UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MASS SPECTROMETRY AS A BIOANALYTICAL TOOL FOR PROTEOMIC AND IONOMIC STUDIES OF HUMAN MILK: ENRICHMENT OF LOW-ABUNDANCE PROTEINS AND CHARACTERIZATION OF PRETERM LACTATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

AMIT SINGH
Norman, Oklahoma
2026

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/342813

MASS SPECTROMETRY AS A BIOANALYTICAL TOOL FOR PROTEOMIC AND IONOMIC
STUDIES OF HUMAN MILK: ENRICHMENT OF LOW-ABUNDANCE PROTEINS AND CHARACTERIZATION OF PRETERM LACTATION

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY THE COMMITTEE CONSISTING OF

Dr. Christina Bourne, Chair

Dr. Luca Fornelli

Dr. John W. Peters

Dr. Liangliang Huang

© Copyright by AMIT SINGH 2026
All Rights Reserved.

To my mother, whose love, strength, and sacrifices made my education and this journey possible.

And to the mothers and their babies, this work is for you.

ACKNOWLEDGEMENTS

This PhD was not a straight path for me. There were long nights when nothing seemed to go right, moments when I felt completely lost, and a time when I was ready to walk away from it all and settle for a Master's degree. If I am standing at the finish line today, it is because a group of people refused to let me give up on myself. This dissertation belongs to them as much as it does to me.

First, I must thank my family, who are the foundation of everything I am. Dad, thank you for being my silent strength. Even while you were fighting your own difficult health battles, your only concern was my education and my future. You supported me in every way a father could, even when you had every reason to focus only on yourself. I hope this achievement makes you as proud as I am to be your son. Mom, you were my first teacher. I still remember you walking me to school as a child, making sure I was safe before you left. That same love and patience followed me across oceans and guided me through this long journey. I carry your sacrifices on every page of this work. To my sister, Divya, thank you for being the hero of our home. You took on all the responsibilities and cared for our parents so that I could chase my dreams here. I was only able to focus on this work because I knew home was safe in your hands.

Dr. Christina Bourne, I owe you more than words can express. When my previous laboratory moved, and my future in this program became uncertain, you did not hesitate. You welcomed me with a warmth and openness that I had not expected but desperately needed. From the very first conversation, you made me feel that I belonged and that my work mattered. Your door was always open, your guidance always thoughtful, and your presence always reassuring. When I faced one of the most difficult situations of my

doctoral journey, you stood by me and helped me navigate it with grace and strength. Beyond all of that, your scientific mind is genuinely fascinating, and every conversation we had about the work opened up new ways of thinking for me. Thank you for being the kind of mentor that every graduate student deserves, but very few are lucky enough to find.

Dr. Luca Fornelli, I truly would not be here without you. When I was ready to give up and master out, you were the one who stopped me. You challenged me to believe in my potential and helped me see that I was capable of finishing what I started. In our field of mass spectrometry and proteomics, you are the example I strive to follow. Thank you for your guidance in science and for being the reason I stayed the course. You have been a mentor and an inspiration.

Dr. Steven Foster, you have been so much more than a teacher; you have been a guardian. I am so grateful for the time I have spent as an intern with you. Your office talks, your guidance, and your genuine care are truly top-notch. You showed me that being a great scientist goes hand in hand with being a kind person. You are an amazing mentor, and I am lucky to call you a friend.

Dr. John W. Peters, thank you for being someone I could count on when things got truly difficult. As a member of my doctoral committee, you played a role in my PhD that went far beyond our committee meetings. There were moments in this journey when I faced challenges that had nothing to do with the science, moments when I needed someone in a position of authority to simply do the right thing. You were that person. Your quiet support, your fairness, and your willingness to look out for me at critical moments gave me the stability I needed to keep going. I am profoundly grateful for the way you

watched over my progress and ensured that I was treated with dignity throughout this process. I am equally grateful to Dr. Liangliang Huang, who has been a member of my doctoral committee from the very beginning of this PhD. Your consistent support, scientific insight, and genuine belief in my abilities meant more than I ever expressed in the moment. Thank you for being there from day one and for never wavering in your encouragement. I want to acknowledge Dr. Nagib Ahsan, whose mentorship and collaboration shaped nearly every project and publication in this dissertation.

