideatlas.org/) and SRMAtlas (http://www.srmatlas.org/). For selection, peptides had to be between 7 and 25 amino acids in length, have a hydrophobicity score between 10 and 40, and have no more than one missed cleavage. A total of 1303 of the candidate biomarkers are represented by three or more peptides per protein; proteins with less than three peptides were keratins and IgGs.

cycle. In addition, the inclusion lists employed by the method can survey for tens of thousands of target precursors making the method easier to implement because it does not require characterization and monitoring of retention time windows.

Analytical Performance of the IS-PRM Assay. In order to optimize the method, we used synthetic peptides to determine the optimum precursor m/z for the IS-PRM inclusion list (Table S2), the intensity thresholds for triggering the identification scan (Table S2), and the fragment ions to be used for peptide identification (Table S3). All targeted peptides were incorporated into a single IS-PRM method. The analytical performance of the IS-PRM method was characterized to ensure sufficient linear dynamic range, sensitivity, and precision for biomarker prioritization studies. We prepared a response curve consisting of a 10-fold serial dilution of human cell (MCF10A) lysate into yeast lysate (100% MCF10A to 0.1%, blanks were prepared using 100% yeast lysate). The MCF10A concentration levels corresponded to an approximate MCF10A cell count of 200000 to 200 cells. Each concentration point underwent proteolytic digestion, addition of SIS peptides, and separation into six bRP fractions. Each fraction was analyzed by the IS-PRM method using triplicate injections (Figure S2a). Peptides meeting the following criteria in at least two of the three replicates were classified as quantified: (i) at least four transitions (light endogenous peptides) or five transitions (heavy SIS peptides) were present in the MS2 spectra, (ii) the ratio dot product of MS2 spectra from heavy and light peptides was >0.98, (iii) at least five points across the peak were profiled in the chromatogram, and (iv) the peak area was >5000. Integrations were manually checked, and 93 (~2%) peptides had a fragment ion with interference in either the heavy or light peptide, which was subsequently removed from the analysis.

Response curve results for the IS-PRM assay are summarized in Figure 3 and data are provided in Table S4. The IS-PRM method triggered the quantification of >98% of the light peptides (Figure 3a). Endogenous signals were quantified for nearly half of the peptides (n = 2443) at the highest

