Targeted Quantification of Proteoforms in Complex Samples by Proteoform Reaction Monitoring

Che-Fan Huang,[1] Jake T. Kline,[2] Fernanda Negrão,[1+] Matthew T. Robey,[1,3] Timothy K. Toby,[1] Kenneth R. Durbin,[1,3] Ryan T. Fellers,[1,3] John J. Friedewald,[4] Josh Levitsky,[4] Michael M. I. Abecassis,[4$] Rafael D. Melani,[1#] Neil L. Kelleher,[1] Luca Fornelli[2,5]*.

[1] Departments of Molecular Biosciences, Chemistry, and the Feinberg School of Medicine, Northwestern University; Evanston, IL, USA.

[2] Department of Biology, University of Oklahoma, Norman, OK, USA

[3] Proteinaceous, Inc., Evanston, IL, USA

[4] Comprehensive Transplant Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[5] Department of Biology, University of Oklahoma, Norman, OK, USA

[+]current address: Sciex, Golden, CO, USA

[$]current address: University of Arizona, Tucson, AZ, USA

[#]current address: Thermo Fisher Scientific, San Jose, CA, USA

ABSTRACT

Existing mass spectrometric assays used for sensitive and specific measurements of target proteins across multiple samples, such as selected/multiple reaction monitoring (SRM/MRM) or parallel reaction monitoring (PRM), are peptide-based methods for bottom-up proteomics. Here, we describe an approach based on the principle of PRM for the measurement of intact proteoforms by targeted top-down proteomics termed Proteoform Reaction Monitoring (PfRM). We explore the ability of our method to circumvent traditional limitations of top-down proteomics such as sensitivity and reproducibility. We also introduce a new software, Proteoform Finder, specifically designed for easy analysis of PfRM data. PfRM was initially benchmarked by quantifying three standard proteins. Linearity of the assay was shown over almost three orders of magnitude in the femtomole range. We later applied our multiplexed PfRM assay to complex samples to quantify biomarker candidates in peripheral blood mononuclear cells (PBMCs) from liver transplanted patients, demonstrating its possible translational applications. These results demonstrate that PfRM has the potential to contribute to the accurate quantification of protein biomarkers for diagnostic purposes and to improve understanding of disease etiology at the proteoform

level.

INTRODUCTION

In translational sciences, quantifying proteins is necessary to characterize alterations to homeostasis associated with aberrant phenotypes. Over the last two decades, dozens of strategies have been established to quantify proteins extracted from cells and tissues using mass spectrometry (MS), taking advantage of its sensitivity, accuracy, and high throughput. Initially, quantitative MS-based proteomics was developed following the peptide-centric, bottom-up approach.[1] Quantitative bottom-up proteomics (BUP) includes both untargeted (or discovery-mode) and targeted methods (which require a priori knowledge).[2] While untargeted methods have developed into two categories (i.e., label-free quantification and labeling-based techniques),[3] targeted methods almost exclusively follow the original concept of selected reaction monitoring (SRM).[4] This MS-based assay is traditionally performed using a triple quadrupole mass spectrometer, where quadrupole-1 is used to select the peptide precursor ion of interest, quadrupole-2 is utilized as a fragmentation cell, and quadrupole-3 selects a single fragment ion for detection and abundance determination. The dual selection of precursor and fragment ion, known as transition, ensures high specificity and selectivity of this assay, even when the targeted analyte is part of a complex mixture. The

SRM method evolved into multiple reaction monitoring (MRM).[5] In MRM, multiple transitions are selected for a given peptide of interest, further improving specificity and quantification accuracy. In 2012, Coon and co-workers introduced parallel reaction monitoring (PRM), where targeted quantification of peptides is performed using mass spectrometers equipped with high-resolution mass analyzers.[6] Originally, a hybrid quadrupole-Orbitrap mass spectrometer (Q Exactive) was utilized. The selection of fragment ions of interest performed by quadrupole-3 in SRM/MRM is replaced in PRM experiments by the acquisition of high-resolution tandem mass spectra. High mass accuracy in Orbitrap Fourier transform MS (FTMS)[7] allows for the simultaneous identification and abundance determination of multiple fragment ions. While data acquisition in Orbitrap FTMS is not as fast as in triple quadrupole instruments, the detection of multiple product ions in a single tandem mass spectrum makes the PRM data acquisition cycle comparable with the overall cycle of an MRM assay.

Thanks to their sensitivity and robustness, MRM and PRM techniques are being applied to a consistently growing number of clinical studies.[8, 9] As a notable example, advances in peptide separation by reversed-phase liquid chromatography (LC) and MS technology have allowed for the quantification

of up to 267 protein families in human plasma, including 61 FDA-approved biomarkers.[10]

Despite the success of MRM and PRM and the enthusiasm brought by recent work by MacCoss and co-workers, which demonstrates that PRM can be performed even using low-resolution (yet highly sensitive) mass analyzers such as ion traps without compromising quantification accuracy,[11] these quantification methods inevitably suffer from the protein inference problem that inherently affects any bottom-up proteomic strategy.[12] Undoubtedly, proteome changes are often caused by variations in the global expression levels of genes, making MRM and PRM assays relevant in clinical studies.[13] However, an increasing number of observations indicates that cellular mechanisms may be regulated not through drastic changes in the global abundance of a given gene product, but rather by modifying translated polypeptides in a specific manner to alter their function.[14-20] Therefore, quantifying proteoforms, or protein forms carrying a distinct set of chemical and genetic modifications,[21] could provide relevant biological information otherwise hard or impossible to obtain using peptide-based quantification strategies. Currently, the quantitative analysis of proteoforms is performed primarily in "discovery mode". Following the path previously traced by BUP,

methods for quantitative top-down proteomics (TDP)[22] in the context of high-throughput studies include both label-free[23, 24] and label-based strategies (the latter comprising both metabolic as well as chemical labeling).[25-28] Examples of targeted quantification of proteoforms have also been described. However, these studies often involve pools of highly abundant proteoforms,[29] are performed using intensity of intact proteoform ions rather than fragment ions,[30] or require lengthy sample preparation procedures (e.g., enrichment steps) to reduce background complexity and boost the signal of the targets.[31] Therefore, all these strategies substantially differ from the large panels of BUP MRM/PRM assays mentioned above.

