

## TECHNICAL BRIEF

# Plasma Extraction by Acetonitrile and Trichloroacetic Acid With Dithiothreitol and Combination for Low Molecular Weight Proteome Analysis

Ting Liu<sup>1,2</sup> | Shenjian Xu<sup>1</sup> | Chengzhuo Chu<sup>1</sup> | Kun Jiang<sup>1</sup> | Jiaojiao Sha<sup>3</sup> | Liangjia Du<sup>4</sup> | Dezhu Chen<sup>5</sup> |  
Minqi Cai<sup>5</sup> | Yiqiang Chen<sup>6</sup> | Bing Bai<sup>1,3,4,7</sup>

<sup>1</sup>Department of Laboratory Medicine, Nanjing Drum Tower Hospital Clinical College of Nanjing Medical University, Nanjing, Jiangsu, China | <sup>2</sup>The Affiliated Suqian First People's Hospital of Nanjing Medical University, Suqian, Jiangsu, China | <sup>3</sup>Department of Laboratory Medicine, Nanjing Drum Tower Hospital Clinical College of Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China | <sup>4</sup>Department of Laboratory Medicine, Nanjing Drum Tower Hospital Clinical College of Nanjing Normal University, Nanjing, Jiangsu, China | <sup>5</sup>Department of Nuclear Medicine, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, Jiangsu, China | <sup>6</sup>Center For Precision Medicine, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, Jiangsu, China | <sup>7</sup>Department of Laboratory Medicine, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, Jiangsu, China

**Correspondence:** Bing Bai (bing.bai@nju.edu.cn)

**Received:** 29 May 2025 | **Revised:** 6 November 2025 | **Accepted:** 8 November 2025

**Keywords:** acetonitrile | DTT | LC-MS/MS | lower molecular weight proteome | plasma | proteome | TCA

## ABSTRACT

The low molecular weight (LMW) proteome fraction comprises small proteins and endogenous peptide fragments present in peripheral circulation that could serve as potential diagnostic markers; however, the coverage of the mass spectrometry-based plasma or serum LMW proteomic analysis is limited. We found that 60%, 70%, 80%, and 90% acetonitrile and 10% trichloroacetic acid (TCA) with or without dithiothreitol (DTT) extracted different LMW proteins (below 50 kD or smaller). Extraction by 60%, 80% acetonitrile, and 10% TCA with or without DTT in the plasma and pooling after tryptic digestion for in-depth proteomic analysis by high pH reverse phase fractionation has led to the identification of 5638 tryptic peptides from 805 proteins. This comprehensive plasma LMW proteomic analysis is expected to promote success in blood biomarker discoveries.

## 1 | Introduction

Lower molecular weight (LMW) protein species in the plasma, including small proteins and peptide fragments, are often involved in diseases and able to reflect physiological homeostasis of the body, such as endorphins, angiotensin, natriuretic peptides, and glucagon-like peptide 1 [1–5]. Analyses of these protein species by the liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) have been intensively studied [6–8] and have achieved great advances [3, 9, 10]. However, due to detection interference from high levels of proteolytically cleaved fragments from abundant proteins (e.g., albumin, immunoglob-

ulins, complements, etc.), and very low levels with instability of disease-related protein fragments or peptides, the coverage of these studies is generally low in which only hundreds or a few thousands of tryptic peptides and tens or a hundred of proteins or so could be identified [11–13], limiting the discovery of plasma endogenous protein fragments or peptides as potential biomarkers, mechanistic molecules, or even therapeutic targets.

Methods for extraction of plasma LMW protein species include organic solvent or acid precipitation [14], ultrafiltration by molecular weight cutoff membranes [15], absorption by solid phase cartridges or C18 resins [16], size exclusion chromatography [17],

and immunoaffinity peptide enrichment [18], each of which has cons and pros [19–21]. As an organic solvent, acetonitrile (ACN) is commonly used in plasma peptide fragments and LMW protein extraction by precipitating large proteins efficiently, and was found to be more effective and reproducible than other solvents [22–24]. ACN is also easily removed by evaporation and is thus adaptable to high-throughput automation pipelines [25]. Trichloroacetic acid (TCA) is also sufficient to precipitate highly abundant proteins from plasma [3, 15, 23]. Both organic solvents and acids are able to denature and pellet large proteins quickly, including proteases, preventing protein species from degradation [3].

Dithiothreitol (DTT) is a common protein-reducing chemical that is frequently used to precipitate abundant disulfide bond-rich proteins in plasma, such as albumin and transferrin [26–29]. It has been demonstrated to be reproducible and quantitative as a distinct approach complementary to other serum extraction methods [29].

We now find that different concentrations of ACN, 10% TCA, and DTT for extraction can yield distinct subsets of LMW protein species and, therefore, combine them to expand the plasma LMW proteome profiling for increased success in blood biomarker discoveries.

## 2 | Materials and Methods

### 2.1 | Human Plasma Sample Collection

This study was approved by the Ethics Committee of our Nanjing Drum Tower Hospital (IRB Review Approval #: 2022-165-01). Human EDTA-anticoagulated blood samples were collected from our clinical laboratory after the tests were all completed. These samples were from individuals who visited our hospital for health screening without any abnormal results in blood cell and biochemical tests, and the infectious disease screening (hepatitis B, C and human immunodeficiency virus). Samples with observable hemolysis were excluded.

0.5 mL of each whole blood sample was centrifuged first at 5000 g for 5 min to pellet cells and platelets, and then the supernatant was taken and centrifuged again at 20,000 g for 5 min before harvest. All samples were processed within 6 h after isolation from body. The plasma from multiple samples was pooled to reach sufficient amount for extraction in the collection day. For the in-depth proteomic analysis, in each collection day, 200-μL plasma from each sample was taken and pooled, and stored at –80°C until all 100 samples were collected. When the in-depth proteomic analysis was performed, all pooled samples were thawed and combined altogether again for ACN and TCA extractions.

### 2.2 | Plasma LMW Protein Extraction and Tryptic Digestion

For plasma extraction by ACN, ACN (Thermo Scientific, USA) was added to the plasma to reach the final concentrations of 60% (v/v), 70% (v/v), 80% (v/v), or 90% (v/v), respectively, and

briefly vortexed, followed by slow rotation at room temperature (RT) for 30 min. For TCA extraction, 100% (w/v) TCA (Sigma-Aldrich, USA) was added to plasma to 10% (v/v) and then vortexed briefly before incubation on ice for 2 h. After incubation at RT or on ice, the samples were centrifuged at 20,000 g for 10 min for collection of the supernatants. The supernatants of acetonitrile extraction were directly dried on Speedvac, while the supernatant of TCA extraction was desalted by a C18 column prior to drying on Speedvac.

When plasma was extracted in the presence of DTT, 1-M DTT (Thermo Scientific, Baltics UAB) was added to the plasma to 10 mM before extraction by acetonitrile or TCA.