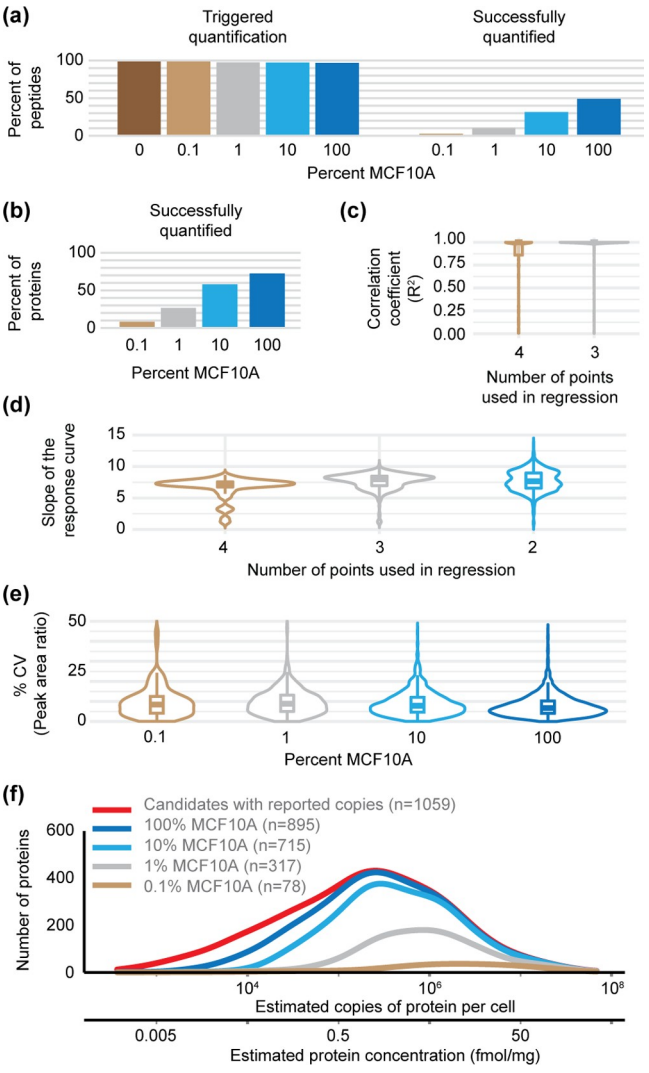


Figure 3. Characterization of the IS-PRM analytical performance. (a) Percent of peptides that triggered quantification (heavy peptides meeting the detection threshold and fragment ion requirement) and were successfully quantified (endogenous peptides meeting all quantification criteria with signal >2× the maximum signal in the blanks). (b) Percentage of the 1314 targeted proteins that were quantified. (c) Distribution of the correlation coefficients squared (R^2) for quantified peptides using the top three (100, 10, and 1% MCF10A) or all four concentration points of the curve. (d) Distribution of slopes for peptides successfully quantified using the top two, three, or all four concentration points. (e) Precision of the replicates of heavy to light peak area ratios for each dilution point. For violin plots, the bold line shows median, box shows inner quartile, vertical line shows 5−95 percentile, density of measurements is indicated by the thin line. (f) Distribution of the number of proteins detected according to the protein level per cell (as reported in Ly et al.).

https://doi.org/10.1021/acs.analchem.1c04382
Anal. Chem. 2022, 94, 9540−9547

concentration of MCF10A (Figure 3a), with an even distribution across the fractions (Figure S2b), resulting in quantification of 953 proteins (73%) of the targeted proteins (Figure 3b). Decreasing the percentage of MCF10A cells resulted in the expected decrease in proteins quantified. The assay panel exhibited excellent linearity across either three or 4 orders of magnitude, with a median $R^2 > 0.999$ (Figure 3c) and a median slope of 7.7 (within 25% of the expected slope of 10; Figure 3d). The method exhibited excellent analytical precision, with a median coefficient of variation (% CV) of 7.7% across all concentration points (Figure 3e). To estimate the sensitivity of the method for detection of low abundance proteins, we used the number of proteins expressed per cell reported in Ly et al.[43] Figure 3f shows a histogram of proteins detected by the IS-PRM method in each dilution point versus the number of proteins per cell. As expected, as the MCF10A cells were diluted, the histogram curve shifts to those proteins that were most abundant. The IS-PRM assay meets Tier 2 requirements[44] by operating at moderate throughput, using internal standards for each analyte, maintaining high specificity through MS2 spectra and PRM transitions, and showing high reproducibility and precision in triplicate analysis of response curves.

Evaluation of the IS-PRM Method for Highly Multiplexed Quantification of Candidate Biomarkers in Human Plasma. We next applied the 5176-plex IS-PRM assay to quantify the biomarker candidates in human plasma from women diagnosed with breast cancer and human plasma from women diagnosed with benign breast lesions (Figure S3a), with a goal of rank-ordering the 1314 candidate biomarkers to identify those meriting further evaluation in larger, case-control validation studies. One challenge in measuring low abundance plasma proteins is the extensive sample preparation required. To measure low abundance proteins we incorporated abundant plasma protein depletion and bRP fractionation, which limited the analytical throughput for analyzing large numbers of samples and necessitated the use of plasma pools instead of individual plasma samples. To address a potential limitation of using pooled samples, where a single outlier patient in one plasma sample can skew the biomarker results from that pool, we devised an experiment to analyze multiple independent pools, each of which includes multiple patients, and aggregate the results.

Samples were assigned to three pools from women diagnosed with breast cancer and three pools from women diagnosed with benign breast lesions, with each pool representing 19−20 women (Figure S3a). A total of 200 μg of each plasma pool underwent reduction, alkylation, and proteolytic digestion. The digested plasma pools were desalted, spiked with all 5176 SIS peptides (~500 fmol), and fractionated into 24 fractions using bRP chromatography.