A few notable exceptions exist: the body of work by Schug and co-workers demonstrates the applicability of MRM performed on a triple quadrupole mass spectrometer for the targeted quantification of intact proteins.[32-35] This research direction recently culminated in quantifying eleven intact protein standards spiked into complex biofluids such as human plasma, serum, and urine. Similarly, Wang et al. reported quantifying the intact and truncated form of the ~7 kDa cytokine stromal cell-derived factor 1α using SRM on a triple quadrupole instrument.[36] Despite these successful examples based on low-resolution instruments, high-resolution MS arguably offers the possibility

of better interpreting the complex tandem mass spectra generated by the gas-phase fragmentation of large polypeptides. Yet, the number of reports discussing the use of PRM for whole proteoform quantification (**Figure 1A**) is very limited: Chen et al. demonstrated that PRM could be used to quantify the growth hormone somatropin from human plasma,[37] while Lefebvre et al. applied the same method to the quantification of variants of staphylococcal enterotoxin A enriched via immunocapture.[38]

Here, we describe our attempt to democratize the targeted quantification of whole proteoforms by introducing a complete analytical pipeline. This includes simple molecular weight (MW)-based fractionation of proteins; a PRM-derived assay, termed proteoform reaction monitoring (PfRM); and a new software, Proteoform Finder, dedicated to the analysis of PfRM data. In addition, we benchmarked this new pipeline using both standard proteins – measured alone or spiked in a complex matrix – and endogenous proteoforms extracted from primary human cells, demonstrating protein quantification linearity across over two orders of magnitudes of concentration and the possibility of performing PfRM starting from relatively low quantities (<100 µg) of whole protein extracts.

EXPERIMENTAL SECTION

Protein standards. Purified ubiquitin (Ub), myoglobin (Myo) and carbonic anhydrase (CA) were purchased from Millipore Sigma. Preparation of stock solutions and serial dilutions are detailed in the Supporting Information.

Preparation of *Escherichia coli* lysate with spiked-in standards. The preparation of *E. coli* proteins extracted from whole-cell lysate is described in Supporting Information. Proteins (300 µg) were pre-fractionated by passively eluting proteins from polyacrylamide gels as intact species (PEPPI)[39] prior to PfRM analysis, and the 0-30 kDa fraction was utilized. A detailed PEPPI protocol is available in the Supporting Information. After clean-up via methanol/chloroform/water precipitation,[40] *E. coli* proteins were resuspended in mobile phase A. Standards (Ub 0.25-25 fmol, Myo 1-100 fmol, CA 4-400 fmol) were spiked-in prior to PfRM analysis.

Sample preparation of PBMCs. Standard human peripheral blood mononuclear cells (PBMCs) were purchased from Stem Cell Technologies. PBMC samples from liver transplanted patients were obtained from the Comprehensive Transplant Center (CTC) of Northwestern University. Peripheral blood was collected in BD Vacutainer® CPT™ Mononuclear Cell Preparation tubes (BD Biosciences). PBMCs were isolated according to the

manufacturer's instructions. Cells were preserved in FBS-rich freeze media (80 % fetal bovine serum, 20 % dimethylsulfoxide). One patient from each of these three outcome groups was analyzed in triplicate: acute rejection (AR), acute dysfunction no rejection (ADNR), and transplant excellent (TX). Detailed protocols for PBMC protein extraction are available in the Supporting Information.

Cytochrome C from equine heart (CytC, Sigma Aldrich, 0.25-2 µg) was spiked into PBMCs lysates (25-75 µg) prior to fractionation of proteins <30 kDa using PEPPI.

Liquid chromatography. Proteins were separated using a Dionex Ultimate 3000 UHPLC chromatographic system (Thermo Scientific). Reversed-phase LC was performed either on a commercial column (FlowChip C4, New Objective) or using a setup based on trap and analytical columns in-house packed with PLRP-S stationary phase (Agilent). Column outlets were online coupled to a nanoelectrospray source. The total run time was 60 minutes (for the analysis of standard proteins) or 90 min (for PBMC samples). Additional details are provided in the Supporting Information.

PfRM mass spectrometry analysis. All mass spectrometry measurements were performed on an Orbitrap Eclipse mass spectrometer (Thermo Scientific)

operating in protein mode using 2 mTorr pressure. Transfer capillary temperature was set at 320 °C, ion funnel RF was set at 30%, and in-source fragmentation was set at 15 V. Data were acquired in targeted-MS^2 mode (tMS^2). For standard proteins (pure or spiked into the *E. coli* lysate), spectra were acquired at 60,000 resolving power (at *m/z* 200), normalized AGC target value of 200%, 1200 ms of maximum injection time, and 2 microscans/spectrum. Higher-energy collisional dissociation (HCD) was used with normalized collision energy (NCE) set at 40% to generate MS^2 fragmentation spectra. Precursors were quadrupole isolated using a 2.5 *m/z* unit-wide isolation window. For standard proteins, the following precursor *m/z* isolation window centers were applied: Ub 714.73 *m/z*, Myo 942.73 *m/z*, CA 908.00 *m/z*. For PBMCs samples, the normalized AGC target was increased to 2000%. Precursor *m/z* isolation window centers for 24 immunoproteoforms and CytC are listed in **Table S1**. A scheduled monitor window (8 min) was created for each proteoform as illustrated in **Figure S1**.

Data Processing. Raw files were analyzed with Proteoform Finder (Proteinaceous, Inc.). Sequences of each proteoform were added, and the *m/z* values of potential fragment ions were calculated. The software, which uses an isotope fitting algorithm (i.e., does not require spectral deconvolution), was

operated in PfRM mode using default parameters, which include 10 ppm fragment *m/z* tolerance, 3 ppm tolerance for isotopologues within a single isotope cluster and 0.3 minimum isotope fitting score. After processing the data, Proteoform Finder displays the extracted ion chromatograms (XIC), spectrum matches, and fragmentation maps. An Excel output file containing the area under the curve (AUC) intensities of the selected fragments for each proteoform can be generated.

AUC was plotted against protein amount (fmol) and fitted with a linear calibration curve in GraphPad Prism. The limit of detection (LOD) and limit of quantification (LOQ) were calculated according to the following definitions: LOD = 3.3 x standard error of Y-intercept / best-fit slope; LOQ = 10 x standard error of Y-intercept / best-fit slope.

