

Mapping the Proteoform Landscape of Five Human Tissues

Bryon S. Drown, Kevin Joos, Rafael D. Melani, Cameron Lloyd-Jones, Jeannie M. Camarillo, and Neil L. Kelleher*

Cite This: *J. Proteome Res.* 2022, 21, 1299–1310

Read Online

ACCESS |



Metrics & More



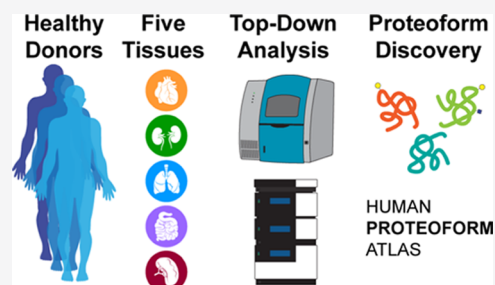
Article Recommendations



Supporting Information

ABSTRACT: A functional understanding of the human body requires structure–function studies of proteins at scale. The chemical structure of proteins is controlled at the transcriptional, translational, and post-translational levels, creating a variety of products with modulated functions within the cell. The term “proteoform” encapsulates this complexity at the level of chemical composition. Comprehensive mapping of the proteoform landscape in human tissues necessitates analytical techniques with increased sensitivity and depth of coverage. Here, we took a top-down proteomics approach, combining data generated using capillary zone electrophoresis (CZE) and nanoflow reversed-phase liquid chromatography (RPLC) hyphenated to mass spectrometry to identify and characterize proteoforms from the human lungs, heart, spleen, small intestine, and kidneys. CZE and RPLC provided complementary post-translational modification and proteoform selectivity, thereby enhancing the overall proteome coverage when used in combination. Of the 11,466 proteoforms identified in this study, 7373 (64%) were not reported previously. Large differences in the protein and proteoform level were readily quantified, with initial inferences about proteoform biology operative in the analyzed organs. Differential proteoform regulation of defensins, glutathione transferases, and sarcomeric proteins across tissues generate hypotheses about how they function and are regulated in human health and disease.

KEYWORDS: top-down proteomics, proteomics, capillary zone electrophoresis, heart, small intestine, kidney, lung, spleen



INTRODUCTION

Mapping the human body is critical to improve our understanding by setting definitive reference points for organs, tissues, and cells of diverse types. In proteomics, a complete understanding of the proteoform¹ diversity requires measurements that systematically capture protein-level complexity. In projects such as the Human Biomolecular Atlas Program (HuBMAP)² and Human Cell Atlas,³ the resolution of mapping can handle single cells in tissues, with several highly multiplexed methods enabled by antibody-based affinity reagents: CODEX,⁴ Immuno-SABER,⁵ CyTOF,⁶ and MIBI,^{7,8} among others. These methods measure the expression of particular epitopes on proteins, although they still fail to capture the full complexity of the proteoforms present. Proteoform-level measurements are more specific for a particular biological state compared to the measurements on the gene or even protein level.^{9,10} While our long-term goal is to develop new technologies that deliver spatial proteoform analysis and build a comprehensive atlas of human proteoforms,¹¹ our goal here is to identify proteoforms present in primary human tissues and provide an initial assessment of their post-translational modifications (PTMs) across tissue types.

Top-down proteomics (TDP), where intact proteins are isolated and fragmented by mass spectrometry (MS), is well suited for the identification and characterization of tissue-specific proteoforms. For the analysis of complex proteome samples, upfront separation and/or fractionation represents a

crucial part in TDP workflows to reduce complexity prior to MS. Reversed-phase liquid chromatography (RPLC) is traditionally employed as the method of choice in TDP, which is due to its reproducibility, separation capacity, and MS compatibility, although capillary zone electrophoresis (CZE) represents an alternative for online MS. In particular, the separation principle of CZE is based on differences in electrophoretic mobilities (*charge-to-size* ratio) and is considered largely “orthogonal” to RPLC, where separation is driven by the hydrophobicity of analyte molecules. For this reason, the combination of information generated by both techniques is anticipated to increase the number of identified proteins and proteoforms.

Here, we report results from two workflows for mapping the proteoform landscape of solid tissues and present the first iteration with five commonly studied human tissues (heart, lungs, kidneys, small intestines, and spleen). Initially, the extracted proteoforms were prefractionated using gel-eluted liquid fraction entrapment electrophoresis (GELFrEE),¹² followed by subsequent CZE-MS and nano-RPLC-MS analysis.

Received: January 18, 2022

Published: April 12, 2022



This study contributes 7373 proteoforms to the Human Proteoform Atlas (HPfA), a FAIR¹³ knowledge base that now contains approximately 60,000 unique proteoforms linked to their biological context.¹⁴

EXPERIMENTAL PROCEDURES

Reagents

All reagents were purchased from Thermo Fisher Scientific at the highest available purity unless otherwise specified.

Tissue Lysate Preparation

Fresh-frozen tissue samples of the human heart, lungs, small intestine, and spleen were obtained from HuBMAP Tissue Mapping Centers (Table S1). The tissue samples were collected under IRB-approved protocols at each institution. Kidney samples were received as 10 μm microtome scrolls embedded in methylcellulose (each ~ 5 mg). All other tissue types were cut into small pieces (~ 5 mm) by the specimen preparer at Mapping Centers. The kidney scrolls were cryopulverized in 2 mL Eppendorf Protein Lo-Bind tubes containing a 5 mm stainless-steel ball (Qiagen, cat. no. 69989) with a CryoMill (Retsch, cat. no. 20.749.001) equipped with a tube adaptor. Nonkidney tissue specimens (50–100 mg) were cryopulverized using the CryoMill equipped with a 25 mL grinding jar containing a 1 inch stainless-steel ball. Three cycles of precooling with liquid nitrogen at 1 Hz for 3 min and grinding at 30 Hz for 1 min were performed. The pulverized tissue was transferred to a 15 mL conical tube and resuspended in 2 mL of cold radio-immunoprecipitation assay lysis buffer [50 mM Tris, 150 mM NaCl, 1% NP-40 (v/v), 0.5% sodium deoxycholate (w/v), 0.1% sodium dodecyl sulfate (w/v), pH 7.4, 1 \times Halt protease and phosphatase inhibitor cocktail (Thermo Scientific)]. The suspension was further disrupted by sonication on ice (40% power, cycle 2 s on, 3 s off, for 30 s total) using a probe sonicator (FisherBrand model 120 with a 1/8 inch probe) and then clarified by centrifugation (3234g, 30 min, 4 $^{\circ}\text{C}$).

Sample Prefractionation and Preparation for MS

The kidney lysates were studied using a $5 \times 4 \times 1 \times 2$ design: five biospecimens from separate donors were GELFrEE-fractionated into four fractions, analyzed by RPLC-MS/MS, and injected in duplicate. The lung lysates were studied in a $3 \times 6 \times 1 \times 3$ design: three samples from a single donor, six fractions, only RPLC, and three injections. The heart lysates were studied in a $2 \times 6 \times 2 \times 3$ design: two donors, six fractions, both CZE and RPLC, and three injections. The small intestine and spleen were studied in a $1 \times 6 \times 2 \times 3$ design: one sample, six fractions, both CZE and RPLC, and three injections. The lysates were fractionated and prepared for MS, as described previously.¹⁵ In brief, the lysates were precipitated by adding four volumes of cold acetone and incubating them at -80 $^{\circ}\text{C}$ for 1 h. The precipitate was collected by centrifugation (20,000g, 30 min, 4 $^{\circ}\text{C}$), and proteins were resolubilized in 1% sodium dodecyl sulfate (w/v). The total protein content was determined by the BCA assay (Thermo Scientific). The samples were fractionated using the GELFrEE 8100 fractionation station (Expedeon). The protein samples (300 μg in 150 μL) were combined with 30 μL of the GELFrEE running buffer and 8 μL of 1 M DTT. The samples were incubated at 95 $^{\circ}\text{C}$ for 5 min, cooled to room temperature, and separated using a 10% GELFrEE cartridge following the manufacturer's protocol. Six (four in the case of kidney samples) 150 μL fractions were collected and stored at -80 $^{\circ}\text{C}$ until immediately prior to analysis. On the day of

analysis, the fractions were thawed on ice and precipitated with methanol–chloroform–water as described.¹⁶ Based on previous experience, each fraction was expected to contain about 5 μg of protein material. The pellets were resuspended in 10 μL of 0.3% acetic acid (HAc) (v/v) and subjected to CZE-MS/MS. When CZE-MS/MS analysis was completed, the samples were diluted with 20 μL of buffer A (5% acetonitrile, 94.8% water, and 0.2% formic acid) and subjected to RPLC-MS/MS analysis. If only RPLC-MS/MS was conducted, the pellets were resuspended directly in 30 μL of buffer A.

Capillary Zone Electrophoresis

CZE was performed using a CESI 8000 Plus (Sciex) equipped with a Neutral OptiMS capillary cartridge (30 μm ID, $L = 90$ cm), neutrally coated. The cartridge was washed and conditioned according to the manufacturer's protocols. Separation conditions: cartridge temperature: 15 $^{\circ}\text{C}$, sample tray temperature: 4 $^{\circ}\text{C}$, background electrolyte: 3% HAc, conductive liquid: 3% HAc, hydrodynamic injection: 2.5 psi for 60 s (corresponds to ~ 20 nL or ~ 10 ng of the protein material). The individual separation method steps are listed in Table S2. Overnight, the capillary was rinsed alternating between high flow (100 psi, 2 min) and low flow (10 psi, 120 min) steps with water. For long-term storage, both separation and conductive lines were rinsed (100 psi) with water for 5 min, respectively, and the cartridge was stored at 4 $^{\circ}\text{C}$.