The IS-PRM method was applied to each of the 24 fractions (per pool), peak integrations were manually reviewed, and interferences removed from 193 (~4%) peptides. Summed transition areas and number of transitions and points per peak are reported in Table S5. In addition to the quantification criteria used for the response curve (above), we required an endogenous signal to be >2× the signal from blank runs (Table S6). On average, endogenous signals were measured for 1708 (33%) of the target peptides (Figure 4a) across the pools, with an even distribution across the fractions (Figure S3b), corresponding to 760 (58%) proteins (Figure 4b). The sum of all proteins quantified across the plasma pools was 893 (68%). Technical variability was estimated using the endogenous

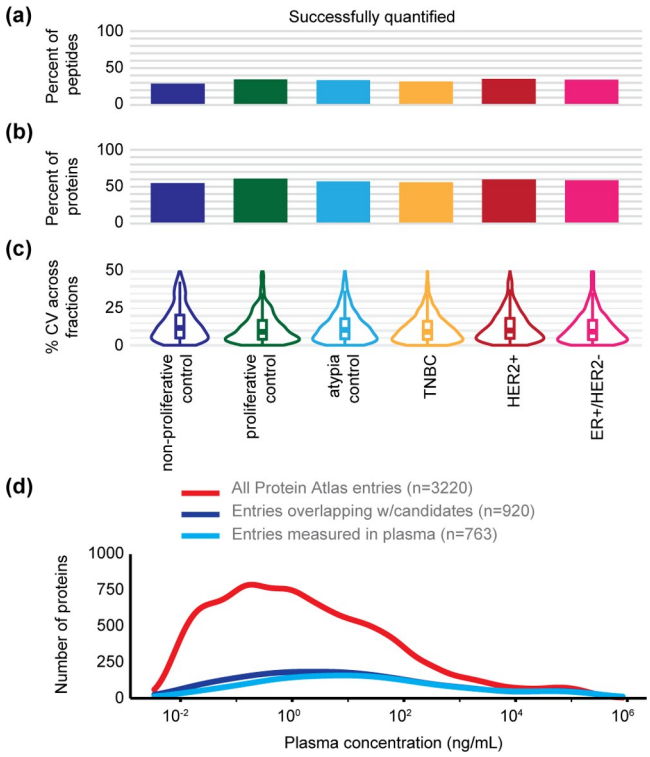


Figure 4. Applying the IS-PRM assay to prioritize the biomarker candidate proteins in plasma of human breast cancer patients. (a) Percent of endogenous light signals meeting quantification criteria with a signal >2× the maximum signal in the blanks in each of the plasma pools. (b) Percent of candidate protein biomarkers with endogenous levels measured in each of the plasma pools. (c) Violin box plot showing the technical variability of the replicate measurements of the heavy to light peak area ratios (PAR), measured by using the PAR in neighboring bRP fractions as technical replicates. Bold line shows median, box shows inner quartile, vertical line shows 5−95 percentile, density of measurements is indicated by the thin line. (d) Distribution of the number of proteins detected according to reported plasma concentration.

measurements in the neighboring bRP fraction as a technical replicate (n = 2). Heavy and light peak areas varied between fractions, but PAR should remain constant. Figure 4c shows the distribution of % CV for each plasma pool (median across all measurements = 11%). Using the plasma concentration for proteins reported in the Human Plasma Peptide Atlas[45] and the median of multiple peptide measurements per protein, we estimated the range of plasma concentrations for the proteins quantified by the IS-PRM assay. Figure 4d shows the distribution of protein concentrations for the candidate biomarkers with concentrations extending to below the ng/mL level. A rigorous QC program was implemented to avoid any system degradation during the analysis (Figure S4).

The overall distributions of the 893 candidate biomarker protein abundances across the six plasma pools, shown in Figure 5a, varied widely. To determine if the candidate biomarker protein signals were higher in the cancer plasma pools, we tested the proteins for significant differences ($p < 0.001$) in each cancer pool compared to at least two of the confounding (i.e., benign) control plasma pools. To identify candidates correlating with biological progression, we used a regression trend approach, which accounted for measurements that were lowest in the nonproliferative control, increasing in the proliferative and the atypia controls, and reaching a maximum in the cancer subtype