RESULTS AND DISCUSSION

Top-down PfRM method development. We developed a robust PfRM assay analyzing intact protein standards of increasing molecular weight (8 kDa Ub, 17 kDa Myo and 29 kDa CA). To create a list of precursor ions, we first acquired the broadband mass spectra (MS^1) of the standards (**Figure 1B**, top chromatogram). From the full scan, we selected one charge state (and its respective m/z) for each standard (Ub: m/z 714.73, z = 12; Myo: m/z 942.73, z = 18; CA: m/z 908.00, z = 32), which would serve as precursor ions for the PfRM method. With this information, we added three targeted MS^2 (tMS^2) experiments to the original full scan run. In each tMS^2 experiment, precursor ions were m/z isolated with the quadrupole and fragmented via HCD, with the resulting fragments detected in the Orbitrap. As expected, the XIC of monitored proteoform fragments entirely agreed with the total ion chromatograms from MS^1 (**Figure 1B**, bottom chromatogram). Importantly, even at low protein concentration, the maximum injection time (IT=1200 ms) was not reached and the AGC function was regulating the IT.

Proteoform Finder software. Proteoform Finder is a new software tool with workflows for processing PfRM data (**Figure 1C**). Mass spectrometry data files are input to the software along with proteoform targets. The proteoforms are

specified by their sequence as well as any additional features such as modifications and truncations. An expected retention time range and fragmentation spectrum filters can be set for each proteoform target. Proteoform Finder computes possible fragmentation and the corresponding theoretical isotopic distributions of each fragment ions using the chemical formulas of the proteoform target. The isotopic distributions are matched to each MS^2 spectrum (i.e., no "m/z to mass" deconvolution is needed). Each distribution is assigned a fit score of between 0 to 1, with 1 representing a perfect match to the theoretical distribution. Distributions passing the minimum score cutoff are used as part of the XIC determination (**Figure 1C**, right). By default, the software sorts fragments according to the percentage of files that they appear in and then selects the top three most abundant fragments using the AUC for each proteoform. Each fragment is composed of the theoretical isotopic distributions for all charge states of the fragment ion that fall within an allowable range. Fragments not initially selected can also be user-selected for inclusion in the quantitation. The AUC of each selected fragment XIC belonging to the same proteoform is summed to create final proteoform AUCs. The underlying Proteoform Finder algorithm calculates the XIC using a 'sliding window' approach (i.e., shifting the

averaged spectra from lower to higher retention times), which can help improve fragment ion signal-to-noise ratios. Here, two spectra were averaged in each window.

A representative Proteoform Finder output XIC and MS^2 spectrum of Ub is shown in **Figure S2**. In this example, y_{40}, y_{58} and y_{24} ions were selected for proteoform quantification because they had both the highest intensities and isotopic fit scores. The chromatogram shows that the three selected ions had overlapping elution times and similar peak shapes, as expected for fragmentation products of the same parent ions. The MS^2 spectrum shows the isotopic fitting of the fragment ions. We note that these fragments are relatively large and that several charge states were detected for each ion. For example, for the most abundant fragment y_{40} (monoisotopic mass = 4561.51 Da), we observed 3 charge states in the MS^2 spectrum in **Figure S2** (blue): 1141.39 m/z (z = 4), 913.31 m/z (z = 5) and 761.26 m/z (z = 6). The fragment intensities of each charge state were combined automatically by the software for proteoform quantification without deconvolution.

Determination of PfRM Limits of Detection and Quantification. Our first goal was to determine the LOD and LOQ of the PfRM assay of three widely used intact protein standards. The proteins were injected in triplicates to

create a calibration curve. The amount of protein loaded on the FlowChip C4 column ranged from 0.25-5 femtomoles for Ub, 0.5-10 femtomoles for Myo and 2-20 femtomoles for CA. We found that the LODs were 0.72, 1.13, and 4.16 fmol for Ub, Myo and CA, respectively; and LOQs were 2.18, 3.43, and 12.6 fmol, respectively (**Figure 2**). To demonstrate that this assay is generalizable for different LC setups, we repeated the experiment using a custom-made nanobore column packed with PLRP-S. The obtained results were similar, with low femtomolar LODs and LOQs calculated for all three protein standards (**Figure S3**).

Measurements of standards in *E. coli* lysate. Most biological samples are complex mixtures. Therefore, our next goal was to demonstrate that the PfRM assay could capture proteoform abundance in the femtomolar range using a whole cell lysate as a protein background. We selected *E. coli* proteins for this test. The 0-30 kDa fraction was obtained via PEPPI.[39] Following protein clean-up, Ub, Myo and CA were spiked in at various amounts and injected for LC separation and PfRM analysis. To demonstrate the complexity of this background, a broadband MS^1 spectrum was collected, and the corresponding total ion chromatogram is shown in **Figure 3A**. All three standards were difficult to observe in the MS^1 spectra since co-eluting *E.coli*

proteins were more abundant. However, Proteoform Finder was able to identify and quantify three fragments from each protein standard. The spiked proteins exhibit a linear relationship ($R^2 \geq 0.98$) between their fragment AUCs and the amount loaded on column (**Figure 3B-D**). Importantly, such linearity was maintained across almost three orders of magnitude of protein loading amounts. Compared to traditional MS^1-based top-down quantification of proteoforms,[24] the PfRM method particularly stands out for the quantification of low-abundance proteoforms in complex samples such as cell lysates.

Optimization of multiplexed PfRM analysis of immunoproteoforms. We previously reported that highly valuable information on immune responses related to transplant organ rejections can be derived from the analysis of proteoform abundances in PBMCs.[41, 42] Here, we aim to demonstrate that the PfRM assay can be applied to the quantification of a panel of 24 immunoproteoforms curated from the previous work (**Table S1**).

To avoid sequence similarity interference, we chose cytochrome c (CytC) from equine heart as an internal standard for this set of experiments. Importantly, CytC was spiked into the commercial PBMC samples before PEPPI fractionation to account for variances in the whole workflow. We first added increasing quantities of CytC (0.25-2 μg, corresponding to 21.3 to 170.4 pmol

per sample) to 75 µg PBMCs and performed PEPPI-PfRM. Note that only 1/6 of the recovered PEPPI material is injected in each LC-MS^2 run. CytC was isolated, fragmented and monitored with an isolation window centered at m/z 734.5 (z = 16). We found that the amount of CytC detected was proportional to the amount spiked in (**Figure 4A**), suggesting that the workflow is suitable for quantitatively comparing PBMC proteoforms between different samples and runs.