Reversed-Phase Liquid Chromatography

RPLC was performed using an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) as described previously.¹⁷ In brief, a self-packed trap column (150 $\mu\text{m} \times 2.5$ cm, PLRP-S 5 μm 1000 \AA pore size) and analytical column (75 $\mu\text{m} \times 25$ cm, PLRP-S 5 μm 1000 \AA pore size) were configured in a vented T setup. The trap and column were kept at 55 $^{\circ}\text{C}$. Buffer A: 94.8% water, 5% acetonitrile, 0.2% formic acid; buffer B: 94.8% acetonitrile, 5% water, 0.2% formic acid. The samples were injected (6 μL , ~ 1 μg total protein) onto the trap column and washed with 5% buffer B at 3 $\mu\text{L}/\text{min}$ for 10 min. Following a valve switch, the proteins were separated on the analytical column according to the following gradient: 5% B at 10 min, 15% B at 13 min, 45% B at 70 min, 95% B at 72 min, 95% B at 76 min, 5% B at 80 min, and 5% B from 80 to 90 min. For fractions 5 and 6, the proteins were separated according to the following gradient: 5% B at 10 min, 15% B at 13 min, 50% B at 70 min, 95% B at 72 min, 95% B at 76 min, 5% B at 80 min, and 5% B from 80 to 90 min. The eluted proteins were ionized in positive ion-mode nanoelectrospray ionization using a pulled-tip nanospray emitter (15 μm i.d. \times 125 mm, New Objective) packed with 1 mm of PLRP-S 5 μm 1000 \AA pore size with a custom nanosource.

Top-Down MS

MS was performed either using a Thermo Scientific Orbitrap Eclipse Tribrid mass spectrometer or a Thermo Scientific Fusion Lumos Orbitrap Tribrid mass spectrometer. For analysis on Eclipse MS, data was acquired using the following global parameters spray voltage: 1600 V, sweep gas: 0, ion transfer tube temperature: 320 $^{\circ}\text{C}$, application mode: intact protein, pressure mode: low pressure (2 mTorr), advanced peak determination: true, default charge state: 15, S-lens RF: 30%, source collision-induced dissociation: 15 eV. The precursor spectra were acquired at a 120,000 resolving power, detect type: Orbitrap, scan range: 600–2000 m/z , mass range: normal, AGC target 2E6, normalized AGC target: 500%, max injection time: 50 ms, microscans: 1. The mass spectrometer was operated using a

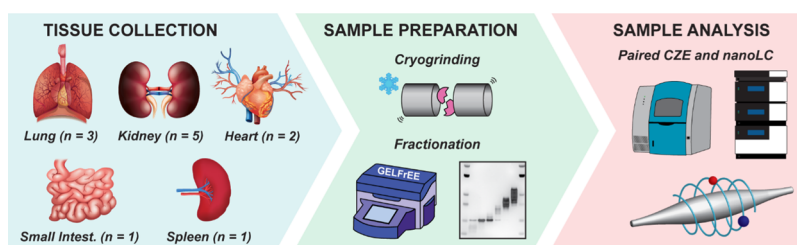


Figure 1. TDP of healthy human tissues. Tissues were obtained from HuBMAP Tissue Mapping Centers. Fresh-frozen tissue was cryogenically pulverized, lysed, and precipitated. Intact proteins were prefractionated using GELFrEE. Each sample was analyzed by CZE-MS/MS and RPLC-MS/MS, respectively.

Table 1. Proteins and Proteoforms Identified from Sampling Five Human Tissue Types

tissue type	biological replicates ^a	separation	MS/MS runs	proteins 1% FDR ^b	unique proteins 1% FDR ^c	proteoforms 1% FDR (C-score >30)	unique proteoforms (C-score >30)
lungs	3	RPLC	49	437	132	5566 (2940)	3601 (1462)
kidneys	5	RPLC	42	307	62	2278 (988)	641 (306)
heart	2	CZE, RPLC	72	305	70	2897 (1346)	1623 (772)
small intestine	1	CZE, RPLC	36	305	43	3101 (1214)	2049 (643)
spleen	1	CZE, RPLC	35	213	36	1869 (972)	870 (589)
total	12		234	1567	343	15,711 (7460)	8784 (3772)
total redundant ^d	12		234	740	343	11,466 (4,906)	8784 (3772)

^aBiological replicate refers to a sample from a single human being. Sample descriptions and metadata are shown in Table S1. ^bThe term “protein” refers to the SwissProt entry mapping to a single human gene. ^cUnique identifications refer to proteins or proteoforms that were only identified in the tissue type indicated. ^dProteins and proteoforms that were observed in more than one human tissue type are counted once in nonredundant totals.

TopN 3 s data-dependent acquisition mode. The precursor ions were filtered by intensity, charge state, and dynamic exclusion. Intensity minimum: 5E3, intensity maximum: 1E20, include charge states: 4–60, include underdetermined charge states: false, dynamic exclusion after n times: 1, dynamic exclusion duration: 60 s, mass tolerance: 0.5 m/z , exclude isotopes: true. The ions for fragmentation were isolated and fragmented via higher energy dissociation (HCD). Detector type: Orbitrap, isolation mode: quadrupole, resolving power: 60,000, scan range: 350–2000 m/z , AGC target: 1E6, normalized AGC target: 2000%, max injection time: 600 ms, microscans: 1, isolation window: 3 m/z , activation type: HCD, collision energy: 32, collision energy mode: fixed.

For analysis on an Orbitrap Fusion Lumos mass spectrometer, data was acquired with the following global parameters: spray voltage: 1600 V, sweep gas: 0, ion transfer tube temperature: 320 °C, application mode: intact protein, pressure mode: low pressure (2 mTorr), advanced peak determination: true, default charge state: 15, S-lens RF: 30%, source collision-induced dissociation: 15 eV. The precursor spectra were acquired at a 120,000 resolving power (at 200 m/z), mass range: normal, detector type: Orbitrap, scan range: 600–2000 m/z , AGC target: 1E6, normalized AGC target: 250%, max injection time: 100 ms, microscans: 4. The mass spectrometer was operated using a Top2 data-dependent acquisition mode. The precursor ions were filtered by intensity, charge state, and dynamic exclusion. Intensity minimum: 2E4, intensity maximum: 1E20, included charge states: 6–60, include undetermined charge states: false, dynamic exclusion after n times: 1, dynamic exclusion duration: 60 s, mass tolerance: 1.5 m/z , exclude isotopes: true. The ions for fragmentation were isolated and fragmented via HCD. Detector type: Orbitrap, isolation mode:

quadrupole, resolving power: 60,000 (at 200 m/z), scan range: 400–2000 m/z , AGC target: 1E6, normalized AGC target: 2000%, max injection time: 400 ms, microscans: 4, isolation window: 3 m/z , activation type: HCD, collision energy: 27, collision energy mode: fixed.

Protein and Proteoform Identification

The raw data files were processed with the publicly available workflow on TDPportal (<https://portal.nrtdp.northwestern.edu>, Code Set 4.0.0) that performed mass inference, searched a database of human proteoforms derived from Swiss-Prot (June 2020) with curated histones, and estimated conservative, context-dependent 1% false discovery rate (FDR) at the protein, isoform, and proteoform levels.¹⁸ Each tissue type was searched separately with its own FDR context. Aggregated search results were used in further data analysis.

Code and Data Availability

Raw files, mzIdentML, and tdReport files were deposited in Massive (Accession MSV000088565). The search results in the tdReport format are viewable using TDViewer—a freeware from Northwestern University (<http://topdownviewer.northwestern.edu>). The search results were further analyzed, and figures were generated with a custom code written for R 4.1.0. The source code for data analysis is available at <https://github.com/bdrown/rplc-cze-tissues>.

RESULTS AND DISCUSSION

The samples were obtained from HuBMAP Tissue Mapping Centers from 10 human donors. The tissue was cryopulverized and lysed, and the proteins were precipitated (Figure 1). To increase the depth of proteome coverage, the proteins were fractionated using GELFrEE prior to MS analysis. Since we