For enzymatic digestion, 8-M urea in 50-mM triethylammonium bicarbonate buffer (TEAB, Sigma-Aldrich, Supelco, USA) was added to the dried sample extracted by acetonitrile or TCA extracts, and then LysC was added to the final concentration of 10-ng/μL (enzyme/substrate = 1/100). The samples were kept at RT for solubilization and digestion. After 2 h, urea in the samples was diluted to 2 M by addition of three volumes of 50-mM TEAB prior to digestion by 10-ng/μL trypsin at RT for 4 h or overnight. After that, the digested samples were reduced by 10-mM DTT for 30 min and then alkylated by 30-mM iodoacetamide (Sigma-Aldrich, USA) for 15 min in the dark at RT. The final digested extracts were acidified by 1% formic acid, desalted through C18 columns, and then dried for LC–MS/MS analysis.

### 2.3 | Extract Combination and Fractionation by Basic pH RPLC

The pooled samples from 100 clinical samples were extracted by ACN at 60% and 80%, and also 10% TCA without or with DTT in the plasma. After BCA quantification, 0.5 mg of each of these six digested and desalted tryptic peptide samples were taken and mixed together. After drying, the mixed tryptic peptides were resuspended in Solvent A (10-mM ammonium formate, pH 8.0), and loaded onto the column (XBridge, C18, Waters, 3.5-μm resin, 4.6 × 250 mm<sup>2</sup>, ~8-mg estimated loading capacity) for RPLC (reverse phase liquid chromatography) at basic pH, where they were then mainly separated by 15%~45% gradient of Solvent B (90% ACN in Solvent A, pH 8.0). About 50 fractions were collected and dried.

### 2.4 | LC–MS/MS Analysis and Data Processing

The digested and dried tryptic peptide samples were resuspended in 5% acetic acid and 1 μg was loaded to the machine (Q Exactive HF-X, Thermo), through a column (50 μm internal diameter, 30 cm long) packed with 1.9-μm C18 resins (Dr. Maisch GmbH, Germany) and separated mainly by a 5%~35% gradient of Solvent B (80% ACN, 0.2% formic acid, 3% DMSO) with Solvent A (0.2% formic acid, 3% DMSO) for 30 min or 1 h. For the tryptic peptide samples fractionated by basic pH RPLC, samples collected in the middle of fractionation were first chosen to run individually by a 1-h gradient. After the MS signals of these samples were found to be not high enough, the sequential neighboring samples were combined to run. The tryptic peptides that were not dissolved in

the loading buffer during the basic pH RPLC process were spun down and run separately as the last sample by a 2-h gradient. Some neighboring fractions might be combined to increase the MS signal, and finally 25 samples were analyzed.

The MS raw data files were processed by the Proteome Discoverer 2.4 software (ThermoFisher) and searched by the Sequest HT engine against the human proteome database (UP000005640, including 20,402 reviewed proteins plus 62,983 unreviewed proteins) from the Uniprot. Specific parameters include the following: semi-trypsin, maximum cleavages of 2, peptide length ranging 6–144, 10 ppm in precursor mass tolerance and 0.02 Da in fragment mass tolerance, b and y ions in spectrum matching, maximum modifications of 3 and the dynamic modification of 15.995 Da (M) on methionine, no dynamic modifications on either terminus of peptides, N-terminal acetylation (+42.011 Da), methionine loss (−131.010 Da) or both (−69.030) in the dynamic modifications of proteins, and the static modification of carbamidomethyl (+57.021 Da).

All MS raw files and the search files were stored at the public server Integrated Proteome Resources iProX (<https://www.iprox.cn/page/home.html>) [30, 31], and the project ID is IPX0011263000 (PXD061482).

## 2.5 | Data Analysis

In protein abundance rank analysis, data were from the report of a deep human serum proteomic study [32] or the public HPA (human protein atlas) database [33]. Proteins from these datasets were ranked according to their number of PSMs in the report or their absolute concentrations provided in the HPA database. Proteins with the largest PSM or the highest concentration were ranked as “1,” and those with smaller PSMs or at lower levels were ranked lower.

The functional annotation analysis was performed by the DAVID (The Database for Annotation, Visualization, and Integrated Discovery) analysis tool available on the website of the American National Institutes of Health with the default parameter settings (<https://davidbioinformatics.nih.gov/tools.jsp>). The biological function, cellular component, and molecular function under gene ontology were selected for analyses, and those with significant *p* values, Bonferroni values, Benjamini values, and false discovery rates simultaneously were selected.

Novel tryptic peptides identified in this study were determined by comparison to the list of 267,383 peptides from the database of The Human Plasma PeptideAtlas (<https://peptideatlas.org/builds/human/plasma/>) [34]. Proteins identified in this study are considered novel if they were not present in the six datasets downloaded from the three public databases [Human Plasma Proteome Project Data (<https://peptideatlas.org/hupo/hppp/>) [34], Human Body Fluid Proteome (HBFP) database (<https://bmbl.bmi.osumc.edu/HBFP/>) [35], Human protein Atlas (<https://www.proteinatlas.org/>) [36, 37], and the three major plasma proteomics studies [32, 38, 39]. Proteins from each dataset were all converted to their primary gene names to remove redundancy.