To my friend and labmate, Annelise Huynh, thank you for being such an incredible person to work alongside. You were with me through the daily grind, and our friendship turned the hardest days in the lab into some of my favorite memories. To Abhishek, my friend since my Master's days: thank you for knowing the "me" before I ever started this PhD. Your video calls were a lifeline to my roots, reminding me of home whenever I felt far away.

To my fiancée, Shivani, you have been my light at the end of the tunnel. For months, we lived our lives through screens and time zones, yet you never felt distant. Your laughter and your care turned these grueling years into a journey I wanted to finish for us. Thank you for waiting for me.

Finally, I thank the Oklahoma Mothers Milk Bank, Dr. Hala Chaaban, and the mothers who donated their milk. This work is for the babies in the NICU, whose will to survive inspired me every day.

And to my dog, Finn. You didn't care about my research; you only cared that I was home. Our walks were the only times my mind was truly at peace. I miss you every day, buddy. This is for you, too.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **v**

TABLE OF CONTENTS ..... **viii**

LIST OF TABLES ..... **xi**

Chapter 2 xi

Chapter 3 xi

Supplementary Tables: Chapter 2 xi

Supplementary Tables: Chapter 3 xi

LIST OF FIGURES ..... **xii**

Chapter 1 xii

Chapter 2 xii

Chapter 3 xiii

Supplementary Figures: Chapter 2 xiv

Supplementary Figures: Chapter 3 xv

ABSTRACT ..... **xvii**

LIST OF KEYWORDS ..... **xix**

LIST OF ABBREVIATIONS ..... **xx**

CHAPTER 1: INTRODUCTION TO HUMAN MILK PROTEOMICS AND MASS SPECTROMETRY **1**

1.1 Human Milk as a Biologically Active Fluid 1

1.2 The Analytical Challenge: Dynamic Range in Human Milk Proteomics 3

1.3 Mass Spectrometry as a Bioanalytical Tool 5

1.4 Preterm Birth and the Clinical Importance of Human Milk in the NICU 9

1.5 Scope and Organization of This Dissertation 11

CHAPTER 2: ENRICHMENT OF LOW-ABUNDANCE AND LOW-MOLECULAR-WEIGHT PROTEINS IN HUMAN MILK USING MASS SPECTROMETRY-BASED PROTEOMICS **14**

Author Contributions 14

2.1 Abstract 15

2.2 Introduction 16

2.3 Materials and Methods 21

2.3.1 Collection of Human Milk Samples 21

2.3.2 Protein Depletion Methods 21

2.3.3 Sample Preparation, LC-MS/MS, and Data Analysis 25

2.3.4 Bioinformatics....26
2.4 Results and Discussion....27
2.4.1 Proteome Landscape of Human Milk Identified by Different Depletion Methodologies ..28
2.4.2 Enrichment of Low-Molecular-Weight Proteins (LMWPs)....30
2.4.3 Comparative Study with Various Milk Proteome Datasets....31
2.4.4 Comparison Among Different Abundant Protein Depletion Methods....34
2.5 Limitations of the Study....34
2.6 Conclusions....35
2.7 Associated Content....36
CHAPTER 3: DYNAMIC PROTEOMIC AND IONOMIC LANDSCAPE OF PRETERM HUMAN MILK ACROSS EARLY LACTATION....37
Author Contributions....37
3.1 Abstract....38
3.2 Introduction....39
3.3 Materials and Methods....41
3.3.1 Study Design and Ethical Approval....41
3.3.2 Sample Collection and Storage....41
3.3.3 Study Population....42
3.3.4 Sample Preparation Strategy for Protein Profiling....42
3.3.5 Preparation of Undepleted Milk Samples (Control)....43
3.3.6 Preparation of Perchloric Acid-Treated Milk Samples (Depleted)....43
3.3.7 Milk Protein Extraction and Sample Preparation for Label-Free Proteomics....43
3.3.8 LC-MS/MS Analysis and Data Processing....44
3.3.9 Acid Digestion of Human Milk for ICP-MS Analysis....45
3.3.10 ICP-MS Elemental Analysis....45
3.3.11 Bioinformatics and Data Visualization....46
3.4 Results....48
3.4.1 Study Population....48
3.4.2 Global Proteome Coverage Across Lactation Stages....49
3.4.3 Multivariate and Quantitative Proteomic Profiling Across Lactation Stages....52
3.4.4 Disease Association and Therapeutic Relevance of Stage-Specific Proteins....54
3.4.5 Elemental Variation Between One-Week and Four-Week Milk Samples....57
3.5 Discussion....59