We next created a scheduled method to detect 24 immunoproteoforms and CytC internal standard in a single LC-MS^2 run of commercial PBMC samples (**Figure S4**). We identified the retention time of each proteoform and narrowed the PfRM monitoring time of each target to an 8 min window using the FlowChip C4 column. This allowed us to improve the data acquisition cycle and increase the MS^2 spectral count of each proteoform target. Upon optimization, the average number of MS^2 spectra for AUC determination was 10. The sequences, m/z of isolation window, and fragments selected for quantifying all 24 immunoproteoforms can be found in **Table S1**.

After optimizing the multiplexed, scheduled PfRM method, increasing quantities of PBMC protein extracts (25, 50, and 75 µg) were supplemented with 1 µg of CytC before PEPPI, and analyzed with PfRM to verify the

extraction yield of PEPPI for the 24 immmunoproteoforms. We found that all 24 targets were detected starting from 75 µg of extract, and they followed the expected abundance pattern, i.e., 75 µg > 50 µg > 25 µg (**Figure 4B** and **Figure S4**). However, this was not a linear relationship likely because the PEPPI workflow includes protein precipitation and protein recovery efficiency generally drops in a non-linear fashion when the input material quantity is particularly low. We found that some lower abundant proteoforms, such as PFR 1027 (translation machinery-associated protein 7) were not well detected and quantified starting from 25 and 50 µg PBMCs. As a point of reference, PEPPI is often performed starting from 100-300 µg of protein.[24, 39] Therefore, these results are not surprising. In conclusion, 75 µg PBMCs with 1 µg CytC is optimal for the PEPPI-PfRM assay of PBMCs proteoforms.

Analysis of PBMC proteoforms from liver-transplant recipients. We acquired PBMCs of liver-transplant recipients categorized according to their transplant outcomes: acute rejection (AR), acute dysfunction no rejection (ADNR), or transplant excellent (TX). We performed PEPPI-PfRM and compared the abundances of all 24 immunoproteoforms between the three groups (**Figure S5**). We observed that the trends were consistent with our previous report.[43] For example, PFR 69028 (actin cytoplasmic 1 – P60709),

PFR 15876 (platelet basic protein – P02775), and PFR 1439 (profilin-1 – P07737) were up-regulated in the TX group both in our previous MS^1-based analysis[43] and the MS^2-based PfRM dataset presented here (**Figure 4C**). We are also excited about the possibility that the PfRM quantification can pick up some nuanced differences between samples more confidently than MS^1-based quantification. For example, PFR 37610 (Histone H2A type 2-B – UniProt accession Q8IUE6) was identified as a potential biomarker that showed differential expression in three transplant outcome groups. However, no significant difference was found in the second set of cohorts for validation.[43] Here, we observed that PFR 37610 is up-regulated in the AR group (**Figure 4C**). We note that in this dataset we only surveyed n=1 in each group and thus biologically significant conclusions cannot be drawn. However, these preliminary results demonstrate the potential of PfRM assay based on its higher sensitivity, specificity, and reproducibility over the more traditional MS^1-based top-down quantitative approach.

Regarding the utility of performing targeted protein quantification at the proteoform level, it is worth noting that over half of the immunoproteoforms in our panel contain one or multiple post-translational modifications, such as initiator methionine removal, acetylation, and phosphorylation (**Table S1**).

Quantifying these PTMs via traditional PRM would require specific tryptic peptide pairs (i.e., with and without modification) and, unless two or more modifications were present in the same peptide, the quantification of proteoforms carrying multiple PTMs would be impossible using a peptide-centric assay. Furthermore, we found that protein truncation is rather common.[43] For example, PFR 69028 shown in **Figure 4C** is a fragment of the actin protein, whose canonical amino acid sequence corresponds to a mass of 42 kDa. Using top-down analysis, we can easily distinguish the 5.2 kDa truncated proteoform from the full-length one. Another example is that PFR 18628 and PFR 18631 in our immunoproteoform panel originate from the platelet factor 4 gene (*PF4* – UniProt accession P02776). PFR 18628 represents the protein's canonical sequence, while PFR 18631 has 4 extra amino acids (FASA) at its N-terminus (**Table S1**). The PfRM assay could distinguish them at two levels. First, distinct precursor isolation windows were used as the intact masses of these two proteoforms differ substantially (Δmass = 376.18 Da). Second, proteoform-specific fragment ions were selected for quantification to eliminate ambiguity. Since PFR 18628 and PFR 18631 are only different in the N-terminus, we utilized *b*-ions to differentiate them together with their different LC retention times (**Figure S6**). We found that

PFR 18628 is the most abundant, with 6 times higher amounts relative to the longer proteoform, in all condition groups. These results are in agreements with our previous publications. Additionally, only ADNR and TX showed differential regulation of the two PF4 proteoforms in this small cohort.

CONCLUSIONS

This work presents a novel approach, PfRM, for the targeted quantification of intact proteoforms. We demonstrate that the assay has femtomole sensitivity and is applicable for quantifying diluted proteoform targets in complex samples such as whole cell lysates. Additionaly, we also introduced a newly developed software, Proteoform Finder, for PfRM data analysis. Our preliminary data on PBMCs from liver-transplant patients is consistent with previously published MS^1-base measurements. The PfRM method will improve the creation of high-value, portable assays to monitor proteoform biology and biomarkers for future diagnostic purposes and clinical research with high quantification accuracy.

ASSOCIATED CONTENT

Supporting Information.

Preparation of protein standards; preparation of *E. coli* lysate; preparation of PBMCs protein extracts; PEPPI-MS protocol; PBMCs scheduled method; Proteoform Finder output chromatogram and spectrum; replications of protein standard calibration curves; 24 immunoproteoform quantifications in standard and patient PBMCs; platelet factor 4 proteoforms distinction; full list of 24 immunoproteoforms, sequences, masses, *m/z* isolation windows and fragments used for quant. (pdf)

AUTHOR INFORMATION

Corresponding Author

*Luca Fornelli: Department of Biology, University of Oklahoma, 730 Van Vleet oval, Richards Hall 417, Norman, Oklahoma 73019; e-mail: luca.fornelli@ou.edu; phone: 405-325-1483; fax: 405-325-6202.

ORCID

Che-Fan Huang: 0000-0002-5799-1533

Jake Kline: 0000-0001-7264-7637

Matthew Robey: 0000-0002-3596-9597

Timothy K. Toby: 0000-0001-9599-7410

Kenneth Durbin: 0000-0001-7163-4877

Rafael D. Melani: 0000-0002-0349-235X

Neil L. Kelleher: 0000-0002-8815-3372

Luca Fornelli: 0000-0001-6334-1046

Notes

The authors declare the following competing financial interest: M.T.R., R.T.F. and K.R.D. work for Proteinaceous Inc., which commercializes the software used for the analysis of top-down mass spectrometry data.