https://doi.org/10.1021/acs.analchem.1c04382
Anal. Chem. 2022, 94, 9540−9547

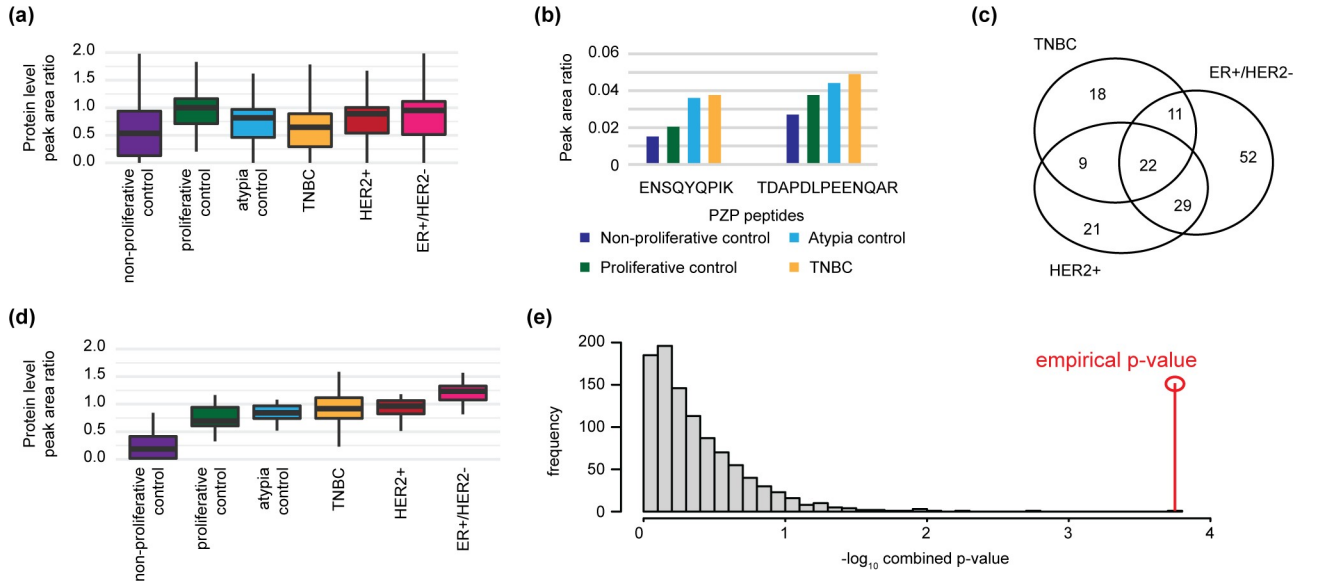


Figure 5. Verification of candidate biomarkers in the breast cancer plasma pools. (a) Endogenous levels in the depleted and fractionated plasma pools, reported as the peak area ratio (PAR; light/heavy) using the median value from multiple measurements of peptides. (b) PAR of quantified peptides from PZP in three control pools and triple negative breast cancer (TNBC). An example of a candidate biomarker meeting significance testing in the TNBC breast cancer subtype with endogenous levels significantly higher ($p < 0.001$) in cancer compared to at least two of the three confounding control plasma pools with a significant regression trend ($p < 0.01$; trend comparing nonproliferative control → proliferative control → atypia control → cancer subtype). (c) Venn diagram of the 162 candidate biomarkers verified in the pooled case/control study. A total of 22 of the candidates passed both cutoffs (p-value < 0.001 and regression trend p-value < 0.1) in all three breast cancer subtypes. (d) Endogenous levels of the 22 proteins found higher in all three breast cancer subtypes compared to the three confounding control samples. For box plots, bold line shows median, box shows inner quartile, vertical line shows 5−95 percentile. (e) Combined p-value from differentiation between cancer subtypes and control plasma pools using randomly sampled subsets of 22 proteins (1000 permutations). The p-value for the set of overlapping 22 proteins verified by IS-PRM assay (p = 0.00016) is shown by the red line. p-Values are based on a student t test.

sample (i.e., candidates whose plasma levels progressively increased as the biology of the breast lesions became more aggressive). An example of an individual protein featuring this trend is shown in Figure 5b, and the results for all proteins are reported in Table S7. Two peptides for PZP show consistent relative quantification (Figure 5b), where the lowest measurement is seen in the nonproliferative control, followed by the proliferative and atypia controls, and finally the TNBC cancer subtype sample shows the highest levels. Overall, there were 162 candidate proteins showing significant differences in at least one of the three cancer subtypes (triple negative, HER2 positive and ER positive/HER2 negative), and 22 were significant in all three (Figure 5c). The distribution of measured abundances for the 22 overlapping proteins (Figure 5d) reflects an improvement in differentiating the cancer pools from control (compared to total proteins measured). Compared to a random sampling of 22 proteins, the candidates overlapping from all three subtypes show a better differentiation from controls (Figure 5e).