## 3 | Results

### 3.1 | LMW Proteome Extraction by ACN and TCA

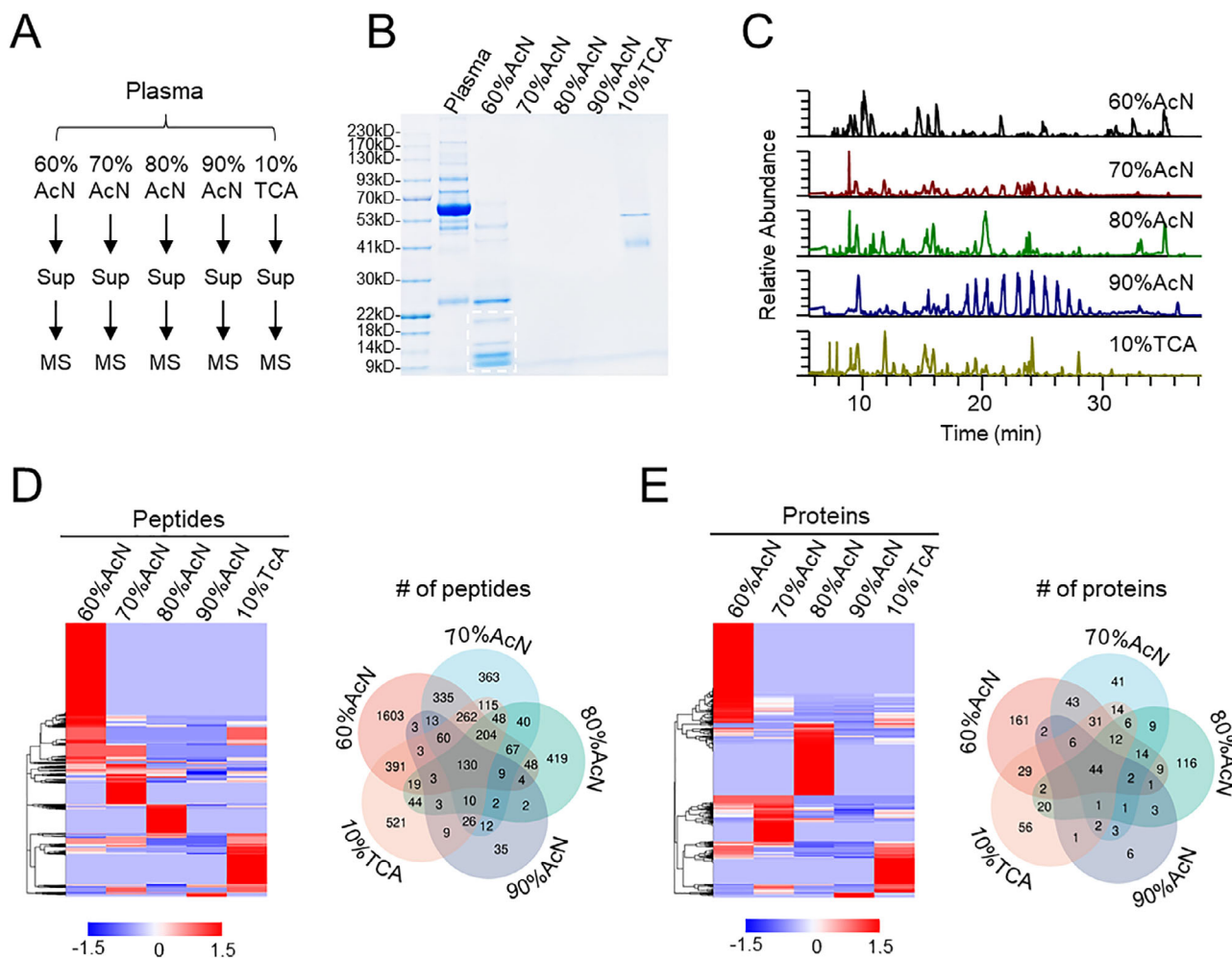
Low concentration (such as 60%) of ACN might extract more LMW protein species. Some clinically important small endogenous peptide fragments (such as angiotensin I and II that only consist 10 and 8 amino acids) are small and low, thus more likely to be enriched at higher concentrations of ACN, increasing the chance of being detected. Besides, acids precipitate proteins through different modes of action, extracting different subsets of protein species for expanded LMW proteome coverage.

We therefore added ACN to the plasma to 60%, 70%, 80%, and 90% respectively, and also TCA to 10% to extract the LMW protein species (Figure 1A), yielding around 70, 50, 40, 25, and 10 µg proteins per 100 µL plasma (Figure S1). By SDS-PAGE, compared to the original plasma, larger proteins were precipitated by 60% ACN, leaving smaller (below 50 kD, especially below 20 kD) proteins in the extract and even fewer in the extracts of 70%, 80%, and 90% ACN (Figure 1B). In contrast, 10% TCA extracted more proteins around 50 and 40 kD but fewer around 10 kD. These results indicate that the ACN at different concentrations and the TCA both extract low molecular mass proteins, exerting differential extraction effects. LC-MS/MS analyses of these five samples showed distinct liquid chromatography patterns (Figure 1C) in which 70%, 80%, and 90% ACN extracts showed much more peaks of singly charged ions (Figure S2) and polymers might exist in the 90% ACN extract, indicating differential contents in the extract under these conditions.

After database searching and data processing, the proteomics analyses identified 3154, 1696, 1052, 324, and 1848 tryptic peptides in the 60%, 70%, 80%, and 90% ACN and 10% TCA extraction samples, respectively, with distinct enrichment and unique tryptic peptides in each (Table S1 and Figure 1D). It is notable that about 60% of the identified tryptic peptides in each sample had two or more PSMs (peptide-spectrum matches) (Figure S3). Correspondingly, proteins mapped by these identified tryptic peptides also displayed different enrichments in these extracts and each contained unique ones (Figures 1E and S4). Hydrophobicity of the proteins extracted by different concentrations of ACN showed statistical significance (Figure S5A), but it could not solely explain the differential extraction effects because the hydrophobicity changed inconsistently with increasing ACN concentrations. These data suggested that each of these extraction conditions could enrich different subsets of LMW protein species and were not completely replaceable by one another. The general sequence coverages seemed higher with a significant difference in extract of 60% ACN and lower in other extracts (Figure S5B,C), indicating 60% ACN extracted more full-length proteins (such as lipoprotein, hemoglobin, cystatin, etc.) while ACN at higher concentrations and 10% TCA extracted more endogenous proteins or peptide fragments from plasma.

### 3.2 | DTT in Plasma for LWM Proteome Extraction

DTT breaks disulfide bonds within proteins, facilitating protein denaturation and precipitation in solution for LWM protein



**FIGURE 1** | LMW plasma proteome extraction under different concentrations of acetonitrile and 10% trichloroacetic acid. (A) Workflow of the extraction procedures from plasma. (B) SDS-PAGE of the extracted samples. (C) LC-MS/MS analyses of these samples. (D, E) Heatmaps and Venn diagrams to demonstrate the relative levels and uniqueness of the peptides and proteins profiled from the proteomics results. PSMs (peptide-spectrum matches) were used as the quantitative results for analysis.