ACKNOWLEDGEMENTS

The authors thank the support from New Objective and their team Mike Lee, Helena Svobodova, Gary Valaskovic, and Emily Ehrenfeld. This work has been supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM147397 to L.F., Award Number R43GM14238 to Proteinaceous, and Award Number P41GM108569 to N.L.K.

Figure legends

Figure 1. (A) Workflow of PfRM-MS of intact proteins. (B) Example of PfRM assay for intact standard proteins. Top: ion trap MS^1 total ion chromatogram of a mixture of Ub, Myo and CA. Bottom: tMS^2 traces of three independent set of fragments specific to each protein. (C) Workflow for PfRM-MS data processing using Proteoform Finder. The inputs to the software include MS data files and a set of proteoforms to quantify. Theoretical fragment ion isotopic distributions are generated for proteoform targets and matched to fragmentation spectra. The matching of isotopic distributions is applied across the file and extracted ion chromatograms (XICs) are produced for each matched fragment to quantify proteoform abundance over time.

Figure 2. Fragment-based quantification of standard proteins at the low femto-mole range. Plots depict triplicate measurements. LOD and LOQ are determined as described in the Experimental Section.

Figure 3. Quantification of protein standards spiked in *E. coli* lysate. (A) Total ion chromatogram from Orbitrap MS^1. The color bars denote the elution time of the corresponding protein standards. (B-D) Calibration curves of Ub (B), Myo (C) and CA (D). All measurements were performed in triplicate. The dotted lines denote 95 % confidence intervals. Error bars: one standard deviation.

Figure 4. (A) 0.25-2.0 µg of CytC were spiked in 75 µg of PBMCs prior to PEPPI extraction and PfRM-MS quantification. (B) Representative immunoproteoforms quantified in 25-75 µg PBMCs by PEPPI-PfRM-MS. (C) Representative immunoproteoforms quantified in PBMCs collected from liver-transplant patients from three different outcome groups (n = 1). All measurements were performed in triplicate. Error bars: one standard deviation.

Figure 1

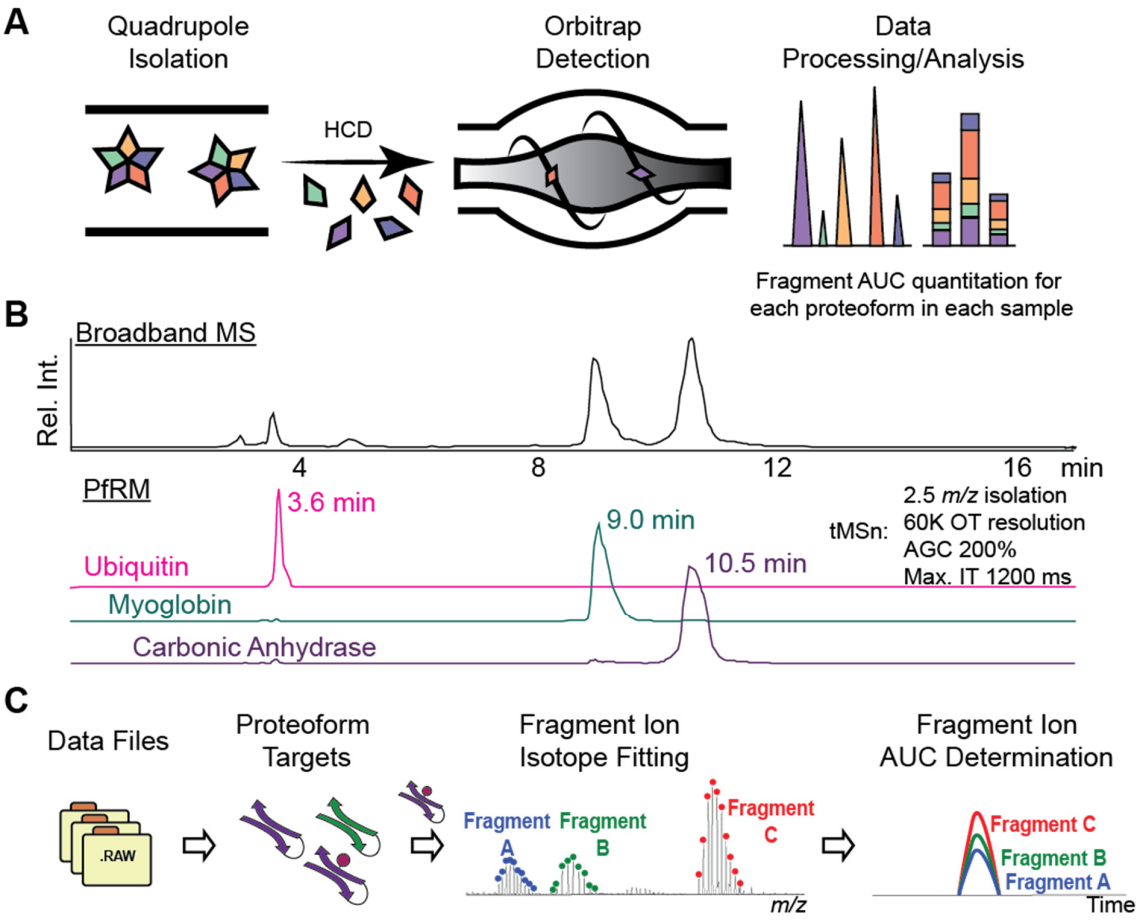


Figure 2.