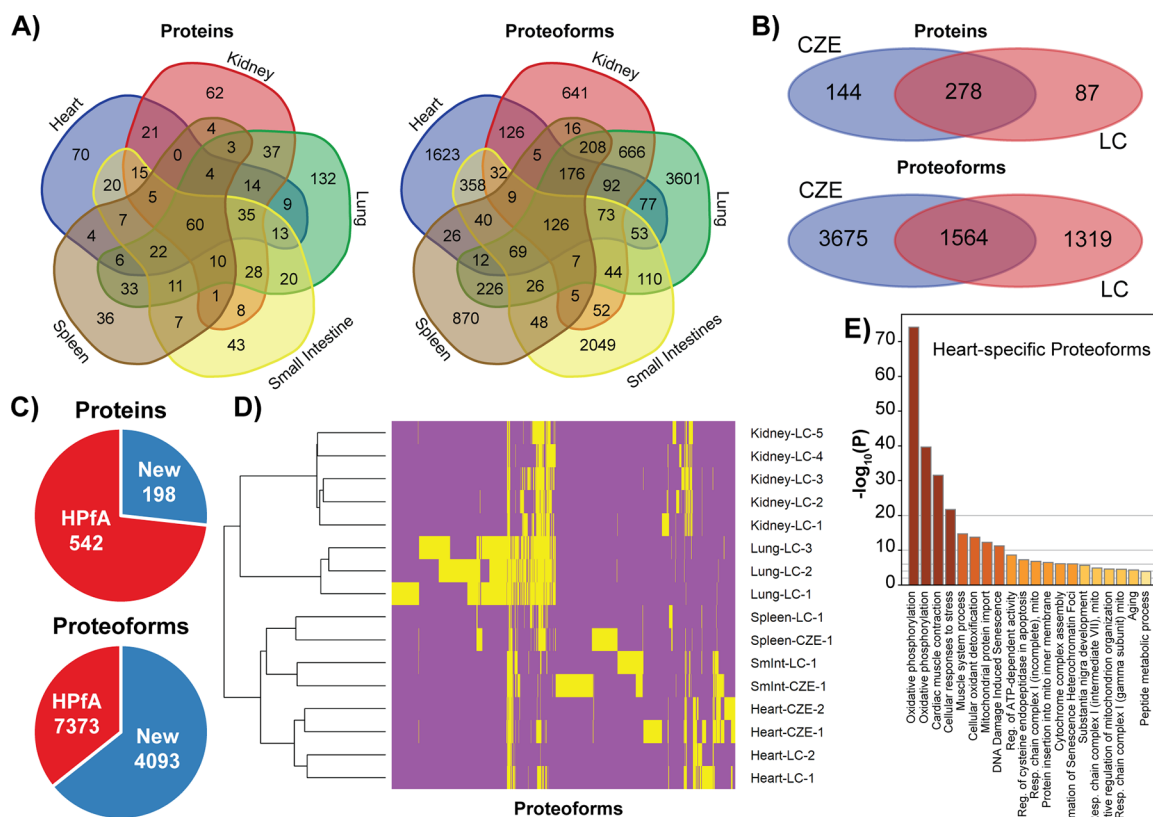


Figure 2. Systematic discovery of unique proteoforms across human tissues. (A) Venn diagrams of shared and unique proteins and proteoforms identified in each tissue. 1% FDR filtering was applied at the PrSM, proteoform, and protein levels for each tissue. (B) Venn diagrams of shared and unique proteins and proteoforms identified in the heart, small intestine, and/or spleen tissues by either CZE or RPLC. (C) Pie charts representing the rediscovery of proteoforms and proteins previously deposited in the HPfA (red) or only this study (New, blue). HPfA was accessed on 8/18/2021. (D) Heat map showing the presence (yellow) and absence (purple) of proteoforms in each tissue sample with hierarchical clustering. (E) Bar graph of top 20 enriched terms from genes associated with proteoforms uniquely identified in the heart tissue using Metascape.

intended to analyze each sample by both CZE and RPLC, we set up two Orbitrap tribrid MS instruments configured with either CZE or RPLC, acquired data for a sample on one system, and immediately acquired data for the same sample on the second one. CZE substantially benefits from a higher scan rate due to generally narrower peak widths. Consequently, the CESI 8000 Plus was hyphenated to the Orbitrap Eclipse, while a Dionex nanoLC was coupled to the Orbitrap Fusion Lumos. Three tissue types (heart, small intestine, and spleen) were analyzed by this paired analysis, while two tissues (lungs and kidneys) were analyzed solely by RPLC-MS on the Orbitrap Eclipse (Table 1).

Identification of Human Proteoforms in Solid Tissues

By searching the TDP data against a database of human proteoforms using TDPportal and 1% conservative FDR, a total of 11,466 proteoforms from 740 proteins were identified (Table 1). Of these annotations, 8784 proteoforms and 343 proteins were unique to a single tissue type (Table 1 and Figure 2A). The lung tissue contained the highest number of proteoforms and proteins (overall and unique), while the kidney tissue contained the fewest unique proteoforms (Figure S1). Despite having the lowest number of proteins identified, the spleen tissue had a high number of proteoforms per protein (Figure S1). While histones and hemoglobin generated the highest number of proteoforms per protein in most tissues, several other proteins populated the top 15 proteins (Figure S2). Among the shared proteins and proteoforms, histones, ribosomal proteins, ATP synthase subunits, and other housekeeping proteins were most frequently

observed (Supporting Information Data 1). Overall, CZE-MS/MS resulted in a higher number of protein and proteoform identifications than RPLC (Figure 2B). However, the difference in MS instrument performance likely contributed to the increased number of identifications characterized by CZE-MS/MS workflow.

We also sought to compare the proteoforms identified in this work to those reported in prior studies. The Human Proteoform Atlas (HPfA, <http://human-proteoform-atlas.org/>) is the most comprehensive collection of characterized proteoforms.¹⁴ The HPfA consists of 49 datasets, which include numerous studies on immortalized cell lines, one study on healthy human solid tissues,¹⁹ two studies on human cancer tissues,^{20,21} and the Blood Proteoform Atlas (<http://blood-proteoform-atlas.org/>).²² Of the 11,466 proteoforms identified in this study, a substantial number of 7373 proteoforms (64.3%) were not previously reported in the HPfA, while 4093 (35.7%) proteoforms were present in this database (Figure 2C). The frequency of rediscovery was higher on the protein level with 198 (26.8%) proteins first reported here and 542 (73.2%) proteins included in the HPfA database (Figure 2C). Thus, while some proteins were identified for the first time in this study, the majority of new proteoforms are differently modified forms of proteins, which were previously detected by TDP. Presence and absence matrices showed clear clustering of tissues at the proteoform (Figure 2D) level, demonstrating that proteoform identifications are more characteristic of the tissues under study.

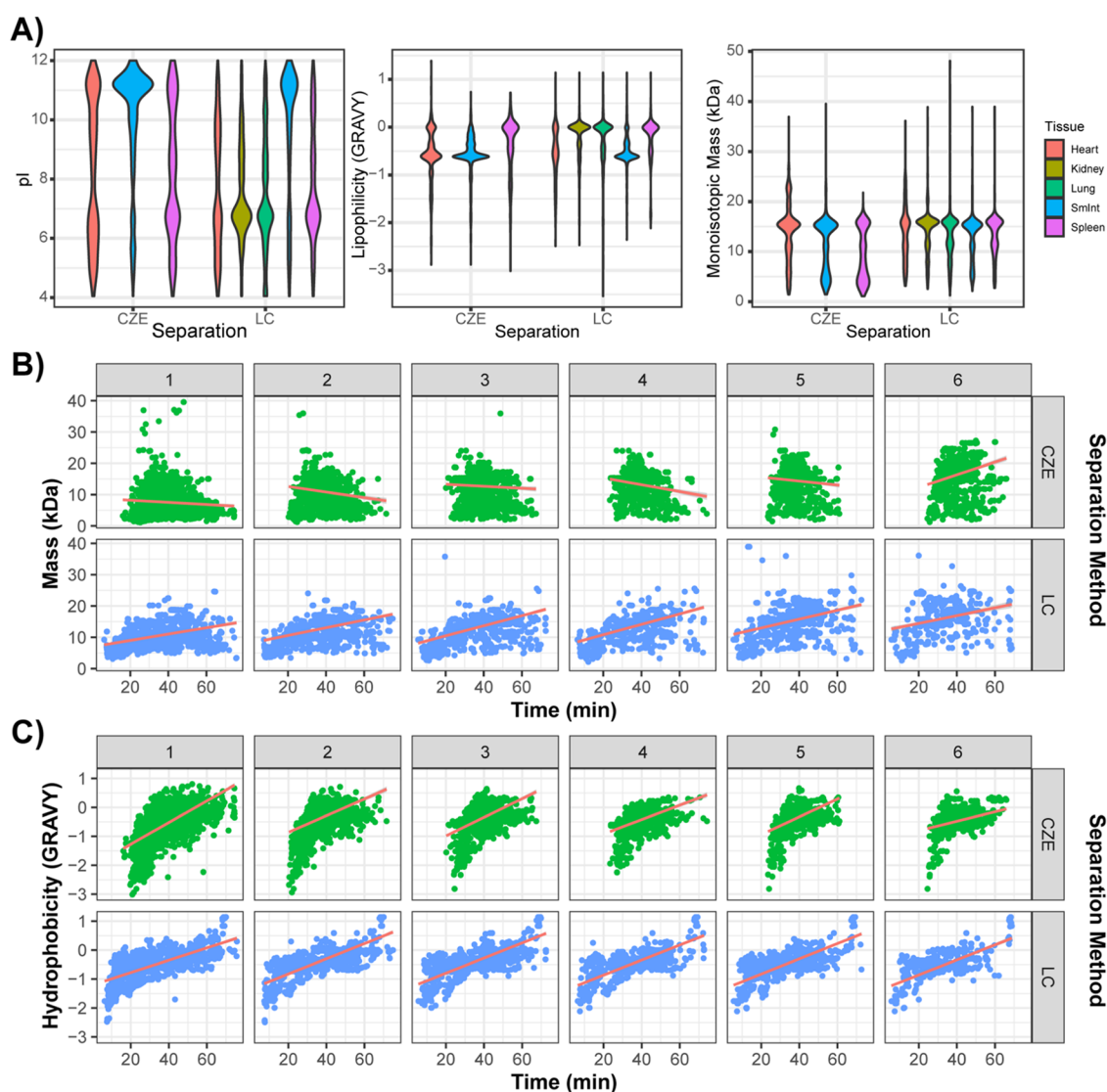


Figure 3. Complementary separation of intact proteins by CZE and RP-nanoLC. (A) Violin plots of proteoform physicochemical properties by the tissue and separation technique. (B) Scatter plots relating the migration/retention time to the monoisotopic mass of proteoforms from the heart and small intestine and the migration/retention time to the monoisotopic mass of proteoforms from the heart, small intestine, and spleen samples subdivided by the separation method and GELFrEE fraction. (C) Scatter plots relating the migration/retention time to the GRAVY score of proteoforms from the heart, small intestine, and spleen samples subdivided by the separation method and GELFrEE fraction. Corresponding correlation coefficients of data presented in panels B and C are listed in Table S3.