3.6 Conclusions 61
3.7 Data Availability Statement 62
CHAPTER 4: CONCLUSIONS 63
4.1 Overview 63
4.2 Summary of Key Findings 63
4.2.1 Chapter 2: A Methodological Foundation for Human Milk Proteomics 63
4.2.2 Chapter 3: Temporal Dynamics of the Preterm Milk Proteome and Ionome 64
4.3 Integrated Interpretation 66
4.4 Implications for the Field 67
4.4.1 Methodological Implications 67
4.4.2 Clinical Implications 68
4.5 Limitations 69
4.6 Closing Remarks 70
REFERENCES 72
APPENDIX A 84
Permission for reproduction of published material 84
Supplementary figures: Chapter 2 85
Supplementary tables: Chapter 2 88
APPENDIX B 90
Supplementary figures: Chapter 3 90
Supplementary tables: Chapter 3 101

LIST OF TABLES

Chapter 2

Table 2-1. Pros and Cons of Different Protein Depletion Methods. 34

Chapter 3

Table 3-1. Maternal and Infant Clinical Characteristics. 48

Supplementary Tables: Chapter 2

Table S2-1. Literature search results on milk proteome depletion methods. A ScienceDirect literature search was conducted on February 5, 2025, using keywords including acetone, centrifugation, methanol, chloroform, perchloric acid, depletion, and milk proteomics/proteome. Results represent the number of published studies retrieved for each search term and method category. 88

Table S2-2. Complete list of proteins identified by each depletion method using LC-MS/MS analysis. 88

Table S2-3. Complete list of peptides identified by each depletion method using LC-MS/MS analysis. 88

Table S2-4. List of unique proteins identified exclusively by the PerCA depletion method. 108

Table S2-5. Comparative milk proteome analysis against recently published datasets. 89

Supplementary Tables: Chapter 3

Table S3-1. Proteins uniquely upregulated at each lactation stage in PerCA-depleted and undepleted datasets. 102

Table S3-2. Therapeutic compounds and drug targets associated with ELANE. ...

.....103

Table S3-5. Comprehensive proteomic datasets for Chapter 3. 103

LIST OF FIGURES

Chapter 1

Figure 1-1. The dynamic range challenge in human milk proteomics and the effect of abundant protein depletion on LC-MS/MS detection. (A) Iceberg schematic illustrating the human milk proteome, where a small number of highly abundant proteins, including caseins, lactoferrin, albumin, α-lactalbumin, and IgA, account for approximately 90% of the total protein mass and mask detection of the low-abundance proteome during standard LC-MS/MS analysis. The submerged low-abundance proteome contains biologically important proteins involved in immune regulation, signaling, growth, and enzymatic activity. (B) Representative undepleted LC-MS/MS spectrum showing dominance of high-abundance protein signals, which suppresses the detection of low-abundance species. (C) Representative LC-MS/MS spectrum following perchloric acid (PerCA) depletion, demonstrating selective removal of abundant proteins and enhanced coverage of low-abundance proteins. 5

Chapter 2

Figure 2-1. Distribution of protein depletion methods used in milk proteomics. (A) A pie chart illustrates the percentage of methods used for protein depletion in milk. The most employed approach is physical and physicochemical methods (54%), followed by organic solvent-based methods (41%) and acid-based methods (5%). (B) Bar chart comparing the cost per sample for processing using each precipitation method, highlighting the differences in expenses between approaches and providing insight into the cost-effectiveness of each technique in milk proteomics research. 18

Figure 2-2. Schematic representation of the experimental workflow and data analysis for proteomics of human milk samples obtained from a milk bank using six methods. (A) Proteins were analyzed from whole, undepleted milk and after the depletion of major proteins using centrifugation, acetone precipitation, methanol–chloroform precipitation, a commercial kit (CK), and perchloric acid precipitation. (B) Workflow of bottom-up proteomics. Proteins were quantified, digested, cleaned, and analyzed by using LC-MS. (C) Data analysis and bioinformatic studies of the LC-MS output. The figure was created in BioRender (https://BioRender.com/ba37fbc). 20