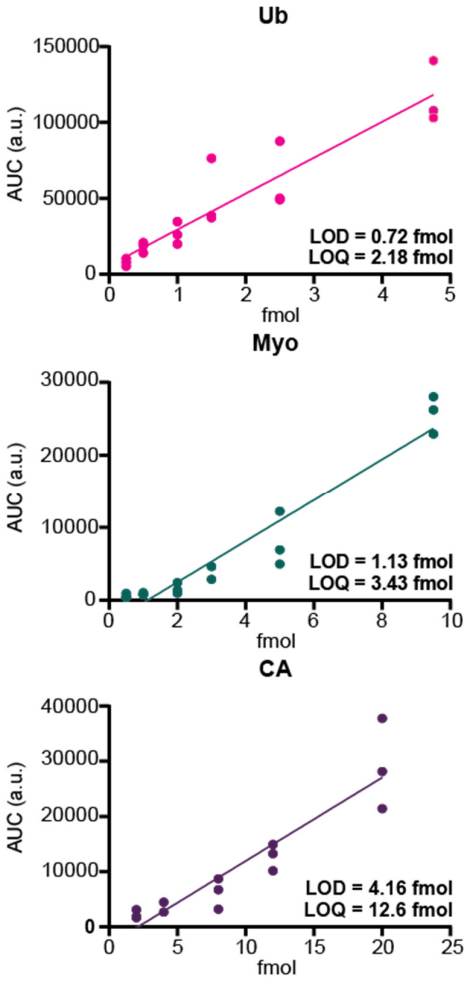


Figure 3.

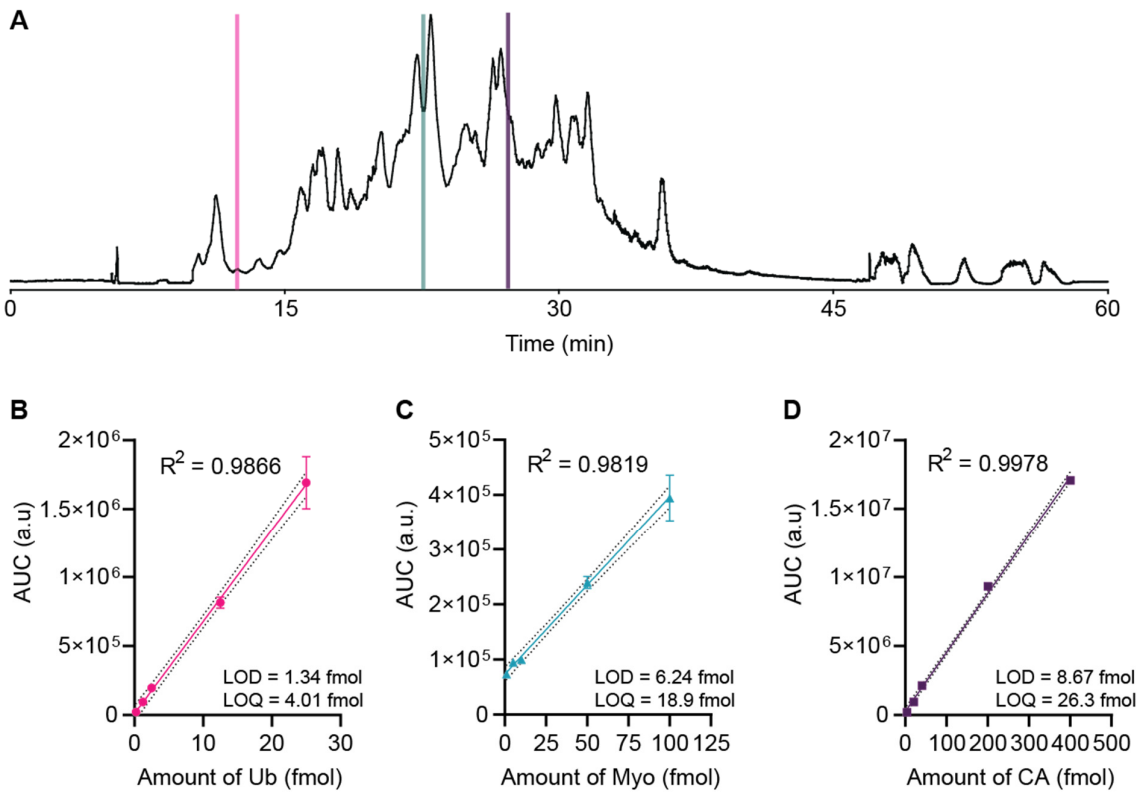
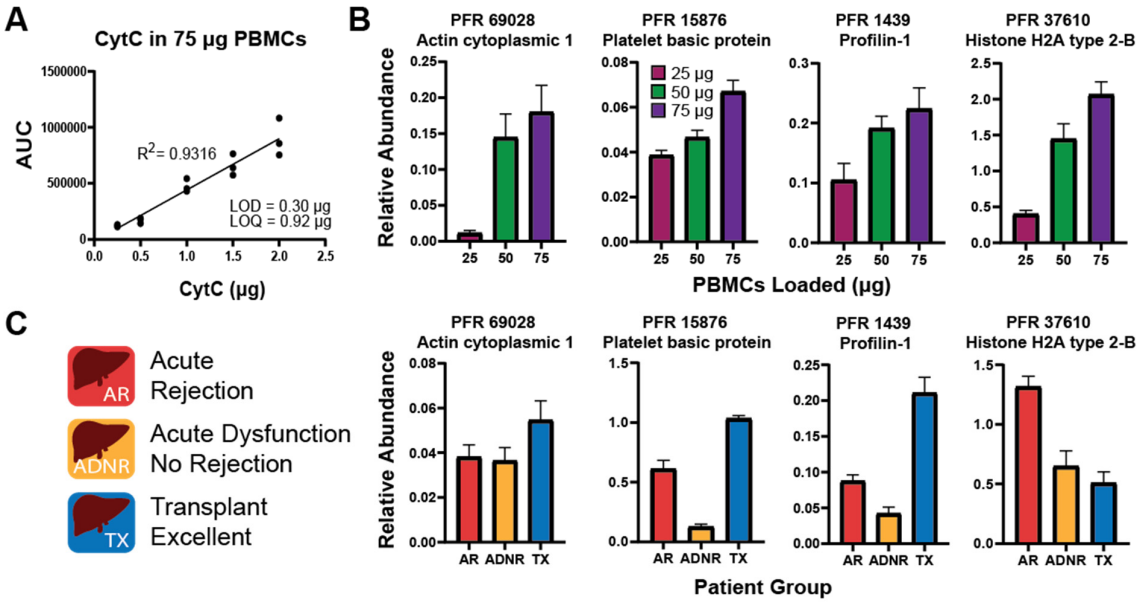


Figure 4.



REFERENCES

1. Schubert, O. T.; Rost, H. L.; Collins, B. C.; Rosenberger, G.; Aebersold, R. Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc* **2017,** *12* (7), 1289-1294.

2. Manes, N. P.; Nita-Lazar, A. Application of targeted mass spectrometry in bottom-up proteomics for systems biology research. *J Proteomics* **2018,** *189*, 75-90.