A “bird’s-eye” view of the physicochemical properties of proteoforms identified in the five different tissue types, including hydrophobicity, monoisotopic mass, and pI value, can be found in Figures 3A and S3. While the kidney, lung, and spleen tissue proteoforms show similar distributions in their violin plots regarding all three investigated characteristics, distinct differences for the heart and especially small intestine tissue were detected. For example, in the case of the small intestine, a high number of proteoforms in the pI range of 10.5–12.0 was observed, which can be explained by a relative increase in histone proteoforms compared to those in the other analyzed tissue types. This is also supported by the negative GRAVY score, showing a large distribution at around -0.6 . On the other hand, proteoforms observed in the heart tissue exhibit a relatively broad distribution of pI values.

Influence of the Separation Technique

While the performances of CZE and RPLC have been compared in numerous contexts,^{23–27} the paired analysis of the heart, small intestine, and spleen provides an opportunity to explore how proteoforms behave regarding these two separation techniques. Despite requiring similarly long acquisition times, the window of separation for CZE was smaller than that for RPLC. The difference in the separation principle was evident in the relationship between proteoform retention/migration times and mass (Figure 3B), as well as time and hydrophobicity (Figure 3C). While there is a strong correlation between mass and retention time with RPLC, no significant correlation was observed between mass and migration time with CZE (Table S3). Both separation methods demonstrate a correlation between hydrophobicity and time, but RPLC has a stronger correlation. While CZE was performed with an acidic background electrolyte (pH 2.4), we observed a positive correlation between the proteoform hydrophobicity and mass-to-charge

Table 2. Frequency of Observation for Different Types of PTMs on Identified Proteoforms Categorized by the Separation Technique Used in TDP

PTM type	CZE		RPLC		χ^2	<i>p</i> -value ^c
	observed ^a	freq. ^b	observed ^a	freq. ^b		
monoacetylation ^d	2723	0.26	1984	0.31	54	2.6×10^{-12}
unmodified ^d	2298	0.22	1123	0.18	44	4.3×10^{-10}
phosphorylation	1644	0.16	1006	0.16	0.057	>1
monomethylation ^d	1201	0.11	556	0.088	31	3.6×10^{-7}
trimethylation ^d	920	0.088	667	0.11	14	2.8×10^{-3}
dimethylation	919	0.088	642	0.10	8.3	4.9×10^{-2}
half-cystine ^d	360	0.034	118	0.019	35	3.8×10^{-8}
nitrosylation	239	0.023	165	0.026	1.6	>1
oxygenated ^d	72	0.0069	5	7.9×10^{-4}	31	3.4×10^{-7}
pyruvic acid iminylated residue	48	0.0046	41	0.0065	2.3	>1
deamidated L-asparagine	42	0.0040	38	0.0060	2.9	>1
S-palmitoylation	14	0.0013	7	0.0011	0.037	>1
total	10,480		6352			

^aNumber of modifications observed on proteoforms at 1% FDR; count does not include N-terminal and C-terminal modifications; multiple PTMs on the same proteoform are counted multiple times. ^bNumber of observations/sum of PTM observations for each separation technique. ^cBonferroni-corrected *p*-value (*n* = 12). ^dStatistically significant difference ($\alpha < 0.01$) in the frequency of observation.

ratio (Figure S31), which helps to explain the increase in hydrophobicity with migration time (less number of “ionizable” amino acids available per size).

In addition to the physicochemical properties of proteoforms identified using CZE and RPLC difference, the distribution of PTMs was similarly asymmetrical. Twelve PTM categories were identified (Table 2), and their identifications differed significantly (Pearson's χ -squared test, $\chi^2 = 196$, *p*-value $< 2 \times 10^{-16}$) depending on the fractionation method. Two-by-two χ -squared tests were performed to determine which PTMs had significant deviations in their identification rates (observed PTM/the sum of all other PTMs), as described previously.²⁸ Monomethylation, half cystines, and oxygenation were elevated on CZE-MS/MS, while on RPLC-MS/MS, the detection of monoacetylated and trimethylation proteoforms was enhanced. PTM observation frequencies at the proteoform spectral match (PrSM) level followed the same trends in observation biases (Table S4). The elevation of half-cystines and oxygenated residues in the CZE-MS/MS data suggests that the electrophoretic process can oxidize some sensitive residues. While the rate of observing oxidized proteoforms is still low overall, this trend should be considered when performing CZE-MS/MS acquisition. The differential rates of methylation and acetylation led us to see if histones were more highly characterized by one separation method. Indeed, the number of histone proteoforms identified and the number of histone PrSMs were elevated in the CZE-MS/MS data compared to those in paired LC-MS/MS data (Figure S4A,C). This trend was maintained even when normalizing for total proteoforms and spectral matches (Figure S4B,D). Summarized, these observations substantiate the benefit of the combination of CZE- and RPLC-derived data by increasing the coverage of the proteoform discovery workflow.

Tissue-Specific Proteoforms and Handling of PTM Ambiguity

Uncertainty in the exact position of a PTM on a proteoform can arise in cases where SwissProt entries have many recorded modifications and amino acid variants and fragmentation data are incomplete to assert an unambiguous level 1 proteoform.²⁹ This phenomenon is exemplified by cardiac troponin C (cTnC),

which was identified in its canonical form (full length, N-terminal acetylated, PFR55232) as a level 1 proteoform (Figure 4A). Nine additional proteoforms had sufficiently high proteoform-level Q-scores to pass FDR cutoffs due to excellent sequence coverage in regions without modifications, and they were classified as level 3 proteoforms with some PTM site ambiguity (Figure 4A). The example of cTnC is not alone; the majority of proteoforms identified in this study are either chemically modified or bear a sequence variant as only 33% are unmodified (Figure 4B). While filtering by C-score can help triage level 3 proteoforms for which PTM localization is ambiguous, the C-score does not help in cases where there is only one possible site of modification.³⁰

To curate a core set of proteoforms uniquely expressed in the five individual tissue types, we implemented a conservative process to select those proteoforms with PTMs with direct fragment ion support (level 1 proteoforms²⁹). To this end, the number of matching fragment ions that bear a PTM (or amino acid variant) were counted for each PrSM. While many mutated and modified proteoforms have supporting fragment ions (level 1), a disproportionate number of modified proteoforms were level 3 with two or fewer ions (Figure 4C,D). Consequently, the requirement of having ≥ 3 supporting fragment ions for modified proteoforms was added in addition to a C-score > 30 . This process culled the set of 8784 unique proteoforms in Table 1 down to 2843 level 1 tissue-specific proteoforms (Figure 4E and Supporting Information Data 1).

More level 1 tissue-specific proteoforms were identified in a subsequence search (previously called BioMarker search that identifies portions of full-length proteoforms^{31,32}) than in absolute mass searches. Specifically, 2548 proteoforms were identified in subsequence searching compared to 295 proteoforms identified in absolute mass searches. Subsequence searches identify proteolytic fragments that often arise from endogenous proteolytic events and can serve as significant biomarkers.²¹ While a portion of these proteoforms may be the product of nonspecific proteolysis, the consensus sequence of cleavage sites varied across tissues (Figure S5). Truncated proteoforms from the heart, kidneys, and small intestine showed enrichment of F, Y, W, and L at P1, which suggests chymotrypsin activity. The spleen proteoforms demonstrated enrichment of