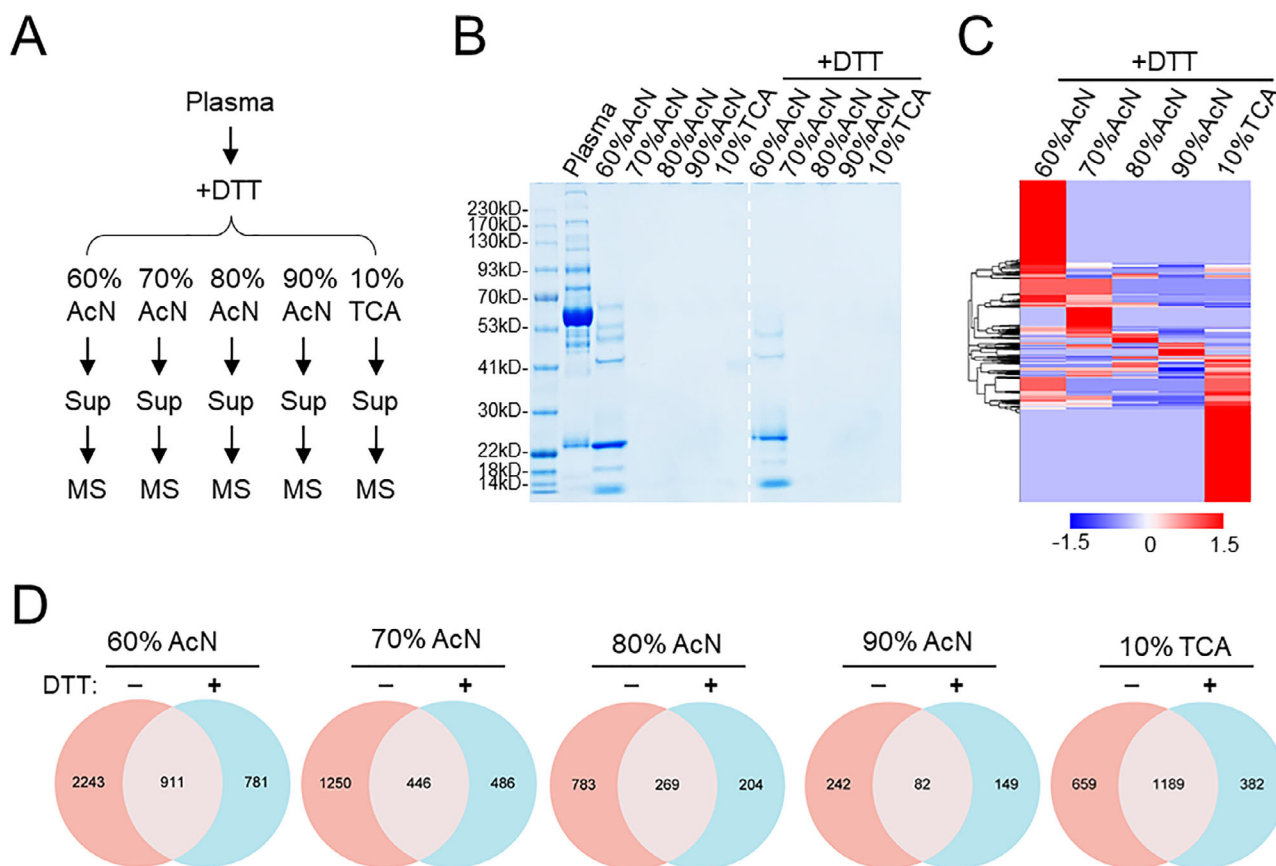
extraction [26, 40], and it was thus utilized in this study (Figure 2A). By SDS-PAGE, DTT in the plasma facilitated protein precipitation as bands around 60 and 18 kD seemed weaker in the extract of 60% ACN (Figure 2B). LC-MS/MS of these samples not only showed differences among their LC chromatographs (Figure S6A), but also difference between each of them and the corresponding extraction sample without DTT in the plasma (Figure S6B). The MS identification results demonstrated again the distinct subsets of tryptic peptides in each sample, especially the large number of tryptic peptides enriched in the samples from 60% ACN and 10% TCA extractions (Figure 2C). Specifically, incorporation of DTT in the extraction led to identification of 781, 486, 204, 149, and 382 additional tryptic peptides in the extractions by 60%, 70%, 80%, 90% ACN and 10% TCA, respectively, occupying around half of all identified tryptic peptides in each extract (Figure 2D). It is notable that the protein coverages among these different extracts were also reduced as the ACN concentration increased but remained barely changed in the extract of 10% TCA (Figure S5D).

### 3.3 | In-Depth Analysis of the Plasma LWM Proteome

As the total number of tryptic peptides identified in all 10 extracted samples had reached more than 4000 (Table S1), we initially combined all 10 extracted samples altogether and analyzed them by LC-MS/MS as a single run to achieve a significantly expanded proteome coverage. However, many base peaks corresponding to monocharged  $m/z$  ions ( $z=1$ ) appeared in the LC chromatograph, and the precursor scans generally seemed noisy (Figure S7A), and only 1966 tryptic peptides mapped to 218 proteins were identified (Figure S7B). We therefore considered fractionation of the digested mixture of extracted samples for the LC-MS/MS analysis.

To simplify the extraction, we selected the extracted samples of 60% and 80% ACN together with the 10% TCA extraction without or with DTT and combined them altogether (Figure 3A). This was because the number of protein species extracted under the condition of 90% ACN was low and those in the 70% ACN extract





**FIGURE 2** | DTT (dithiothreitol) for LMW plasma proteome extraction. (A) Workflow of the extraction procedures from plasma in the presence of DTT. (B) SDS-PAGE of the samples extracted from plasma with or without DTT. (C) Heatmap to demonstrate the relative levels of the peptides profiled by the LC-MS/MS analyses. (D) The number of unique and shared peptides identified by the LC-MS/MS analyses in the absence and presence of DTT under each extraction condition.

would still be largely detectable by the in-depth proteomic analysis; in addition, these three conditions yielded deeper detection of peptide fragments from normally very low level of proteins in the plasma (Figure S4); besides, some abundant and large common proteins in the 60% ACN extract could serve as the carrier to avoid loss of low but clinically important protein species.

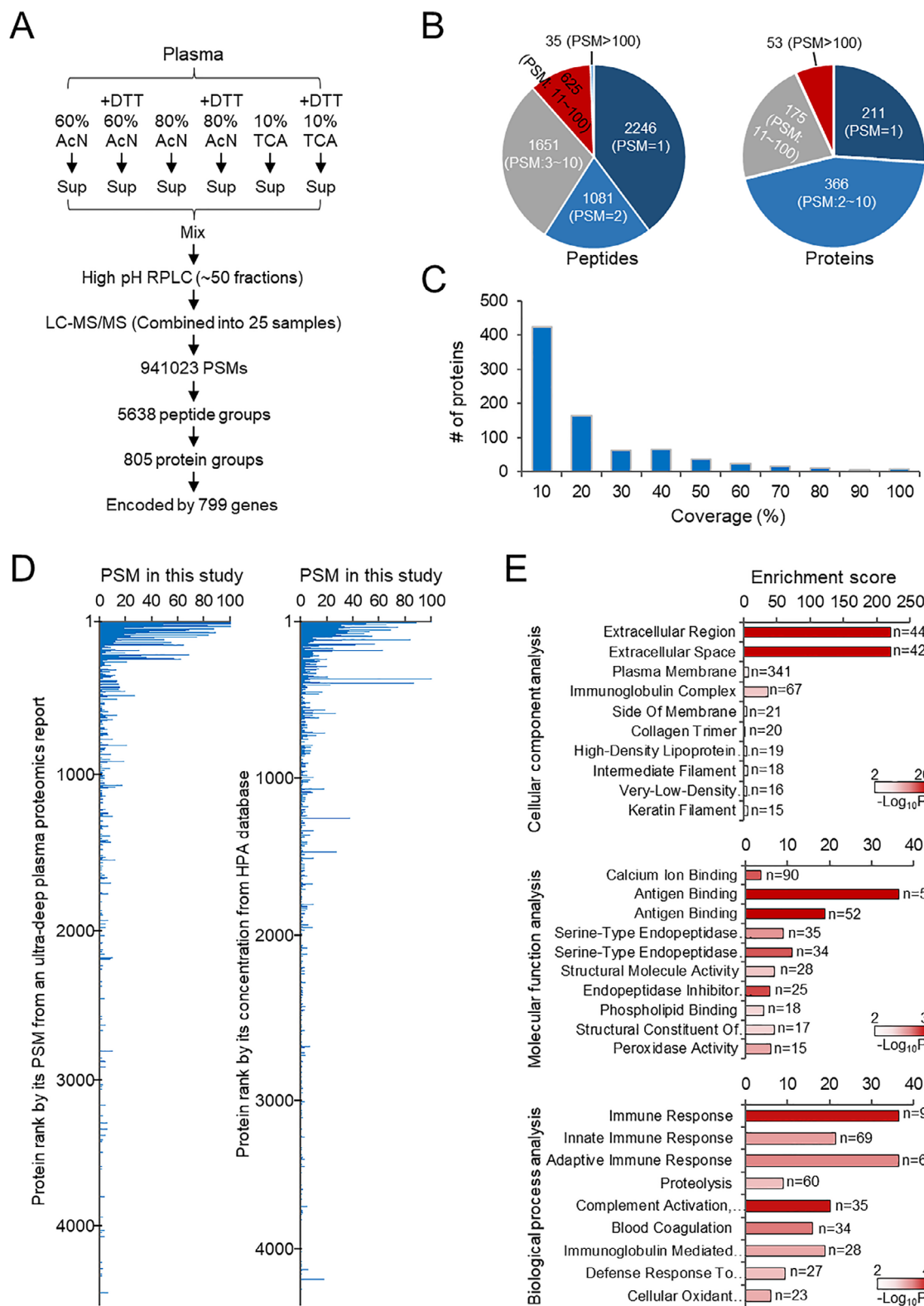
We therefore combined the tryptic peptides in equal amounts from these six extracted samples and fractionated them for in-depth proteomic analysis, yielding identification of 5638 tryptic peptides from 805 proteins in the results (Figures 3A and S8A). Among these, 2343 tryptic peptides were detected only once [PSM (peptide-spectrum match) = 1], 1131 were detected twice (PSM = 2), and the rest tryptic peptides were detected more than twice; for the identified proteins, 213 had only one PSM, and the rest had two or more PSMs (Figures 3B and S8B).