3. Patel, V. J.; Thalassinos, K.; Slade, S. E.; Connolly, J. B.; Crombie, A.; Murrell, J. C.; Scrivens, J. H. A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J Proteome Res* **2009,** *8* (7), 3752-9.

4. Picotti, P.; Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **2012,** *9* (6), 555-66.

5. Anderson, L.; Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* **2006,** *5* (4), 573-88.

6. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* **2012,** *11* (11), 1475-88.

7. Scigelova, M.; Hornshaw, M.; Giannakopulos, A.; Makarov, A. Fourier transform mass spectrometry. *Mol Cell Proteomics* **2011,** *10* (7), M111 009431.

8. Schiess, R.; Wollscheid, B.; Aebersold, R. Targeted proteomic strategy for clinical biomarker discovery. *Mol Oncol* **2009,** *3* (1), 33-44.

9. Whiteaker, J. R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R. M.; Zhao, L.; Voytovich, U. J.; Kelly-Spratt, K. S.; Krasnoselsky, A.; Gafken, P. R.; Hogan, J. M.; Jones, L. A.; Wang, P.; Amon, L.; Chodosh, L. A.; Nelson, P. S.; McIntosh, M. W.; Kemp, C. J.; Paulovich, A. G. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotechnol* **2011,** *29* (7), 625-34.

10. Gaither, C.; Popp, R.; Mohammed, Y.; Borchers, C. H. Determination of the concentration range for 267 proteins from 21 lots of commercial human plasma using highly multiplexed multiple reaction monitoring mass spectrometry. *Analyst* **2020,** *145* (10), 3634-3644.

11. Heil, L. R.; Remes, P. M.; MacCoss, M. J. Comparison of Unit Resolution Versus High-Resolution Accurate Mass for Parallel Reaction Monitoring. *J Proteome Res* **2021,** *20* (9), 4435-4442.

12. Nesvizhskii, A. I.; Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **2005,** *4* (10), 1419-

40.

13. Loven, J.; Orlando, D. A.; Sigova, A. A.; Lin, C. Y.; Rahl, P. B.; Burge, C. B.; Levens, D. L.; Lee, T. I.; Young, R. A. Revisiting global gene expression analysis. *Cell* **2012,** *151* (3), 476-82.

14. Ntai, I.; Fornelli, L.; DeHart, C. J.; Hutton, J. E.; Doubleday, P. F.; LeDuc, R. D.; van Nispen, A. J.; Fellers, R. T.; Whiteley, G.; Boja, E. S.; Rodriguez, H.; Kelleher, N. L. Precise characterization of KRAS4b proteoforms in human colorectal cells and tumors reveals mutation/modification cross-talk. *Proc Natl Acad Sci U S A* **2018,** *115* (16), 4140-4145.

15. Duan, G.; Walther, D. The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol* **2015,** *11* (2), e1004049.

16. Bah, A.; Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J Biol Chem* **2016,** *291* (13), 6696-705.

17. Wei, L.; Gregorich, Z. R.; Lin, Z.; Cai, W.; Jin, Y.; McKiernan, S. H.; McIlwain, S.; Aiken, J. M.; Moss, R. L.; Diffee, G. M.; Ge, Y. Novel Sarcopenia-related Alterations in Sarcomeric Protein Post-translational Modifications (PTMs) in Skeletal Muscles Identified by Top-down Proteomics.

Mol Cell Proteomics **2018,** *17* (1), 134-145.

18. Cai, W.; Hite, Z. L.; Lyu, B.; Wu, Z.; Lin, Z.; Gregorich, Z. R.; Messer, A. E.; McIlwain, S. J.; Marston, S. B.; Kohmoto, T.; Ge, Y. Temperature-sensitive sarcomeric protein post-translational modifications revealed by top-down proteomics. *J Mol Cell Cardiol* **2018,** *122*, 11-22.

19. Wilkins, J. T.; Seckler, H. S.; Rink, J.; Compton, P. D.; Fornelli, L.; Thaxton, C. S.; LeDuc, R.; Jacobs, D.; Doubleday, P. F.; Sniderman, A.; Lloyd-Jones, D. M.; Kelleher, N. L. Spectrum of Apolipoprotein AI and Apolipoprotein AII Proteoforms and Their Associations With Indices of Cardiometabolic Health: The CARDIA Study. *J Am Heart Assoc* **2021,** *10* (17), e019890.

20. Moors, T. E.; Maat, C. A.; Niedieker, D.; Mona, D.; Petersen, D.; Timmermans-Huisman, E.; Kole, J.; El-Mashtoly, S. F.; Spycher, L.; Zago, W.; Barbour, R.; Mundigl, O.; Kaluza, K.; Huber, S.; Hug, M. N.; Kremer, T.; Ritter, M.; Dziadek, S.; Geurts, J. J. G.; Gerwert, K.; Britschgi, M.; van de Berg, W. D. J. The subcellular arrangement of alpha-synuclein proteoforms in the Parkinson's disease brain as revealed by multicolor STED microscopy. *Acta Neuropathol* **2021,** *142* (3), 423-448.

21. Smith, L. M.; Kelleher, N. L.; Consortium for Top Down, P. Proteoform:

a single term describing protein complexity. *Nat Methods* **2013,** *10* (3), 186-7.

22. Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev Anal Chem (Palo Alto Calif)* **2016,** *9* (1), 499-519.

23. Ntai, I.; Kim, K.; Fellers, R. T.; Skinner, O. S.; Smith, A. D. t.; Early, B. P.; Savaryn, J. P.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L. Applying label-free quantitation to top down proteomics. *Anal Chem* **2014,** *86* (10), 4961-8.

24. Kline, J. T.; Belford, M. W.; Huang, J.; Greer, J. B.; Bergen, D.; Fellers, R. T.; Greer, S. M.; Horn, D. M.; Zabrouskov, V.; Huguet, R.; Boeser, C. L.; Durbin, K. R.; Fornelli, L. Improved Label-Free Quantification of Intact Proteoforms Using Field Asymmetric Ion Mobility Spectrometry. *Anal Chem* **2023,** *95* (23), 9090-9096.

25. Hung, C. W.; Tholey, A. Tandem mass tag protein labeling for top-down identification and quantification. *Anal Chem* **2012,** *84* (1), 161-70.