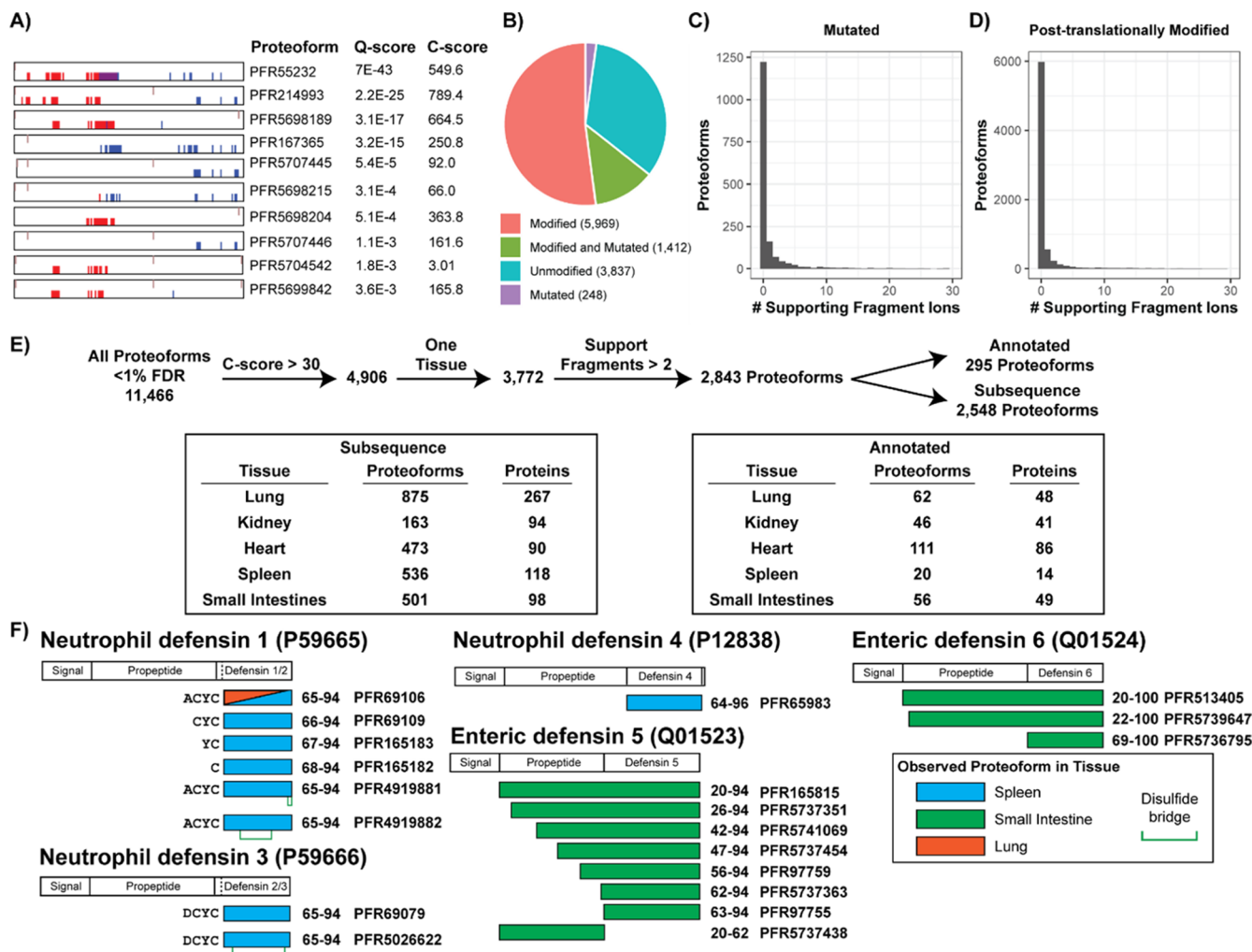


Figure 4. Selection of tissue-specific proteoforms. (A) Cigar depiction of cTnC proteoforms identified in the human heart tissue. Red, blue, and purple marks on the bottom of cigars indicate b, y, and both b and y fragment ions. Tan marks on top of cigars indicate the presence of a PTM or sequence variant. (B) Distribution of proteoforms identified with PTMs or sequence variance. Proteolytic cleavage and N-terminal acetylation are excluded from consideration as PTMs in this panel. (C) Histogram of proteoforms and the number of matching fragment ions that support the presence of a sequence variant (e.g., a polymorphism). (D) Histogram of proteoforms and the number of matching fragment ions that support the presence of a PTM. (E) Sequential filtering of proteoforms to identify high-confidence tissue-specific proteoforms. (F) Identification of tissue-specific defensin proteoforms.

hydrophobic residues but no apparent sequence specificity. This lack of specificity combined with a high proteoform-to-protein ratio agrees well with the role of the spleen for scavenging senescent blood cells.³³ Lung proteoforms had a higher propensity of cysteine at P1, which is not commonly observed for specific proteases. This enrichment was driven by 24 of the 715 lung-specific proteoforms with N-terminal cleavage. 9 of these 24 proteoforms originate from collapsing response mediator protein 2 (CRMP-2 and Q16555), with cleavage occurring at C439 (Figure S6). CRMP-2 has largely been studied in the context of neurological diseases due to its role in microtubule assembly and axon growth.³⁴ Indeed, C-terminal truncation of CRMP-2 has been linked to neurodegeneration,³⁵ and the cleavage site was later localized to S517.³⁶ As the function of CRMP-2 in the lung tissue has only recently begun to be characterized,³⁷ this novel truncation at C439 may assist in elucidating its role.

Subsequence searching also identified a proteolytic cleavage site in CDGSH iron-sulfur domain-containing protein 1 (mitoNEET and Q9NZ45) at L47 (Figure S7). MitoNEET is a mitochondrial outer membrane protein that was initially

discovered as an off-target interactor of the PPAR- γ agonist pioglitazone.³⁸ With its iron-sulfur cluster oriented toward the cytosol, mitoNEET acts as a redox sensor and regulator of mitochondrial iron.³⁹⁻⁴¹ Downregulation of mitoNEET has been associated with aging and increased risk of heart failure.⁴² The canonical proteoform of mitoNEET was observed in both the small intestine and heart tissue, while both proteolytic products were observed solely in the heart tissue (Figure S7). Cleavage at L47 does not disrupt the iron-sulfur cluster binding site but does separate this reactive center from the protein's transmembrane domain. Thus, proteolytic cleavage may act as a means for regulating mitoNEET or a mechanism by which full-length mitoNEET abundance declines in aging cardiomyocytes.

Unique Proteoforms Are Reflective of Tissue Central Functions

Many of the tissue-specific proteoforms originate from genes involved in the core function of these tissues, as indicated by gene ontology enrichment (Figures 2E and S8). The subsequence proteoform search identified a series of proteoforms associated with defensins with distinct expression patterns (Figures 4F and S9). Defensins are a family of small cationic host

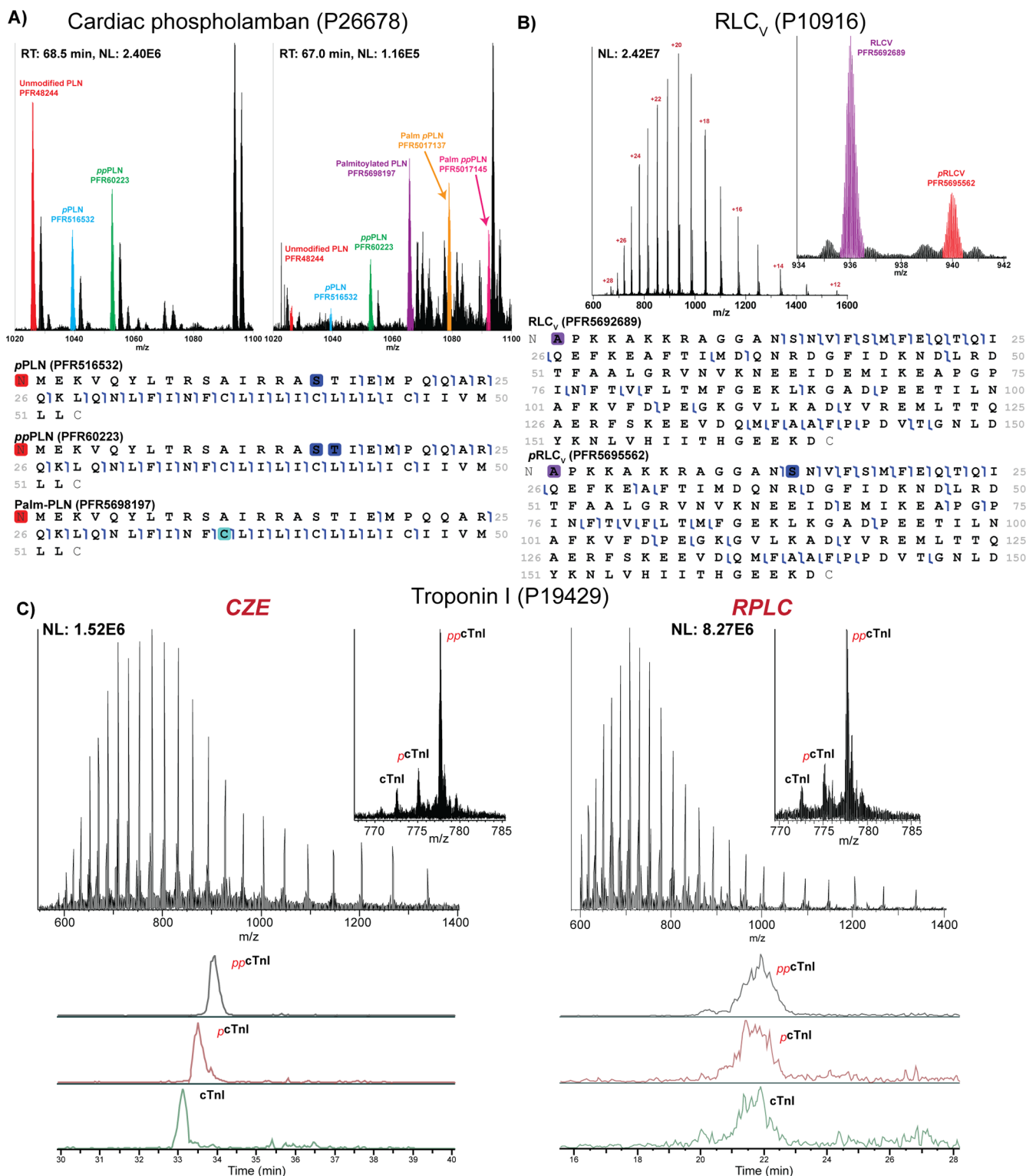


Figure 5. Unique cardio-proteforms identified in paired RPLC- and CZE-MS/MS analysis. (A) Phosphorylated and palmitoylated proteoforms of PLN (P26678) were observed by RPLC-MS/MS late in the chromatogram. (B) Phosphorylation of the ventricular myosin regulatory light chain (RLCV and P10916). HCD fragmentation precisely localized the phosphorylation to S15. (C) cTnI (P19429) was observed by CZE- and RPLC-MS/MS as three phosphoproteoforms, which correlate with enlargement of the heart in a model of hypertrophic cardiomyopathy (ref 60). Both CZE- and RPLC-TDPs successfully resolved and quantified all three proteoforms.