Full-length proteins and their larger fragments tend to have higher sequence coverage in the proteomic results. The data demonstrated that about 50% of the identified proteins were below 10% coverage, and about 25% were within 10%~30% coverage (Figure 3C). Compared to the proteomic analysis of the whole plasma that we just reported [41], the averaged protein sequence coverages in this LMW proteome study were much higher and reached ~90% for proteins smaller than 10 kD, but quickly decreased to around 20% and became almost consistently

lower for proteins larger than 20 kD (Figure S8C), indicating that the proteins extracted by ACN and TCA were highly likely small intact proteins (~10 kD, such as lipoproteins) and fragments from larger protein (>20 kD).

To evaluate the depth of the analysis, we utilized two datasets of plasma proteins with relative abundance (PSMs) or absolute concentrations as the reference, and found that some proteins were actually at very low levels in the plasma, indicating this in-depth proteomics analysis was indeed deep (Figure 3D). Gene ontology analyses demonstrated that the identified proteins from the extract could be enriched in certain subsets of cellular components, molecular functions, and biological processes (Figure 3E).

To see what factors that might contribute to the presence of these LMW protein species, we first compared their abundance with the levels of their entire proteins in the intact plasma as reported [32] or from a public database [33] and both demonstrated significant correlations (Figure S9A,B), indicating that the levels of the LMW protein species are related to their total level or the level of their precursor proteins. We then examined whether the longer proteins might more likely generate the LMW protein species; however, there was no significant correlation between their PSMs and the number of amino acids of their precursor proteins (Figure S9C). Further analysis of the location of the



**FIGURE 3** | In-depth analysis of the combined plasma LMW proteome extracts. (A) Workflow of the extraction procedures and in-depth proteomics analysis. (B) The number of peptides and proteins identified in the in-depth proteomics analysis in different ranges of PSMs. (C) Distribution of protein coverages. (D) Depth analysis of the proteins identified in this proteomics analysis. The data were from the report [32] or the public HPA (human protein atlas) database [37], in which proteins were ranked according to their number of PSMs in the report or their concentrations provided in the HPA database. Proteins with the largest PSM or the highest concentration were ranked as “1.” (E) Gene ontology analyses of the identified proteins in this analysis.

**TABLE 1** | Novel plasma peptide sequence identified in this study. The list is generated after comparison with the database of The Human Plasma Peptide Atlas, and only identified peptides with two or more PSMs are shown. The peptides from proteins that are not included in the six datasets from reported plasma proteomics studies [32, 38, 39] and public plasma protein databases [34–37] are highlighted.

	# PSMs	Accession	Gene	Protein
VGPVSLPR	37	A0A087WTD7	AKAP13	A-kinase anchoring protein 13
KQFLETIEK	28	P20941	PDC	Phosducin
AVQLAIGVQAVHYNGVLAAG	15	F5H563	EPB42	Erythrocyte membrane protein band 4.2
DTVCQCKEGTFR	15	O14798	TNFRSF10C	Tumor necrosis factor receptor superfamily member 10C
KETLETISNEEQTPLLKK	8	Q8NHP6	MOSPD2	Motile sperm domain-containing protein 2
VLDLGPITR	7	P55259	GP2	Pancreatic secretory granule membrane major glycoprotein GP2
RAYVEQIEK	7	Q9BSJ2	TUBGCP2	Gamma-tubulin complex component 2
YMIHEPEPHILLFR	6	P33552	CKS2	Cyclin-dependent kinases regulatory subunit 2
PSRAAPRAPSAALSSSPLLTAPHK	5	Q70SY1	CREB3L2	Cyclic AMP-responsive element-binding protein 3-like protein 2
QAAALTR	5	P78423	CX3CL1	Fractalkine
GQQGAADKR	4	J3KRS3	ARHGAP44	Rho GTPase activating protein 44
LIVQPPFLQSVHHPEFR	4	Q9H4D0	CLSTN2	Calsyntenin-2
SSCTMTR	4	O14798	TNFRSF10C	Tumor necrosis factor receptor superfamily member 10C
YFRNHQVEELR	4	Q7Z4J2	GLT6D1	Putative glycosyltransferase 6 domain-containing protein 1
GVQAVHYNGVLAAG	3	F5H563	EPB42	Erythrocyte membrane protein band 4.2
SIPENNIMR	3	P22466	GAL	Galanin peptides
VLDLGPITRR	3	P55259	GP2	Pancreatic secretory granule membrane major glycoprotein GP2
TYTLAVQGISGLPLKK	3	Q2NL82	TSR1	Pre-rRNA-processing protein TSR1 homolog
LLLYESKSR	3	Q07002	CDK18	Cyclin-dependent kinase 18
PQNTLPRVAVPEEQQGSLEQVSR	2	C9J7W8	ABHD14A	Abhydrolase domain containing 14A
GSTNLGVNQAGFTLHSAIYAARPDVK	2	E7ENY0	ADD1	Adducin 1
MGPLMVLFCLLFLYPGLADSAP	2	F2Z3N2	C2	Complement C2
MGEDAAQAEKFQHPG	2	O95810	CAVIN2	Caveolae-associated protein 2
IVQPPFLQSVHHPEFR	2	Q9H4D0	CLSTN2	Calsyntenin-2
LAIGVQAVHYNGVLAAG	2	F5H563	EPB42	Erythrocyte membrane protein band 4.2
LVSQVGR	2	Q6UWH4	GASK1B	Golgi-associated kinase 1B
WVWAQPVTTGK	2	P40197	GP5	Platelet glycoprotein V
TEGFPKDSR	2	Q9ULI3	HEG1	Protein HEG homolog 1
CFPAGAQVR	2	Q14623	IHH	Indian hedgehog protein
EKLCGHHFVR	2	M0QXQ3	INSL3	Insulin like 3
DAVPGEAALQAR	2	Q96MG2	JSRP1	Junctional sarcoplasmic reticulum protein 1
SSFSEHKPR	2	Q9HCM3	KIAA1549	UPF0606 protein KIAA1549
FGALTAEK	2	Q5QPQ0	LYPLA2	Palmitoyl-protein hydrolase
FAGLPETGR	2	A0A087WZS5	MMP25	HCG15613, isoform CRA_b
NVDTLTPLIKK	2	Q9ULW6	NAPIL2	Nucleosome assembly protein 1-like 2
LPICPGGAAR	2	A0A494C1P2	PRL	Prolactin
AVGLLTVISK	2	O43653	PSCA	Prostate stem cell antigen
ETESIEK	2	Q14D04	VEPH1	Ventricular zone-expressed PH domain-containing protein homolog 1
PSDLAELR	2	Q96HV5	TMEM41A	Transmembrane protein 41A