Follow-up studies will be required to determine if the PDX models enabled discovery of clinically translatable biomarkers. Patient-derived xenografts of human cancer have emerged as powerful tools for clinical/translational science due to their recapitulation of many aspects of the biology of tumors derived from patients, including treatment responses, genomic mutation and copy number alterations, as well as RNA and protein expression.[46−48] This high degree of biological consistency with clinical samples may make PDX-bearing mice a potential discovery platform for identification of tumor-derived proteins in plasma, where human sequences can be distinguished from mouse peptides using mass spectrometry.[49,50]

CONCLUSIONS

We demonstrate that IS-PRM can be deployed on plasma samples to credential large numbers (i.e., thousands) of candidate plasma biomarkers for follow up validation studies. The method was capable of targeting >1300 proteins in a highly reproducible manner, measuring >800 proteins in human plasma over several orders of magnitude with high specificity and sensitivity, far exceeding the scale of previous demonstration studies. Endogenous measurements across six human plasma pools included 200 proteins with reported plasma concentrations <1 ng/mL. As expected,[51−53] differences in protein expression levels between case and control pools were relatively small, highlighting the need for highly precise measurements (and perhaps multiprotein panels and longitudinal sampling of individual patients over time)[54] to provide clinical diagnoses. IS-PRM quantification showed excellent analytical precision in both the triplicate analysis of a response curve (median % CV of 7.7%) and the analysis of neighboring fractions of the human plasma pools (median % CV across fractions of 11.0%), proving the method is capable of high precision. Overall, this workflow was able to interrogate the list of candidates and prioritize a subset for follow up studies that is more amenable to workflows like multiplex immuno-MRM,[11,55−57] which can be used to support clinical validation studies in a high throughput manner for the most promising candidates.

ASSOCIATED CONTENT

Supporting Information

. The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.1c04382.

https://doi.org/10.1021/acs.analchem.1c04382
Anal. Chem. 2022, 94, 9540−9547

Expanded materials and methods; Figures S1−S4: The IS-PRM workflow; The workflow for the characterization of the analytical performance of the IS-PRM assay; A summary of the results applying the IS-PRM assay to plasma of human breast cancer patients; Levey−Jennings plot of peak areas and retention times from a representative yeast QC LC-MS/MS analysis (PDF)

Tables S1−S7: Candidate biomarkers and selected proteotypic peptides; The precursor m/z and intensity thresholds used in the IS-PRM method; The fragment ion, charge state, and m/z used in the IS-PRM method; IS-PRM results from MCF10A dilution curves; IS-PRM results from pooled depleted human plasma; Light to heavy peak area ratios (PAR) from pooled depleted human plasma; The relative PAR for each candidate biomarker, the z-score and the p-values for each cancer subtype compared to 2/3 controls and compared using the regression trend approach (XLSX)

AUTHOR INFORMATION

Corresponding Author

Amanda G. Paulovich − *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States;* orcid.org/0000-0001-6532-6499; Phone: 206-667-1912; Email: apaulovi@fredhutch.org; Fax: 206-667-2277

Authors

Jacob J. Kennedy − *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States;* orcid.org/0000-0003-2532-3861

Jeffrey R. Whiteaker − *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States;* orcid.org/0000-0001-5042-8580

Richard G. Ivey − *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States*

Aura Burian − *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States*

Shrabanti Chowdhury − *Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States*

Chia-Feng Tsai − *Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0002-6514-6911

Tao Liu − *Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0001-9529-6550

ChenWei Lin − *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States*

Oscar D. Murillo − *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States*

Rachel A. Lundeen − *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States*

Lisa A. Jones − *Proteomics and Metabolomics Shared Resources, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States*

Philip R. Gafken − *Proteomics and Metabolomics Shared Resources, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States*

Gary Longton − *Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States*

Karin D. Rodland − *Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0001-7070-6541

Steven J. Skates − *MGH Biostatistics Center, Harvard Medical School, Boston, Massachusetts 02114, United States*

John Landua − *Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, Texas 77030, United States*

Pei Wang − *Department of Genetics and Genomic Sciences, Mount Sinai Hospital, New York, New York 10065, United States*

Michael T. Lewis − *Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, Texas 77030, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.1c04382

Author Contributions

#These authors contributed equally to this work.