Figure 2-3. Comparison of protein and peptide identification across different precipitation (depletion) methods. (A) Bar graph showing the total number of proteins and peptides identified using each depletion method, including pooled proteomics data from the three fractions of the commercial kit (CK) and perchloric acid at 2.5, 3, and 3.5% concentrations. (B) Upset plot displaying the overlap and unique proteins identified across all methods, highlighting that perchloric acid (pooled across all concentrations) resulted in the highest number of unique proteins, followed by CK (pooled across all fractions), while undepleted samples identified the fewest unique proteins. (C) KEGG pathway analysis of the PerCA-specific 93 unique protein sets. .....
28

Figure 2-4. Efficacy evaluation of major protein depletion methodologies in human milk samples. (A) A heat map illustrates the efficacy of each depletion method in precipitating the nine most abundant proteins found in human milk. (B) A bar diagram displays the percentage of unique LMWPs identified by using each depletion method. (C) A pie chart shows the number of unique LMWPs identified by using each depletion method.30

Figure 2-5. Comprehensive analysis of unique proteins identified in our study compared to previous proteomic studies on human milk over the past decade. (A) A Venn diagram compares the unique proteins identified in our study with those reported in previous studies, including Herwijnen et al. (2016), Zhu et al. (2018), Dayon et al. (2021), and Viitaharju et al., (2024). (B) KEGG pathway analysis shows the functional roles of the unique protein sets identified in the current study (green circle). (C) Protein–protein interaction map of the unique protein sets identified in this study. (D) Example of method-specific target protein (red box in panel (C)) sets that showed a strong association with other proteins in the protein–protein network and are related to various neonatal developmental processes. Proteins present and absent in a specific method are indicated by red- and black-filled circles, respectively. 32

Chapter 3

Figure 3-1. Experimental workflow for integrated proteomic and ionomic analysis of preterm human milk across early lactation stages. The upper panel illustrates the proteomics workflow including perchloric acid (PerCA) depletion, protein denaturation, reduction and alkylation, enzymatic digestion using trypsin, peptide separation by Dionex UltiMate 3000 UHPLC, and protein identification and quantitation by Q Exactive HF-X mass spectrometry. The lower panel illustrates the ionomics workflow including acid digestion of human milk using HCl and HNO3, dilution with Milli-Q water, and elemental quantification by PerkinElmer NexION 2000 ICP-MS. Both workflows were applied to paired milk samples collected at one week and four weeks postpartum, enabling integrated analysis of the preterm milk proteome and ionome. 47

Figure 3-2. Comparative proteomic coverage, dataset characteristics, and biological interpretation of undepleted and PerCA-depleted human milk proteomes. (A) Total number of proteins identified across all samples (1,043), with comparison between undepleted (878 proteins) and PerCA-depleted datasets (648 proteins). (B) Distribution of shared and unique proteins between workflows, showing 462 proteins common to both datasets, 405 uniquely identified in undepleted samples, and 176 uniquely identified following PerCA depletion. (C) Relationship between protein sequence coverage and peptide spectral matches (PSMs) in the undepleted dataset. (D) Sequence coverage versus PSM distribution in the PerCA-depleted dataset. (E) Comparative analysis of the current dataset with previously published human milk proteomics studies including Dekker et al. (2024), Medic et al. (2025), and Zhao et al. (2024). (F) Integrated pathway enrichment analysis of the 153 proteins uniquely identified in the current study, combining KEGG, Reactome, and disease ontology annotations, revealing enrichment of immune response, epithelial differentiation, host defense, and cell death-related pathways. (G) Protein-protein interaction network of the same 153 unique proteins, illustrating functional clustering into biologically connected modules including muscle/myosin-associated complexes and immunoglobulin-related pathways. 50