identified tryptic peptides on their precursor proteins showed that these peptides were nearly evenly distributed at any region of the proteins without a clear bias to the N-terminal, middle, or C-terminal regions overall (Figure S9D). We also looked to see whether it was possibly related to half-lives of their precursor proteins in plasma, but the correlation was insignificant (Figure S9E). Besides, we performed the INTERPRO analysis and found that the “Ig-like\_fold” domain was enriched in the 155 proteins (among the total 820 proteins identified) (Figure S9F), indicating that the generation of LMW protein species might be related to particular domains within the proteins.

### 3.4 | Novel Tryptic Peptide Sequences and Proteins in the In-Depth Analysis

To validate whether the tryptic peptides identified were reliable, we checked if they were present in the database of The Human Plasma PeptideAtlas (267,383 peptide records) [34]. Among the 5638 identified tryptic peptides, 5303 could directly be searched out. After consideration of partial trypticity and miscleavages, more sequences of tryptic peptides could be found in this database, leaving 142 not recorded, including 39 with two or more PSMs in the in-depth proteomic analysis of this study (Table 1). It was notable that among the 39 tryptic peptides, there were six from novel proteins that are not reported in the deep plasma proteomic studies [32, 38, 39] or recorded in the public plasma protein databases [34–37] (Table 1).

## 4 | Discussion and Conclusion

In this study, we used acetonitrile at different concentrations and TCA with or without dithiothreitol to extract proteins from plasma and then combined the extracts for in-depth proteomics analysis. This yielded an expanded proteome coverage larger than many reported studies in recent years [11, 13, 42, 43], comparable to the largest scales in the field [15, 21, 44].

The distinct extraction effects of acetonitrile at different concentrations were unexpected because protein species extracted by 60% acetonitrile should cover all those extracted by acetonitrile at higher concentrations. This is probably because the plasma endogenous peptide fragments extracted by acetonitrile at higher concentrations are generally low (Figure S1), and thus are not easily detectable in the presence of highly abundant protein species in the 60% ACN extraction. Besides, different concentrations of acetonitrile extract different subsets of nonpeptidic substances (such as lipids, sugars, and other metabolites) as indicated by MS1 scans in which many prominent  $m/z$  ions bear only “1” charge and their masses are different across extracted samples (Figures 1C and S2). This might also contribute to different interfering effects in peptide detection.

Indeed, these hydrophobic interfering substances coextracted especially by acetonitrile is a major problem in the plasma LMW proteome analysis, making a simple combination of the extracts fail to achieve an expanded proteome, probably because it also adds up the interfering substances in the samples (Figure S7). This common challenge in plasma peptide fragment extraction

by organic solvents [3, 5, 45] can be overcome by basic pH RPLC fractionation prior to LC–MS/MS in our study.

The observation that distinct subsets of tryptic peptides identified in the TCA extract may be explained as that these protein species are held to larger proteins through salt bridges or other associations, which are destroyed by the strong acid TCA.

The DTT has exerted a dramatic effect in the extraction of a large number of extra protein species. Besides the expected mechanism that DTT disrupts the internal disulfide bond of a protein and thus leads to structural collapse, the LMW protein species might also crosslink to large proteins through disulfide bonds between them and therefore co-precipitate with them. Upon breakage of these disulfide bonds by DTT, these small protein species will be released from the precipitated proteins, retained in the supernatant of the organic solvent. Indeed, for example, in the precipitation by 60% acetonitrile, there were 578 identified tryptic peptides (2814 PSMs) that contain cysteines, comprising 34.2% of the total tryptic peptides identified and 40.8% of the total PSMs; in contrast, in the 60% acetonitrile extraction without DTT, only 24.4% tryptic peptides (27.1% PSMs) contain cysteines (Table S1).

In this study, we have used data-dependent acquisition (DDA) for the mass spectrometric analysis; however, evaluation using data-independent acquisition (DIA) is currently becoming a standard for deeper proteomic analysis as it excels in its high reproducibility, quantitative accuracy, and comprehensive analytical capability. With DIA, the number of identified tryptic peptides might be more remarkably increased [46–48]. Besides, the LMW protein extraction efficiency could be even more improved by adjusting the salts, pH, and temperature conditions [14]. Moreover, tryptic peptides bearing mutations could also be identified with the peptide-centric proteomics software like PepQuery [49].

Overall, a combination of the extractions by acetonitrile and TCA followed by fractionation could expand the plasma lower molecular proteome coverage. Compared to other plasma extraction reagents and methodologies (such as membrane ultrafiltration, solid phase extraction, and size exclusion chromatography) [3, 19], this strategy might be more practical as a general protocol for clinical plasma LMW protein biomarker discoveries.

### Author Contributions

B.B. conceptualized this study. T.L., S.X., and C.C. performed the experiments mainly. All authors were involved in validation, methodology, investigation, formal analysis, data curation, writing, and approved the final manuscript.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (82172354 and 82472333, to B.B.), the Research Foundation of Jiangsu Provincial Commission of Health and Family Planning (M2021012, to B.B.), the Nanjing Medical Science and Technology Development Foundation (ZKX22013, to B.B.), and the fundings for Novel technology and Clinical Trials from the Affiliated Drum Tower Hospital, Medical School of Nanjing University (XJSFZLX202331 and 2022-LCYJ-MS-02, to B.B.). The authors thank the Nanjing Jiangbei New Area



Biopharmaceutical Public Service Platform (Nanjing, China) for technical support and all the lab members for helpful discussion.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<https://proteomecentral.proteomexchange.org>) via the iProX partner repository [30, 31] with the dataset identifier PXD061482. Other data will be made available on request.