defense proteins characterized by three conserved intramolecular disulfide bonds.⁴³ Six human α -defensins have been identified to date and are subdivided into human neutrophil peptides 1–4 (HNP1–4) and human (enteric) defensins

(HDS–6). HNPs are stored as mature peptides in granules of neutrophils and released upon activation by exocytosis.⁴⁴ HNP1 (PFR69106) was identified in both lung and spleen tissues, as expected for tissues with high neutrophil content. HNP2

(PFR69109), HNP3 (PFR69079), HNP4 (PFR65983), and truncation products of HNP2 (PFR165182 and PFR165183) were observed exclusively in the spleen tissue. No β -defensin proteoforms were identified. HD5 and HD6 are produced in Paneth cells at the base of small-intestinal crypts.⁴⁵ Accordingly, HD5 and HD6 were detected exclusively in the small-intestinal tissue. Unlike other defensins, HD5 is stored as a propeptide, and the fully mature peptides are thought to be produced by intracellular trypsin.⁴⁶ Consequently, the HD5 propeptide (PFR165815) and several truncated products were observed. Several of these truncated proteoforms (PFR5737351, PFR97759, and PFR97755) correspond to trypsin cleavage sites (R25, R55, and R62), while others (PFR5741069, PFR5737454, and PFR5737363) seem to correspond to other mechanisms of cleavage considering the residues at the P1 positions (D41, F46, and A61). Defensins are important components of the host innate immunity, so observing new proteoforms on mucosal surfaces is important in understanding their regulation and design of therapeutic mimetics.^{47,48} Furthermore, these findings are a good showcase for the capabilities of the presented setup to evaluate tissue-specific proteoform-related questions.

Glutathione *S*-transferases are a family of proteins involved in inflammation and the cellular defense against toxic and carcinogenic compounds.^{49,50} Proteoforms from this protein family were broadly observed but with distinct tissue distributions (Figure S10). Glutathione *S*-transferase A1 (P08263) and A2 (P09210) were observed primarily in the small intestine and kidneys, respectively. The polymorphism E210A (rs6577) was observed in a single kidney sample (Biorep 3), which was derived from a 53-year-old African American male (Table S1). This coding SNP occurs with much higher frequency in African Americans (56.5%) compared to the global population (9.9%).⁵¹ Microsomal glutathione *S*-transferases (MGSTs) 1, 2, and 3 were observed in the small intestine and lungs (1), small intestine and kidneys (2), and heart tissue (3), respectively (Figure S10C,D). These glutathione transferases are polytopic membrane proteins located in the endoplasmic reticulum membrane with both glutathione conjugation and peroxidase activity.^{52,53} A novel MGST3 proteoform (PFR5719232) that lacks the C-terminal cysteine necessary for *S*-palmitoylation was the predominant form observed in the heart tissue.⁵⁴

Enrichment of functionally relevant genes from the identified proteoforms was particularly notable for the heart tissue, with terms associated with ATP synthesis and muscle contraction leading the list (Figure 2E). Six proteoforms of cardiac phospholamban (PLN), a key regulator of cardiac contraction via inhibition of the sarcoplasmic reticulum calcium pump (SERCA), were identified by RPLC-MS/MS (Figure 5A).⁵⁵ While unmodified PLN and palmitoylated PLN have both been reported previously,⁵⁶ this study is the first report of phosphorylated PLN and combined phosphorylation and palmitoylation. Phosphorylation and palmitoylation of PLN have both been shown to control the impact localization, complexation, and inhibition of SERCA, so accurate measurement of their combination will help clarify PLN's role in health and disease.⁵⁷

We also present evidence for phosphorylation of the ventricle myosin regulatory light chain (RLC_V). Prior reports by the Ge group have established N-terminal trimethylation of RLC_V and phosphorylation of swine RLC_V, but phosphorylation of human RLC_V was unlocalized.^{58,59} By calculating the area-under-curve

from extracted ion chromatograms of each proteoform, phosphorylated RLC_V is estimated to be at 9% relative abundance. The removal of N-terminal methionine and trimethylation was confirmed by tandem HCD fragmentation, and the site of phosphorylation was localized to S15, which is analogous to the site identified on swine RLC_V (Figure 5B). On a last analytical note, phosphoproteoforms of cardiac troponin I (cTnI)⁶⁰ were not resolved by RPLC but were baseline-separated by CZE (Figure 5C); proteoform quantitation by both techniques showed <10% coefficient of variation between them. Better separation of charge variants such as phospho-troponin by CZE should translate into better on-the-fly sequence coverage and proteoform characterization with tandem MS scan speeds.

CONCLUSIONS

We have described the combination of TDP data collected with online separation by RPLC and CZE to expand the depth of human proteome coverage. All proteomics methods face the challenge of measuring low-abundance analytes, so identifying robust approaches that introduce new proteoform selectivity is highly sought. RPLC and CZE were shown to possess differential proteoform selectivity that manifests as different physiochemical properties and PTM profiles. In a TDP study of five human tissues, we dramatically expanded the number of proteoforms associated with these tissues by combining the two methods.

Confident assignment of proteoforms bearing PTMs or sequence variations becomes more challenging as query proteoforms get larger and the search databases contain more candidate PTM sites. Unambiguous level 1 proteoform assignments are particularly troublesome when seeking proteoforms specific to a particular biological context (e.g., tissue types), but this can be significantly mitigated with the inclusion of fragment-ion data quality standards. Even at the current levels of proteoform characterization quality, organ-specific proteoforms achieve robust tissue type identification.

The genes from the tissue-specific proteoforms identified in this study were tied to the core function of the tissues, as broadly indicated by GEO analysis. This is further supported by specific examples such as proteins that regulate muscle contractility (PLN, RLCV, and cardiac troponins), host–pathogen interaction (defensins), cytoskeletal reorganization (CRMP-2), and metabolic detoxification (family of glutathione transferases). In many cases, these unique proteoforms were detected with only one of the upfront separation methods. Thus, proper exploration of our hypothesis that proteoform-level measurements more fully capture biological context than protein-level measurement requires an increased depth of proteome coverage.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00034>.

Performance metrics of proteoform search across tissues, proteoform annotation frequency for top 15 proteins in each tissue, distribution of physiochemical properties of proteoforms, differential histone characterization by CZE-MS and LC-MS, analysis of cleavage sites on proteoforms discovered by subsequence search, proteoforms of CRMP2 identified in the human lung tissue with subsequence search, identification of mitoNEET proteo-

forms following proteolytic cleavage at L47, gene ontology enrichment of genes associated with unique proteoforms, characterization of neutrophil defensin proteoforms with fragment maps, identification and tissue distribution of glutathione transferase proteoforms, tissue samples analyzed by top-down proteomics, program for the CZE separation method, correlation coefficients of retention/migration time and proteoform mass by the separation method and GELFrEE fraction, and post-translationally modified proteoform identification rates at the PrSM level (PDF)

List of proteoforms identified in this study (XLSX)

AUTHOR INFORMATION

Corresponding Author

Neil L. Kelleher – Departments of Molecular Biosciences, Chemistry, and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, United States; orcid.org/0000-0002-8815-3372; Email: n-kelleher@northwestern.edu

Authors

Bryon S. Drown – Departments of Molecular Biosciences, Chemistry, and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, United States; orcid.org/0000-0002-5631-512X

Kevin Jooß – Departments of Molecular Biosciences, Chemistry, and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, United States; orcid.org/0000-0003-1669-7837

Rafael D. Melani – Departments of Molecular Biosciences, Chemistry, and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, United States

Cameron Lloyd-Jones – Departments of Molecular Biosciences, Chemistry, and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, United States

Jeannie M. Camarillo – Departments of Molecular Biosciences, Chemistry, and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, United States

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jproteome.2c00034>

Author Contributions

Data acquisition was performed by B.S.D., K.J., and R.D.M with support from C.L.J. Data analysis and visualization was performed by B.S.D. with additional input from K.J. and N.L.K. J.M.C and N.L.K. collected funding support. B.S.D., K.J., R.D.M., and N.L.K. wrote and edited the manuscript. All authors critically reviewed and given approval to the final version of the manuscript.

Funding

This material is based upon the work supported by the NIH Common Fund, through the Office of Strategic Coordination/Office of the Director under award UH3 CA246635 (N.L.K.), National Institute of General Medical Sciences of the National Institutes of Health grant P41 GM108569 (N.L.K.), and National Institute of Cancer of the National Institutes of Health grant F32 CA246894 (B.S.D.).

Notes

The authors declare the following competing financial interest(s): NLK is involved in entrepreneurial activities in top-down proteomics and consults for Thermo Fisher Scientific.

ACKNOWLEDGMENTS

The authors thank the principal investigators at HuBMAP Tissue Mapping Centers, Mark Atkinson (University of Florida), Shin Lin (University of Washington), Mike Snyder (Stanford University), Jeff Spraggins (Vanderbilt University), and Gloria Pryhuber (University of Rochester), for contributing tissue samples. We thank SCIEX for their support throughout this research project.