Author Contributions

J.J.K., J.R.W., R.G.I., A.B., S.C., C.-F.T., T.L., C.W.L., O.M., R.L., L.A.J., P.R.G., G.L., K.D.R., S.S., J.L., P.W., and A.G.P. acquired, analyzed, or interpreted the data. G.L., K.D.R., S.S., J.L., P.W., M.T.L., and A.G.P. supervised studies. J.J.K., J.R.W., and A.G.P. drafted and revised the manuscript. S.S., J.L., P.W., M.T.L., and A.G.P. conceived of and designed the work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Kevin Schauer, Aaron Gajadhar, and Sebastien Gallien from ThermoFisher Scientific for technical support in configuring the method. This research was funded by the National Cancer Institute (NCI) Early Detection Research Network (EDRN) under Grant No. U01CA214172, the NCI Research Specialist program (Grant No. R50CA211499), and a generous donation from the Aven Foundation. The Proteomics and Metabolomics shared resource of the Fred Hutch/University of Washington Cancer Consortium is supported by NCI Grant P30 CA015704. Scientific Computing Infrastructure at Fred Hutch was funded by ORIP Grant S10OD028685. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Figure 1 and the TOC graphic were created with BioRender.com.

REFERENCES

(1) Bhawal, R.; Oberg, A. L.; Zhang, S.; Kohli, M. *Cancers (Basel)* **2020**, *12* (9), 2428.
(2) Landegren, U.; Hammond, M. *Mol. Oncol* **2021**, *15* (6), 1715−1726.
(3) Keshishian, H.; Burgess, M. W.; Specht, H.; Wallace, L.; Clauser, K. R.; Gillette, M. A.; Carr, S. A. *Nat. Protoc* **2017**, *12* (8), 1683−1701.
(4) Omenn, G. S.; et al. *Proteomics* **2005**, *5* (13), 3226−45.
(5) Geyer, P. E.; Holdt, L. M.; Teupser, D.; Mann, M. *Mol. Syst. Biol.* **2017**, *13* (9), 942.