Figure 3-3. Multivariate analysis, correlation, and differential protein abundance in undepleted and PerCA-depleted human milk samples. (A) PLS-DA of undepleted samples showing distribution of one-week and four-week groups. (B) PLS-DA of PerCA-depleted samples showing distribution of one-week and four-week groups. (C) Heatmap with hierarchical clustering of protein abundance in undepleted samples. (D) Heatmap with hierarchical clustering of protein abundance in PerCA-depleted samples. (E) Pearson correlation matrix of undepleted samples. (F) Pearson correlation matrix of PerCA-depleted samples. (G) Volcano plot of undepleted samples comparing one-week versus four-week groups, with upregulated proteins (red, n=49), downregulated proteins (green, n=21), and non-significant proteins (gray, n=808). (H) Volcano plot of PerCA-depleted samples comparing one-week versus four-week groups, with upregulated proteins (red, n=29), downregulated proteins (blue, n=17), and non-significant proteins (gray, n=602). (I) Bar plot showing the number of upregulated proteins at one-week and four-week stages in undepleted and PerCA-depleted datasets, with inset donut plots indicating uniquely regulated proteins. (J) Bubble plot of pathway enrichment analysis generated from significantly regulated proteins in the PerCA-depleted dataset, with bubble size representing protein count and color indicating log10 p-value. 53

Figure 3-4. Stage-specific upregulated proteins and disease association analysis in the PerCA-depleted dataset. (A) Bar plot showing average abundance of selected proteins (BPI, ELANE, MINDY1, MUC5B, COX5A, and LY6E) in one-week (red) and four-week (blue) samples. (B) Sankey diagram illustrating disease associations of ELANE, a protein upregulated at four weeks and classified as a TClin target, with connections to cancer, immune/inflammatory, pulmonary, and other disease categories. (C) Sankey diagram illustrating disease associations of MUC5B, a protein upregulated at one week and classified as a TBio target, with connections to cancer, pulmonary, immune/inflammatory, and other disease categories. 56

Figure 3-5. Elemental ionomics analysis of preterm human milk samples across lactation stages by ICP-MS. (A) PCA score plot showing distribution of one-week (red) and four-week (blue) samples based on elemental profiles. (B) Circular heatmap displaying relative abundance patterns of measured elements across samples with color scale representing normalized values. (C through L) Bar plots showing abundance of individual elements in one-week (red) and four-week (blue) samples: (C) Fe, (D) P, (E) Mg, (F) Ca, (G) K, (H) Na, (I) Zn, (J) Cu, (K) Mn, and (L) Hg. Error bars represent variability across samples and p-values are indicated above each comparison. Elemental concentrations are reported in micrograms per milliliter (µg/mL). 58

Supplementary Figures: Chapter 2

Figure S2-1. Current status of milk proteome depletion methods in published literature. A ScienceDirect literature search conducted on February 5, 2025 using keywords including acetone, centrifugation, methanol chloroform, perchloric acid, trichloroacetic acid, and milk proteomics revealed that physical and physicochemical methods are the most commonly used approaches for protein depletion in milk proteomics (54%), followed by organic solvent-based methods (41%) and acid-based methods (5%). Within the physical and physicochemical category, centrifugation accounts for 94% of methods with isoelectric precipitation accounting for the remaining 6%. Among organic solvent-based

approaches, acetone precipitation accounts for 28%, methanol-chloroform for 12%, and other methods primarily ethanol-based for 60%. Acid-based precipitation is predominantly performed using trichloroacetic acid (84%) while perchloric acid is used in 16% of studies. Detailed search results are provided in Table S2-1. 85

Figure S2-2. Protein and peptide identification at individual PerCA concentrations and CK fractions. The number of proteins and peptides identified using each individual condition including the three fractions of the commercial kit (CK-F1, CK-F2, CK-F3) and perchloric acid at three concentrations (PerCA 2.5%, PerCA 3.0%, and PerCA 3.5%). Results are shown separately for each fraction and concentration to illustrate the contribution of each individual condition to overall proteome coverage. Detailed protein and peptide lists are provided in Table S2-2 and Table S2-3, respectively. 86

Figure S2-3. Pairwise milk proteome comparison obtained from different depletion methods. Venn diagrams illustrate the comparison of proteins identified by different methods against the control (undepleted milk). 87