ABBREVIATIONS

BCA, bicinchoninic acid; CZE, capillary zone electrophoresis; CRMP-2, collapsing response mediator protein 2; cTnC, cardiac troponin C; cTnI, cardiac troponin I; cTnT, cardiac troponin T; DTT, dithiothreitol; FDR, false-discovery rate; GELFrEE, gel-eluted liquid fraction entrapment electrophoresis; HAC, acetic acid; HCD, higher-energy collisional dissociation; HD, human enteric defensin; HNP, human neutrophil defensin peptide; HPfA, Human Proteoform Atlas; HuBMAP, Human BioMolecular Atlas Program; MGST, microsomal glutathione S-transferase; MS, mass spectrometry; PLN, phospholamban; PPAR γ , peroxisome proliferator-activated receptor γ ; PTM, post-translational modification; RLC $_v$, ventricle myosin regulatory light chain; RPLC, reversed-phase liquid chromatography; SERCA, sarcoplasmic reticulum calcium pump; SNP, single-nucleotide polymorphism; TDMS, top-down mass spectrometry; TDP, top-down proteomics; TMC, tissue mapping center

REFERENCES

- (1) Smith, L. M.; Kelleher, N. L.; Kelleher, N. L. Consortium for Top Down, P., Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186–187.
- (2) Consortium, H. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **2019**, *574*, 187–192.
- (3) Regev, A.; Teichmann, S. A.; Lander, E. S.; Amit, I.; Benoist, C.; Birney, E.; Bodenmiller, B.; Campbell, P.; Carninci, P.; Clatworthy, M.; Clevers, H.; Deplancke, B.; Dunham, I.; Eberwine, J.; Eils, R.; Enard, W.; Farmer, A.; Fugger, L.; Götting, B.; Hacohen, N.; Haniffa, M.; Hemberg, M.; Kim, S.; Klenerman, P.; Kriegstein, A.; Lein, E.; Linnarsson, S.; Lundberg, E.; Lundberg, J.; Majumder, P.; Marioni, J. C.; Merad, M.; Mhlanga, M.; Nawijn, M.; Netea, M.; Nolan, G.; Pe'er, D.; Philippakis, A.; Ponting, C. P.; Quake, S.; Reik, W.; Rozenblatt-Rosen, O.; Sanes, J.; Satija, R.; Schumacher, T. N.; Shalek, A.; Shapiro, E.; Sharma, P.; Shin, J. W.; Stegle, O.; Stratton, M.; Stubbington, M. J. T.; Theis, F. J.; Uhlen, M.; van Oudenaarden, A.; Wagner, A.; Watt, F.; Weissman, J.; Wold, B.; Xavier, R.; Yosef, N. Human Cell Atlas Meeting, P., The Human Cell Atlas. *Elife* **2017**, *6*, No. e27041.
- (4) Neumann, E. K.; Patterson, N. H.; Allen, J. L.; Migas, L. G.; Yang, H.; Brewer, M.; Anderson, D. M.; Harvey, J.; Gutierrez, D. B.; Harris, R. C.; deCaestecker, M. P.; Fogo, A. B.; Van de Plas, R.; Caprioli, R. M.; Spraggins, J. M. Protocol for multimodal analysis of human kidney tissue by imaging mass spectrometry and CODEX multiplexed immunofluorescence. *STAR Protoc.* **2021**, *2*, 100747.
- (5) Saka, S. K.; Wang, Y.; Kishi, J. Y.; Zhu, A.; Zeng, Y.; Xie, W.; Kirli, K.; Yapp, C.; Cicconet, M.; Beliveau, B. J.; Lapan, S. W.; Yin, S.; Lin, M.; Boyden, E. S.; Kaeser, P. S.; Pihan, G.; Church, G. M.; Yin, P. ImmunoSABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **2019**, *37*, 1080–1090.
- (6) Giesen, C.; Wang, H. A. O.; Schapiro, D.; Zivanovic, N.; Jacobs, A.; Hattendorf, B.; Schüffler, P. J.; Grolimund, D.; Buhmann, J. M.;

Brandt, S.; Varga, Z.; Wild, P. J.; Günther, D.; Bodenmiller, B. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **2014**, *11*, 417–422.

(7) Keren, L.; Bosse, M.; Thompson, S.; Risom, T.; Vijayaragavan, K.; McCaffrey, E.; Marquez, D.; Angoshtari, R.; Greenwald, N. F.; Fienberg, H.; Wang, J.; Kambham, N.; Kirkwood, D.; Nolan, G.; Montine, T. J.; Galli, S. J.; West, R.; Bendall, S. C.; Angelo, M. MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci. Adv.* **2019**, *5*, No. eaax5851.

(8) Ptacek, J.; Locke, D.; Finck, R.; Cvijic, M.-E.; Li, Z.; Tarolli, J. G.; Aksoy, M.; Sigal, Y.; Zhang, Y.; Newgren, M.; Finn, J. Multiplexed ion beam imaging (MIBI) for characterization of the tumor microenvironment across tumor types. *Lab. Invest.* **2020**, *100*, 1111–1123.

(9) 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*, 2458–2467.

(10) 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*, 2156–2164.

(11) Smith, L. M.; Agar, J. N.; Chamot-Rooke, J.; Danis, P. O.; Ge, Y.; Loo, J. A.; Paša-Tolić, L.; Tsybin, Y. O.; Kelleher, N. L. Consortium for Top-Down, P., The Human Proteoform Project: Defining the human proteome. *Sci. Adv.* **2021**, *7*, No. eabk0734.

(12) Lee, J. E.; Kellie, J. F.; Tran, J. C.; Tipton, J. D.; Catherman, A. D.; Thomas, H. M.; Ahlf, D. R.; Durbin, K. R.; Vellaichamy, A.; Ntai, I.; Marshall, A. G.; Kelleher, N. L. A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2183–2191.

(13) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018.

(14) Hollas, M. A. R.; Robey, M. T.; Fellers, R. T.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L. The Human Proteoform Atlas: a FAIR community resource for experimentally derived proteoforms. *Nucleic Acids Res.* **2021**, *50*, D526–D533.

(15) Toby, T. K.; Fornelli, L.; Srzentić, 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*, 119–152.

(16) Wessel, D.; Flüggé, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **1984**, *138*, 141–143.

(17) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **2017**, *16*, 609–618.

(18) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Shams, D. P.; Thomas, P. M.; Kelleher, N. L. Accurate Estimation of Context-Dependent False Discovery Rates in Top-Down Proteomics. *Mol. Cell. Proteomics* **2019**, *18*, 796–805.

(19) Chen, Y.-C.; Sumandea, M. P.; Larsson, L.; Moss, R. L.; Ge, Y. Dissecting human skeletal muscle troponin proteoforms by top-down mass spectrometry. *J. Muscle Res. Cell Motil.* **2015**, *36*, 169–181.

(20) 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*, 4140–4145.

(21) Ntai, I.; LeDuc, R. D.; Fellers, R. T.; Erdmann-Gilmore, P.; Davies, S. R.; Rumsey, J.; Early, B. P.; Thomas, P. M.; Li, S.; Compton, P. D.; Ellis, M. J. C.; Ruggles, K. V.; Fenyö, D.; Boja, E. S.; Rodriguez, H.; Townsend, R. R.; Kelleher, N. L. Integrated Bottom-Up and Top-Down Proteomics of Patient-Derived Breast Tumor Xenografts. *Mol. Cell. Proteomics* **2016**, *15*, 45–56.

(22) Melani, R. D.; Gerbasi, V. R.; Anderson, L. C.; Sikora, J. W.; Toby, T. K.; Hutton, J. E.; Butcher, D. S.; Negrão, F.; Seckler, H. S.; Srzentić, 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*, 411–418.

(23) Faserl, K.; Sarg, B.; Kremser, L.; Lindner, H. Optimization and evaluation of a sheathless capillary electrophoresis-electrospray ionization mass spectrometry platform for peptide analysis: comparison to liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chem.* **2011**, *83*, 7297–7305.

(24) Li, Y.; Champion, M. M.; Sun, L.; Champion, P. A. D.; Wojcik, R.; Dovichi, N. J. Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry as an alternative proteomics platform to ultraperformance liquid chromatography-electrospray ionization-tandem mass spectrometry for samples of intermediate complexity. *Anal. Chem.* **2012**, *84*, 1617–1622.

(25) McCool, E. N.; Liangliang, S. Comparing nanoflow reversed-phase liquid chromatography-tandem mass spectrometry and capillary zone electrophoresis-tandem mass spectrometry for top-down proteomics. *Se Pu* **2019**, *37*, 878–886.

(26) Han, X.; Wang, Y.; Aslanian, A.; Fonslow, B.; Graczyk, B.; Davis, T. N.; Yates, J. R., 3rd In-line separation by capillary electrophoresis prior to analysis by top-down mass spectrometry enables sensitive characterization of protein complexes. *J. Proteome Res.* **2014**, *13*, 6078–6086.

(27) Nowak, P. M.; Sekula, E.; Kościelniak, P. Assessment and Comparison of the Overall Analytical Potential of Capillary Electrophoresis and High-Performance Liquid Chromatography Using the RGB Model: How Much Can We Find Out? *Chromatographia* **2020**, *83*, 1133–1144.

(28) Latta, S. C.; Howell, C. A.; Dettling, M. D.; Cormier, R. L. Use of data on avian demographics and site persistence during overwintering to assess quality of restored riparian habitat. *Conserv. Biol.* **2012**, *26*, 482–492.

(29) Smith, L. M.; Thomas, P. M.; Shortreed, M. R.; Schaffer, L. V.; Fellers, R. T.; LeDuc, R. D.; Tucholski, T.; Ge, Y.; Agar, J. N.; Anderson, L. C.; Chamot-Rooke, J.; Gault, J.; Loo, J. A.; Paša-Tolić, L.; Robinson, C. V.; Schlüter, H.; Tsybin, Y. O.; Vilaseca, M.; Vizcaino, J. A.; Danis, P. O.; Kelleher, N. L. A five-level classification system for proteoform identifications. *Nat. Methods* **2019**, *16*, 939–940.