(6) Duffy, M. J.; Sturgeon, C. M.; Soletormos, G.; Barak, V.; Molina, R.; Hayes, D. F.; Diamandis, E. P.; Bossuyt, P. M. *Clin Chem.* **2015**, *61* (6), 809−20.
(7) Whiteaker, J. R.; et al. *Nature biotechnology* **2011**, *29* (7), 625−34.
(8) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. *Mol. Syst. Biol.* **2008**, *4*, 222.
(9) Pan, S.; Aebersold, R.; Chen, R.; Rush, J.; Goodlett, D. R.; McIntosh, M. W.; Zhang, J.; Brentnall, T. A. *J. Proteome Res.* **2009**, *8* (2), 787−97.
(10) Faria, S. S.; Morris, C. F.; Silva, A. R.; Fonseca, M. P.; Forget, P.; Castro, M. S.; Fontes, W. *Front Oncol* **2017**, *7*, 13.
(11) Whiteaker, J. R.; Zhao, L.; Abbatiello, S. E.; Burgess, M.; Kuhn, E.; Lin, C.; Pope, M. E.; Razavi, M.; Anderson, N. L.; Pearson, T. W.; Carr, S. A.; Paulovich, A. G. *Mol. Cell Proteomics* **2011**, *10* (4), M110 005645.
(12) Rifai, N.; Gillette, M. A.; Carr, S. A. *Nature biotechnology* **2006**, *24* (8), 971−83.
(13) Chen, J.; Zheng, N. *Bioanalysis* **2020**, *12* (20), 1469−1481.
(14) Kumar, V.; Ray, S.; Ghantasala, S.; Srivastava, S. *Front Oncol* **2020**, *10*, 543997.
(15) Yoo, M. W.; Park, J.; Han, H. S.; Yun, Y. M.; Kang, J. W.; Choi, D. Y.; Lee, J. W.; Jung, J. H.; Lee, K. Y.; Kim, K. P. *Proteomics* **2017**, *17* (6), 1600332.
(16) Young, M. R.; Wagner, P. D.; Ghosh, S.; Rinaudo, J. A.; Baker, S. G.; Zaret, K. S.; Goggins, M.; Srivastava, S. *Pancreas* **2018**, *47* (2), 135−141.
(17) Ozcan, S.; Cooper, J. D.; Lago, S. G.; Kenny, D.; Rustogi, N.; Stocki, P.; Bahn, S. *Sci. Rep* **2017**, *7*, 45178.
(18) Percy, A. J.; Michaud, S. A.; Jardim, A.; Sinclair, N. J.; Zhang, S.; Mohammed, Y.; Palmer, A. L.; Hardie, D. B.; Yang, J.; LeBlanc, A. M.; Borchers, C. H. *Proteomics* **2017**, *17* (7), 1600097.
(19) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. *Mol. Cell Proteomics* **2012**, *11* (11), 1475−88.
(20) Gallien, S.; Bourmaud, A.; Kim, S. Y.; Domon, B. *J. Proteomics* **2014**, *100*, 147−59.
(21) Burgess, M. W.; Keshishian, H.; Mani, D. R.; Gillette, M. A.; Carr, S. A. *Mol. Cell Proteomics* **2014**, *13* (4), 1137−49.
(22) Jaffe, J. D.; Keshishian, H.; Chang, B.; Addona, T. A.; Gillette, M. A.; Carr, S. A. *Mol. Cell Proteomics* **2008**, *7* (10), 1952−62.
(23) Addona, T. A.; Shi, X.; Keshishian, H.; Mani, D. R.; Burgess, M.; Gillette, M. A.; Clauser, K. R.; Shen, D.; Lewis, G. D.; Farrell, L. A.; Fifer, M. A.; Sabatine, M. S.; Gerszten, R. E.; Carr, S. A. *Nature biotechnology* **2011**, *29* (7), 635−43.
(24) Gallien, S.; Kim, S. Y.; Domon, B. *Mol. Cell Proteomics* **2015**, *14* (6), 1630−44.
(25) Stopfer, L. E.; Flower, C. T.; Gajadhar, A. S.; Patel, B.; Gallien, S.; Lopez-Ferrer, D.; White, F. M. *Cancer Res.* **2021**, *81* (9), 2495−2509.
(26) Marholz, L. J.; Federspiel, J. D.; Suh, H.; Fernandez Ocana, M. *J. Proteome Res.* **2021**, *20*, 4272−4283.
(27) *Guide for the Care and Use of Laboratory Animals*; The National Academies Collection; National Institutes of Health, 2011.
(28) Lv, X.; Dobrolecki, L. E.; Ding, Y.; Rosen, J. M.; Lewis, M. T.; Chen, X. *J. Vis. Exp.* **2020**, No. 159, e61173.
(29) Janecki, D. J.; Pomerantz, S. C.; Beil, E. J.; Nemeth, J. F. *J. Chromatogr B Analyt Technol. Biomed Life Sci.* **2012**, *902*, 35−41.
(30) Shi, T.; Fillmore, T. L.; Sun, X.; Zhao, R.; Schepmoes, A. A.; Hossain, M.; Xie, F.; Wu, S.; Kim, J. S.; Jones, N.; Moore, R. J.; Pasa-Tolic, L.; Kagan, J.; Rodland, K. D.; Liu, T.; Tang, K.; Camp, D. G., 2nd; Smith, R. D.; Qian, W. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (38), 15395−400.