## References

1. F. M. Gribble and F. Reimann, "Metabolic Messengers: Glucagon-Like Peptide 1," *Nature Metabolism* 3 (2021): 142–148, <https://doi.org/10.1038/s42255-020-00327-x>.
2. R. Hellinger, A. Sigurdsson, W. Wu, et al., "Peptidomics," *Nature Reviews Methods Primers* 3 (2023): 25.
3. R. E. Foreman, A. L. George, F. Reimann, F. M. Gribble, and R. G. Kay, "Peptidomics: A Review of Clinical Applications and Methodologies," *Journal of Proteome Research* 20 (2021): 3782–3797, <https://doi.org/10.1021/acs.jproteome.1c00295>.
4. R. Vitorino, M. Choudhury, S. Guedes, et al., "Peptidomics and Proteogenomics: Background, Challenges and Future Needs," *Expert Review of Proteomics* 18 (2021): 643–659, <https://doi.org/10.1080/14789450.2021.1980388>.
5. L. Li, J. Wu, C. J. Lyon, L. Jiang, and T. Y. Hu, "Clinical Peptidomics: Advances in Instrumentation, Analyses, and Applications," *BME Frontiers* 4 (2023): 0019.
6. L. P. Aristoteli, M. P. Molloy, and M. S. Baker, "Evaluation of Endogenous Plasma Peptide Extraction Methods for Mass Spectrometric Biomarker Discovery," *Journal of Proteome Research* 6 (2007): 571–581, <https://doi.org/10.1021/pr0602996>.
7. D. W. Greening, E. A. Kapp, and R. J. Simpson, "The Peptidome Comes of Age: Mass Spectrometry-Based Characterization of the Circulating Cancer Peptidome," *Enzymes* 42 (2017): 27–64.
8. Y. Du, D. Wu, Q. Wu, and Y. Guan, "Quantitative Evaluation of Peptide-Extraction Methods by HPLC–Triple-Quad MS–MS," *Analytical and Bioanalytical Chemistry* 407 (2015): 1595–1605, <https://doi.org/10.1007/s00216-014-8389-0>.
9. B. He, Z. Huang, C. Huang, and E. C. Nice, "Clinical Applications of Plasma Proteomics and Peptidomics: Towards Precision Medicine," *Proteomics Clinical Applications* 16 (2022): 2100097, <https://doi.org/10.1002/prca.202100097>.
10. J. Dufresne, P. Bowden, T. Thavarajah, et al., "The Plasma Peptidome," *Clinical Proteomics* 15 (2018): 39, <https://doi.org/10.1186/s12014-018-9211-3>.
11. L. Das, V. Murthy, and A. K. Varma, "Comprehensive Analysis of Low Molecular Weight Serum Proteome Enrichment for Mass Spectrometric Studies," *ACS Omega* 5 (2020): 28877–28888, <https://doi.org/10.1021/acsomega.0c04568>.
12. X. Jiang, X. Liu, X. Qu, et al., "Integration of Metabolomics and Peptidomics Reveals Distinct Molecular Landscape of Human Diabetic Kidney Disease," *Theranostics* 13 (2023): 3188–3203.
13. G. N. Zhang, Y. J. Xu, and L. Jin, "Peptidomics Analysis of Plasma in Patients With Ankylosing Spondylitis," *Frontiers in Immunology* 14 (2023): 1104351.
14. L. Cassidy, P. T. Kaulich, and A. Tholey, "Depletion of High-Molecular-Mass Proteins for the Identification of Small Proteins and Short Open Reading Frame Encoded Peptides in Cellular Proteomes," *Journal of Proteome Research* 18 (2019): 1725–1734, <https://doi.org/10.1021/acs.jproteome.8b00948>.
15. B. L. Parker, J. G. Burchfield, D. Clayton, et al., "Multiplexed Temporal Quantification of the Exercise-Regulated Plasma Peptidome," *Molecular & Cellular Proteomics: MCP* 16 (2017): 2055–2068, <https://doi.org/10.1074/mcp.RA117.000020>.
16. L. Pont, F. Benavente, J. Barbosa, and V. Sanz-Nebot, "An Update for Human Blood Plasma Pretreatment for Optimized Recovery of Low-Molecular-Mass Peptides Prior to CE–MS and SPE–CE–MS," *Journal of Separation Science* 36 (2013): 3896–3902, <https://doi.org/10.1002/jssc.201300838>.
17. D. J. Harney, A. T. Hutchison, Z. Su, et al., "Small-Protein Enrichment Assay Enables the Rapid, Unbiased Analysis of Over 100 Low Abundance Factors From Human Plasma," *Molecular & Cellular Proteomics: MCP* 18 (2019): 1899–1915.
18. A. W. Purcell, S. H. Ramarathnam, and N. Ternette, "Mass Spectrometry–Based Identification of MHC-Bound Peptides for Immunopeptidomics," *Nature Protocols* 14 (2019): 1687–1707, <https://doi.org/10.1038/s41596-019-0133-y>.
19. S. Mahboob, A. Mohamedali, S. B. Ahn, P. Schulz-Knappe, E. Nice, and M. S. Baker, "Is Isolation of Comprehensive Human Plasma Peptidomes an Achievable Quest?," *Journal of Proteomics* 127 (2015): 300–309, <https://doi.org/10.1016/j.jprot.2015.05.010>.
20. J. Dufresne, A. Florentinus-Mefailoski, P. Bowden, and J. G. Marshall, "A Method for the Extraction of the Endogenous Tryptic Peptides (peptidome) From Human EDTA Plasma," *Analytical Biochemistry* 549 (2018): 188–196, <https://doi.org/10.1016/j.ab.2018.02.025>.
21. M. Tucholska, S. Scozzaro, D. Williams, et al., "Endogenous Peptides From Biophysical and Biochemical Fractionation of Serum Analyzed by Matrix-Assisted Laser Desorption/Ionization and Electrospray Ionization Hybrid Quadrupole Time-of-Flight," *Analytical Biochemistry* 370 (2007): 228–245, <https://doi.org/10.1016/j.ab.2007.07.029>.
22. R. Kay, C. Barton, L. Ratcliffe, et al., "Enrichment of Low Molecular Weight Serum Proteins Using Acetonitrile Precipitation for Mass Spectrometry Based Proteomic Analysis," *Rapid Communications in Mass Spectrometry* 22 (2008): 3255–3260, <https://doi.org/10.1002/rcm.3729>.
23. C. Polson, P. Sarkar, B. Incledon, V. Raguvaran, and R. Grant, "Optimization of Protein Precipitation Based Upon Effectiveness of Protein Removal and Ionization Effect in Liquid Chromatography–Tandem Mass Spectrometry," *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences* 785 (2003): 263–275, [https://doi.org/10.1016/S1570-0232\(02\)00914-5](https://doi.org/10.1016/S1570-0232(02)00914-5).
24. D. N. Potier, J. R. Griffiths, R. D. Unwin, et al., "An Assessment of Peptide Enrichment Methods Employing mTRAQ Quantification Approaches," *Analytical Chemistry* 84 (2012): 5604–5610, <https://doi.org/10.1021/ac300584y>.
25. J. Ma, J. Shi, H. Le, et al., "A Fully Automated Plasma Protein Precipitation Sample Preparation Method for LC–MS/MS Bioanalysis," *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences* 862 (2008): 219–226, <https://doi.org/10.1016/j.jchromb.2007.12.012>.
26. J. Wu, Y. An, H. Pu, et al., "Enrichment of Serum Low-Molecular-Weight Proteins Using C18 Absorbent Under Urea/Dithiothreitol Denatured Environment," *Analytical Biochemistry* 398 (2010): 34–44, <https://doi.org/10.1016/j.ab.2009.10.047>.
27. M. Isokawa, T. Kanamori, T. Funatsu, and M. Tsunoda, "Analytical Methods Involving Separation Techniques for Determination of Low-Molecular-Weight Biothiols in Human Plasma and Blood," *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences* 964 (2014): 103–115, <https://doi.org/10.1016/j.jchromb.2013.12.041>.
28. J. R. De Jesus, R. Da Silva Fernandes, G. De Souza Pessôa, I. M. Raimundo, and M. A. Z. Arruda, "Depleting High-Abundant and Enriching Low-Abundant Proteins in Human Serum: An Evaluation of Sample Preparation Methods Using Magnetic Nanoparticle, Chemical Depletion and Immunoaffinity Techniques," *Talanta* 170 (2017): 199–209, <https://doi.org/10.1016/j.talanta.2017.03.091>.