(30) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top down proteomics. *J. Proteome Res.* **2014**, *13*, 3231–3240.

(31) LeDuc, R. D.; Kelleher, N. L. Using ProSight PTM and related tools for targeted protein identification and characterization with high mass accuracy tandem MS data. *Current Protocols in Bioinformatics*; Wiley, 2007, Chapter 13, Unit 13.6.

(32) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y.-B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **2007**, *35*, W701–W706.

(33) Klei, T. R. L.; Meinderts, S. M.; van den Berg, T. K.; van Bruggen, R. From the Cradle to the Grave: The Role of Macrophages in Erythropoiesis and Erythrophagocytosis. *Front. Immunol.* **2017**, *8*, 73.

- (34) Zhang, J.-N.; Michel, U.; Lenz, C.; Friedel, C. C.; Köster, S.; d'Hedouville, Z.; Tönges, L.; Urlaub, H.; Bähr, M.; Lingor, P.; Koch, J. C. Calpain-mediated cleavage of collapsin response mediator protein-2 drives acute axonal degeneration. *Sci. Rep.* **2016**, *6*, 37050.
- (35) Taghian, K.; Lee, J. Y.; Petratos, S. Phosphorylation and cleavage of the family of collapsin response mediator proteins may play a central role in neurodegeneration after CNS trauma. *J. Neurotrauma* **2012**, *29*, 1728–1735.
- (36) Shinkai-Ouchi, F.; Yamakawa, Y.; Hara, H.; Tobiume, M.; Nishijima, M.; Hanada, K.; Hagiwara, K. i. Identification and structural analysis of C-terminally truncated collapsin response mediator protein-2 in a murine model of prion diseases. *Proteome Sci.* **2010**, *8*, 53.
- (37) Morales, X.; Peláez, R.; Garasa, S.; Ortiz de Solórzano, C.; Rouzaut, A. CRMP2 as a Candidate Target to Interfere with Lung Cancer Cell Migration. *Biomolecules* **2021**, *11*, 1533.
- (38) Colca, J. R.; McDonald, W. G.; Waldon, D. J.; Leone, J. W.; Lull, J. M.; Bannow, C. A.; Lund, E. T.; Mathews, W. R. Identification of a novel mitochondrial protein ("mitoNEET") cross-linked specifically by a thiazolidinedione photoprobe. *Am. J. Physiol. Endocrinol. Metab.* **2004**, *286*, E252–E260.
- (39) Kusminski, C. M.; Holland, W. L.; Sun, K.; Park, J.; Spurgin, S. B.; Lin, Y.; Askew, G. R.; Simcox, J. A.; McClain, D. A.; Li, C.; Scherer, P. E. MitoNEET-driven alterations in adipocyte mitochondrial activity reveal a crucial adaptive process that preserves insulin sensitivity in obesity. *Nat. Med.* **2012**, *18*, 1539–1549.
- (40) Habener, A.; Chowdhury, A.; Echtermeyer, F.; Lichtinghagen, R.; Theilmeier, G.; Herzog, C. MitoNEET Protects HL-1 Cardiomyocytes from Oxidative Stress Mediated Apoptosis in an In Vitro Model of Hypoxia and Reoxygenation. *PLoS One* **2016**, *11*, No. e0156054.
- (41) Wiley, S. E.; Paddock, M. L.; Abresch, E. C.; Gross, L.; van der Geer, P.; Nechushtai, R.; Murphy, A. N.; Jennings, P. A.; Dixon, J. E. The outer mitochondrial membrane protein mitoNEET contains a novel redox-active 2Fe-2S cluster. *J. Biol. Chem.* **2007**, *282*, 23745–23749.
- (42) Furihata, T.; Takada, S.; Kakutani, N.; Maekawa, S.; Tsuda, M.; Matsumoto, J.; Mizushima, W.; Fukushima, A.; Yokota, T.; Enzan, N.; Matsushima, S.; Handa, H.; Fumoto, Y.; Nio-Kobayashi, J.; Iwanaga, T.; Tanaka, S.; Tsutsui, H.; Sabe, H.; Kinugawa, S. Cardiac-specific loss of mitoNEET expression is linked with age-related heart failure. *Commun. Biol.* **2021**, *4*, 138.
- (43) Xu, D.; Lu, W. Defensins: A Double-Edged Sword in Host Immunity. *Front. Immunol.* **2020**, *11*, 764.
- (44) Faurschou, M.; Sørensen, O. E.; Johnsen, A. H.; Askaa, J.; Borregaard, N. Defensin-rich granules of human neutrophils: characterization of secretory properties. *Biochim. Biophys. Acta* **2002**, *1591*, 29–35.
- (45) Sankaran-Walters, S.; Hart, R.; Dills, C. Guardians of the Gut: Enteric Defensins. *Front. Microbiol.* **2017**, *8*, 647.
- (46) Ghosh, D.; Porter, E.; Shen, B.; Lee, S. K.; Wilk, D.; Drazba, J.; Yadav, S. P.; Crabb, J. W.; Ganz, T.; Bevins, C. L. Paneth cell trypsin is the processing enzyme for human defensin-5. *Nat. Immunol.* **2002**, *3*, 583–590.
- (47) Varney, K. M.; Bonvin, A. M. J. J.; Pazgier, M.; Malin, J.; Yu, W.; Ateh, E.; Oashi, T.; Lu, W.; Huang, J.; Diepeveen-de Buin, M.; Bryant, J.; Breukink, E.; Mackerell, A. D., Jr.; de Leeuw, E. P. H. Turning defense into offense: defensin mimetics as novel antibiotics targeting lipid II. *PLoS Pathog.* **2013**, *9*, No. e1003732.
- (48) Pachón-Ibáñez, M. E.; Smani, Y.; Pachón, J.; Sánchez-Céspedes, J. Perspectives for clinical use of engineered human host defense antimicrobial peptides. *FEMS Microbiol. Rev.* **2017**, *41*, 323–342.
- (49) Mannervik, B.; Awasthi, Y. C.; Board, P. G.; Hayes, J. D.; Di Ilio, C.; Ketterer, B.; Listowsky, I.; Morgenstern, R.; Muramatsu, M.; Pearson, W. R.; Pickett, C. B.; Sato, K.; Widersten, M.; Wolf, C. R. Nomenclature for human glutathione transferases. *Biochem. J.* **1992**, *282*, 305–306.
- (50) Oakley, A. Glutathione transferases: a structural perspective. *Drug Metab. Rev.* **2011**, *43*, 138–151.
- (51) Phan, L.; Jin, Y.; Zhang, H.; Qiang, W.; Shekhtman, E.; Shao, D.; Revoe, D.; Villamarin, R.; Ivanchenko, E.; Kimura, M.; Wang, Z. Y.; Hao, L.; Sharopova, N.; Bihan, M.; Sturcke, A.; Lee, M.; Popova, N.; Wu, W.; Bastiani, C.; Ward, M.; Holmes, J. B.; Lyoshin, V.; Kaur, K.; Moyer, E.; Feolo, M.; Kattman, B. L. ALFA: Allele Frequency Aggregator. www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/ (accessed Dec 13, 2021).
- (52) Jakobsson, P.-J.; Mancini, J. A.; Riendeau, D.; Ford-Hutchinson, A. W. Identification and characterization of a novel microsomal enzyme with glutathione-dependent transferase and peroxidase activities. *J. Biol. Chem.* **1997**, *272*, 22934–22939.
- (53) Morgenstern, R.; Zhang, J.; Johansson, K. Microsomal glutathione transferase 1: mechanism and functional roles. *Drug Metab. Rev.* **2011**, *43*, 300–306.
- (54) Forrester, M. T.; Hess, D. T.; Thompson, J. W.; Hultman, R.; Moseley, M. A.; Stamler, J. S.; Casey, P. J. Site-specific analysis of protein S-acylation by resin-assisted capture. *J. Lipid Res.* **2011**, *52*, 393–398.
- (55) Frank, K.; Kranias, E. G. Phospholamban and cardiac contractility. *Ann. Med.* **2000**, *32*, 572–578.
- (56) Brown, K. A.; Chen, B.; Guardado-Alvarez, T. M.; Lin, Z.; Hwang, L.; Ayaz-Guner, S.; Jin, S.; Ge, Y. A photocleavable surfactant for top-down proteomics. *Nat. Methods* **2019**, *16*, 417–420.
- (57) Zhou, T.; Li, J.; Zhao, P.; Liu, H.; Jia, D.; Jia, H.; He, L.; Cang, Y.; Boast, S.; Chen, Y.-H.; Thibault, H.; Scherrer-Crosbie, M.; Goff, S. P.; Li, B. Palmitoyl acyltransferase Aph2 in cardiac function and the development of cardiomyopathy. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 15666–15671.
- (58) Gregorich, Z. R.; Cai, W.; Lin, Z.; Chen, A. J.; Peng, Y.; Kohmoto, T.; Ge, Y. Distinct sequences and post-translational modifications in cardiac atrial and ventricular myosin light chains revealed by top-down mass spectrometry. *J. Mol. Cell. Cardiol.* **2017**, *107*, 13–21.
- (59) 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.
- (60) Zhang, J.; Guy, M. J.; Norman, H. S.; Chen, Y.-C.; Xu, Q.; Dong, X.; Guner, H.; Wang, S.; Kohmoto, T.; Young, K. H.; Moss, R. L.; Ge, Y. Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure. *J. Proteome Res.* **2011**, *10*, 4054–4065.