(31) Wang, L. B.; et al. *Cancer Cell* **2021**, *39* (4), 509−528.
(32) Ziv, I.; Matiuhin, Y.; Kirkpatrick, D. S.; Erpapazoglou, Z.; Leon, S.; Pantazopoulou, M.; Kim, W.; Gygi, S. P.; Haguenauer-Tsapis, R.; Reis, N.; Glickman, M. H.; Kleifeld, O. *Mol. Cell Proteomics* **2011**, *10* (5), M111.009753.
(33) Feng, Z.; Kagan, J.; Pepe, M.; Thornquist, M.; Ann Rinaudo, J.; Dahlgren, J.; Krueger, K.; Zheng, Y.; Patriotis, C.; Huang, Y.; Sorbara, L.; Thompson, I.; Srivastava, S. *Clin Chem.* **2013**, *59* (1), 68−74.
(34) Marks, J. R.; Anderson, K. S.; Engstrom, P.; Godwin, A. K.; Esserman, L. J.; Longton, G.; Iversen, E. S.; Mathew, A.; Patriotis, C.; Pepe, M. S. *Cancer Epidemiol Biomarkers Prev* **2015**, *24* (2), 435−41.
(35) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. *Nat. Methods* **2017**, *14* (5), 513−520.
(36) Lam, H.; Deutsch, E. W.; Eddes, J. S.; Eng, J. K.; Stein, S. E.; Aebersold, R. *Nat. Methods* **2008**, *5* (10), 873−5.
(37) Deutsch, E. W.; et al. *Nucleic Acids Res.* **2020**, *48* (D1), D1145−D1152.
(38) Sharma, V.; Eckels, J.; Schilling, B.; Ludwig, C.; Jaffe, J. D.; MacCoss, M. J.; MacLean, B. *Mol. Cell Proteomics* **2018**, *17* (6), 1239−1244.
(39) Mertins, P.; et al. *Nature* **2016**, *534* (7605), 55−62.
(40) van Bentum, M.; Selbach, M. *Mol. Cell Proteomics* **2021**, *20*, 100165.
(41) Kusebauch, U.; et al. *Cell* **2016**, *166* (3), 766−778.
(42) Kennedy, J. J.; et al. *Nat. Methods* **2014**, *11* (2), 149−55.
(43) Ly, T.; Endo, A.; Brenes, A.; Gierlinski, M.; Afzal, V.; Pawellek, A.; Lamond, A. I. *Wellcome Open Res.* **2018**, *3*, 51.
(44) Carr, S. A.; et al. *Mol. Cell Proteomics* **2014**, *13* (3), 907−17.
(45) Schwenk, J. M.; Omenn, G. S.; Sun, Z.; Campbell, D. S.; Baker, M. S.; Overall, C. M.; Aebersold, R.; Moritz, R. L.; Deutsch, E. W. *J. Proteome Res.* **2017**, *16* (12), 4299−4310.
(46) Sun, H.; et al. *Nat. Commun.* **2021**, *12* (1), 5086.
(47) Woo, X. Y.; et al. *Nat. Genet.* **2021**, *53* (1), 86−99.
(48) Petrosyan, V.; et al. *BioRxiv 2021.08.20.457116* **2021**, 1.
(49) Tang, H. Y.; Beer, L. A.; Chang-Wong, T.; Hammond, R.; Gimotty, P.; Coukos, G.; Speicher, D. W. *J. Proteome Res.* **2012**, *11* (2), 678−91.
(50) Sinha, A.; Hussain, A.; Ignatchenko, V.; Ignatchenko, A.; Tang, K. H.; Ho, V. W. H.; Neel, B. G.; Clarke, B.; Bernardini, M. Q.; Ailles, L.; Kislinger, T. *Cell Syst.* **2019**, *8* (4), 345−351.
(51) Anderson, N. L.; Anderson, N. G. *Mol. Cell Proteomics* **2002**, *1* (11), 845−67.
(52) Liu, Y.; Buil, A.; Collins, B. C.; Gillet, L. C.; Blum, L. C.; Cheng, L. Y.; Vitek, O.; Mouritsen, J.; Lachance, G.; Spector, T. D.; Dermitzakis, E. T.; Aebersold, R. *Mol. Syst. Biol.* **2015**, *11* (1), 786.
(53) Stenemo, M.; Teleman, J.; Sjostrom, M.; Grubb, G.; Malmstrom, E.; Malmstrom, J.; Nimeus, E. *Proteomics* **2016**, *16* (13), 1928−37.
(54) Berek, J. S.; Bast, R. C., Jr. *Cancer* **1995**, *76* (S10), 2092−6.
(55) Whiteaker, J. R.; Zhao, L.; Anderson, L.; Paulovich, A. G. *Mol. Cell Proteomics* **2010**, *9* (1), 184−96.
(56) Whiteaker, J. R.; Zhao, L.; Yan, P.; Ivey, R. G.; Voytovich, U. J.; Moore, H. D.; Lin, C.; Paulovich, A. G. *Mol. Cell Proteomics* **2015**, *14* (8), 2261−73.
(57) Zhang, B.; Whiteaker, J. R.; Hoofnagle, A. N.; Baird, G. S.; Rodland, K. D.; Paulovich, A. G. *Nat. Rev. Clin Oncol* **2019**, *16* (4), 256−268.

https://doi.org/10.1021/acs.analchem.1c04382
Anal. Chem. 2022, 94, 9540−9547