29. S. E. Warder, L. A. Tucker, T. J. Strelitzer, et al., "Reducing Agent-Mediated Precipitation of High-Abundance Plasma Proteins," *Analytical Biochemistry* 387 (2009): 184–193, <https://doi.org/10.1016/j.ab.2009.01.013>.
30. J. Ma, T. Chen, S. Wu, et al., "iProX: An Integrated Proteome Resource," *Nucleic Acids Research* 47 (2019): D1211–D1217, <https://doi.org/10.1093/nar/gky869>.
31. T. Chen, J. Ma, Y. Liu, et al., "iProX in 2021: Connecting Proteomics Data Sharing With Big Data," *Nucleic Acids Research* 50 (2022): D1522–D1527.
32. K. K. Dey, H. Wang, M. Niu, et al., "Deep Undepleted Human Serum Proteome Profiling Toward Biomarker Discovery for Alzheimer's Disease," *Clinical Proteomics* 16 (2019): 16, <https://doi.org/10.1186/s12014-019-9237-1>.
33. A. Digre and C. Lindskog, "The Human Protein Atlas—Spatial Localization of the Human Proteome in Health and Disease," *Protein Science* 30 (2021): 218–233, <https://doi.org/10.1002/pro.3987>.
34. E. W. Deutsch, G. S. Omenn, Z. Sun, et al., "Advances and Utility of the Human Plasma Proteome," *Journal of Proteome Research* 20 (2021): 5241–5263, <https://doi.org/10.1021/acs.jproteome.1c00657>.
35. D. Shao, L. Huang, Y. Wang, et al., "HBFP: A New Repository for Human Body Fluid Proteome," *Database (Oxford)* 2021 (2021): baab065.
36. M. Uhlen, P. Oksvold, L. Fagerberg, et al., "Towards a Knowledge-Based Human Protein Atlas," *Nature Biotechnology* 28 (2010): 1248–1250.
37. P. J. Thul and C. Lindskog, "The Human Protein Atlas: A Spatial Map of the Human Proteome," *Protein Science* 27 (2018): 233–244, <https://doi.org/10.1002/pro.3307>.
38. H. Keshishian, M. W. Burgess, M. A. Gillette, et al., "Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury," *Molecular & Cellular Proteomics: MCP* 14 (2015): 2375–2393.
39. Y. Bian, F. P. Bayer, Y. C. Chang, et al., "Robust Microflow LC-MS/MS for Proteome Analysis: 38 000 Runs and Counting," *Analytical Chemistry* 93 (2021): 3686–3690.
40. Y. Kawashima, T. Fukutomi, T. Tomonaga, et al., "High-Yield Peptide-Extraction Method for the Discovery of Subnanomolar Biomarkers From Small Serum Samples," *Journal of Proteome Research* 9 (2010): 1694–1705, <https://doi.org/10.1021/pr9008018>.
41. Y. Wang, M. Cai, Y. Song, et al., "PEG Precipitation Followed by Albumin Depletion for Plasma Proteomics Analysis," *Analytical Chemistry* 97 (2025): 12133–12142, <https://doi.org/10.1021/acs.analchem.5c00593>.
42. Z. Miao, K. Ding, S. Jin, L. Dai, C. Dai, and X. Li, "Using Serum Peptidomics to Discover the Diagnostic Marker for Different Stage of Ulcerative Colitis," *Journal of Pharmaceutical and Biomedical Analysis* 193 (2021): 113725, <https://doi.org/10.1016/j.jpba.2020.113725>.
43. Z. Song, C. Liu, Y. Liu, et al., "Long-Term Dysregulation of Plasma Peptidome in Mild and Multiple COVID-19 Recovered Patients Revealed by a Novel Efficient Peptidomics Workflow," *Analytical and Bioanalytical Chemistry* 417 (2025): 733–746, <https://doi.org/10.1007/s00216-024-05684-0>.
44. T. Taguchi, Y. Kodera, K. Oba, et al., "Suprabasin-Derived Bioactive Peptides Identified by Plasma Peptidomics," *Scientific Reports* 11 (2021): 1047.
45. A. Descamps, K. Van Der Borgh, A. De Spiegeleer, E. Wynendaele, and B. De Spiegeleer, "Peptidomics: LC–MS Operational Parameters Do Matter," *Journal of Pharmaceutical and Biomedical Analysis* 229 (2023): 115348, <https://doi.org/10.1016/j.jpba.2023.115348>.
46. J. D. Venable, M.-Q. Dong, J. Wohlschlegel, A. Dillin, and J. R. Yates, "Automated Approach for Quantitative Analysis of Complex Peptide Mixtures From Tandem Mass Spectra," *Nature Methods* 1 (2004): 39–45, <https://doi.org/10.1038/nmeth705>.
47. R. Bruderer, O. M. Bernhardt, T. Gandhi, et al., "Extending the Limits of Quantitative Proteome Profiling With Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues," *Molecular & Cellular Proteomics: MCP* 14 (2015): 1400–1410, <https://doi.org/10.1074/mcp.M114.044305>.
48. B. C. Collins, C. L. Hunter, Y. Liu, et al., "Multi-Laboratory Assessment of Reproducibility, Qualitative and Quantitative Performance of SWATH-Mass Spectrometry," *Nature Communications* 8 (2017): 291, <https://doi.org/10.1038/s41467-017-00249-5>.
49. B. Wen, X. Wang, and B. Zhang, "PepQuery Enables Fast, Accurate, and Convenient Proteomic Validation of Novel Genomic Alterations," *Genome Research* 29 (2019): 485–493, <https://doi.org/10.1101/gr.235028.118>.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.

**Supporting Figure 1:** pmic70083 sup 0001 figuresS1 S9.pdf. **Supporting Table 1:** pmic70083 sup 0001 tableS1.xlsx. **Supporting Table 2:** pmic70083 sup 0001 tableS1.xlsx.