26. Yu, D.; Wang, Z.; Cupp-Sutton, K. A.; Guo, Y.; Kou, Q.; Smith, K.; Liu, X.; Wu, S. Quantitative Top-Down Proteomics in Complex Samples Using Protein-Level Tandem Mass Tag Labeling. *J Am Soc Mass Spectrom* **2021,** *32* (6), 1336-1344.

27. Collier, T. S.; Sarkar, P.; Rao, B.; Muddiman, D. C. Quantitative top-down proteomics of SILAC labeled human embryonic stem cells. *J Am Soc Mass Spectrom* **2010,** *21* (6), 879-89.

28. Rhoads, T. W.; Rose, C. M.; Bailey, D. J.; Riley, N. M.; Molden, R. C.; Nestler, A. J.; Merrill, A. E.; Smith, L. M.; Hebert, A. S.; Westphall, M. S.; Pagliarini, D. J.; Garcia, B. A.; Coon, J. J. Neutron-encoded mass signatures for quantitative top-down proteomics. *Anal Chem* **2014,** *86* (5), 2314-9.

29. Holt, M. V.; Wang, T.; Young, N. L. High-Throughput Quantitative Top-Down Proteomics: Histone H4. *J Am Soc Mass Spectrom* **2019,** *30* (12), 2548-2560.

30. Seckler, H. D. S.; Fornelli, L.; Mutharasan, R. K.; Thaxton, C. S.; Fellers, R.; Daviglus, M.; Sniderman, A.; Rader, D.; Kelleher, N. L.; Lloyd-Jones, D. M.; Compton, P. D.; Wilkins, J. T. A Targeted, Differential Top-Down Proteomic Methodology for Comparison of ApoA-I Proteoforms in Individuals with High and Low HDL Efflux Capacity. *J Proteome Res* **2018,** *17* (6), 2156-2164.

31. Adams, L. M.; DeHart, C. J.; Drown, B. S.; Anderson, L. C.; Bocik, W.; Boja, E. S.; Hiltke, T. M.; Hendrickson, C. L.; Rodriguez, H.; Caldwell, M.; Vafabakhsh, R.; Kelleher, N. L. Mapping the KRAS proteoform landscape

in colorectal cancer identifies truncated KRAS4B that decreases MAPK signaling. *J Biol Chem* **2023,** *299* (1), 102768.

32. Wang, E. H.; Combe, P. C.; Schug, K. A. Multiple Reaction Monitoring for Direct Quantitation of Intact Proteins Using a Triple Quadrupole Mass Spectrometer. *J Am Soc Mass Spectrom* **2016,** *27* (5), 886-96.

33. Wang, E. H.; Appulage, D. K.; McAllister, E. A.; Schug, K. A. Investigation of Ion Transmission Effects on Intact Protein Quantification in a Triple Quadrupole Mass Spectrometer. *J Am Soc Mass Spectrom* **2017,** *28* (9), 1977-1986.

34. Marakova, K.; Rai, A. J.; Schug, K. A. Effect of difluoroacetic acid and biological matrices on the development of a liquid chromatography-triple quadrupole mass spectrometry method for determination of intact growth factor proteins. *J Sep Sci* **2020,** *43* (9-10), 1663-1677.

35. Wang, E. H.; Nagarajan, Y.; Carroll, F.; Schug, K. A. Reversed-phase separation parameters for intact proteins using liquid chromatography with triple quadrupole mass spectrometry. *J Sep Sci* **2016,** *39* (19), 3716-3727.

36. Wang, W.; Choi, B. K.; Li, W.; Lao, Z.; Lee, A. Y.; Souza, S. C.; Yates, N. A.; Kowalski, T.; Pocai, A.; Cohen, L. H. Quantification of intact and truncated stromal cell-derived factor-1alpha in circulation by immunoaffinity

enrichment and tandem mass spectrometry. *J Am Soc Mass Spectrom* **2014,** *25* (4), 614-25.

37. Chen, Y.; Mao, P.; Wang, D. Quantitation of Intact Proteins in Human Plasma Using Top-Down Parallel Reaction Monitoring-MS. *Anal Chem* **2018,** *90* (18), 10650-10653.

38. Lefebvre, D.; Blanco-Valle, K.; Feraudet-Tarisse, C.; Merda, D.; Simon, S.; Fenaille, F.; Hennekinne, J. A.; Nia, Y.; Becher, F. Quantitative Determination of Staphylococcus aureus Enterotoxins Types A to I and Variants in Dairy Food Products by Multiplex Immuno-LC-MS/MS. *J Agric Food Chem* **2021,** *69* (8), 2603-2610.

39. Takemori, A.; Butcher, D. S.; Harman, V. M.; Brownridge, P.; Shima, K.; Higo, D.; Ishizaki, J.; Hasegawa, H.; Suzuki, J.; Yamashita, M.; Loo, J. A.; Loo, R. R. O.; Beynon, R. J.; Anderson, L. C.; Takemori, N. PEPPI-MS: Polyacrylamide-Gel-Based Prefractionation for Analysis of Intact Proteoforms and Protein Complexes by Mass Spectrometry. *J Proteome Res* **2020,** *19* (9), 3779-3791.

40. Wessel, D.; Flugge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **1984,** *138* (1), 141-3.

41. Toby, T. K.; Fornelli, L.; Srzentic, K.; DeHart, C. J.; Levitsky, J.; Friedewald, J.; Kelleher, N. L. A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nat Protoc* **2019,** *14* (1), 119-152.

42. Toby, T. K.; Abecassis, M.; Kim, K.; Thomas, P. M.; Fellers, R. T.; LeDuc, R. D.; Kelleher, N. L.; Demetris, J.; Levitsky, J. Proteoforms in Peripheral Blood Mononuclear Cells as Novel Rejection Biomarkers in Liver Transplant Recipients. *Am J Transplant* **2017,** *17* (9), 2458-2467.

43. Melani, R. D.; Gerbasi, V. R.; Anderson, L. C.; Sikora, J. W.; Toby, T. K.; Hutton, J. E.; Butcher, D. S.; Negrao, F.; Seckler, H. S.; Srzentic, K.; Fornelli, L.; Camarillo, J. M.; LeDuc, R. D.; Cesnik, A. J.; Lundberg, E.; Greer, J. B.; Fellers, R. T.; Robey, M. T.; DeHart, C. J.; Forte, E.; Hendrickson, C. L.; Abbatiello, S. E.; Thomas, P. M.; Kokaji, A. I.; Levitsky, J.; Kelleher, N. L. The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells. *Science* **2022,** *375* (6579), 411-418.