Supplementary Figures: Chapter 3

Figure S3-1. Principal component analysis (PCA) and scatter matrix analysis of protein abundance in undepleted and PerCA-depleted datasets. (A) PCA score plot of undepleted samples showing clustering of one-week and four-week groups based on global protein abundance profiles. (B) PCA score plot of PerCA-depleted samples showing corresponding lactation stage separation. (C and D) Scatter matrix plots illustrating pairwise relationships in protein abundance between individual undepleted samples at one week and four weeks. (E and F) Scatter matrix plots illustrating pairwise relationships in protein abundance between individual PerCA-depleted samples at one week and four weeks. Consistent clustering across both multivariate approaches confirms that lactation stage is the primary source of proteome variation in both sample preparation workflows. 90

Figure S3-2. Overview of proteins uniquely identified in PerCA-depleted samples. (A) Venn diagram showing the 176 proteins exclusively detected in PerCA-depleted samples relative to the undepleted dataset. (B) Gene Ontology enrichment analysis of the 176 PerCA-unique proteins showing enriched biological process, molecular function, and cellular component terms. (C) Functional classification of PerCA-unique proteins by biological category, highlighting representation across immune defense, metabolic regulation, and structural functions. 92

Figure S3-3. Vitamin D receptor (VDR) signaling pathway enrichment of PerCA-unique proteins. Network visualization showing the integration of PerCA-unique proteins within the curated VDR signaling pathway, with detected proteins highlighted and pathway connectivity illustrated through protein-protein interaction nodes. 94

Figure S3-4. Apoptosis pathway enrichment of PerCA-unique proteins. Network visualization highlighting detected PerCA-unique proteins within the curated apoptosis signaling pathway, showing their positions within programmed cell death cascades including both intrinsic and extrinsic apoptotic routes. 95

Figure S3-5. Gastric cancer pathway network of PerCA-unique proteins. Network diagram showing the integration of PerCA-unique proteins within the curated gastric cancer pathway, illustrating their connectivity to oncogenic and tumor-suppressive nodes within this disease-associated network. 96

Figure S3-6. Breast cancer protein-protein interaction network of PerCA-unique proteins. Network visualization highlighting PerCA-unique proteins within curated breast cancer interaction modules, showing functional clustering and connectivity among proteins with documented roles in breast cancer pathways. 97

Figure S3-7. Lymphocyte cell-cell adhesion pathway network of PerCA-unique proteins. Network diagram illustrating immune-related PerCA-unique proteins within the lymphocyte cell-cell adhesion pathway, highlighting their roles in immune cell interactions and adhesion-mediated signaling relevant to mucosal immunity in the neonatal gut. 98

Figure S3-8. Overlap and filtered heatmap analysis of differentially regulated proteins. (A) Venn diagram showing the overlap of upregulated and downregulated proteins between undepleted and PerCA-depleted datasets at one week versus four weeks, illustrating the largely complementary nature of the two preparation workflows in capturing regulated proteins. (B) Heatmap of differentially regulated proteins in the undepleted dataset after applying a 50% detection filter. (C) Corresponding filtered heatmap for the PerCA-depleted dataset showing stage-specific abundance patterns after the same detection threshold was applied. 99

Figure S3-9. Extended disease association Sankey diagrams for stage-specific proteins identified in the PerCA-depleted dataset. Sankey diagrams illustrating the disease category mapping of proteins uniquely upregulated at one week and four weeks postpartum, showing distribution across cancer, immune/inflammatory, pulmonary, neurological, and other disease categories derived from NIH Pharos and Target Central Resource Database classifications. 100

ABSTRACT

Human milk is biologically extraordinary, carrying hundreds of immune-active proteins, enzymes, and minerals that collectively protect and shape the developing infant. For a researcher, however, this complexity presents a significant analytical challenge. A handful of highly abundant proteins casein, α-lactalbumin, lactoferrin, and secretory IgA dominate the fluid at concentrations spanning six to eight orders of magnitude above the rarer proteins that often carry the most important biological signals. This dissertation applies high-resolution mass spectrometry not simply as an analytical instrument, but as a means to look past these dominant proteins and uncover the hidden molecular landscape of human milk.

The first study addresses a methodological necessity: finding a reliable, cost-effective way to see what is hidden. We conducted the first systematic head-to-head evaluation of five protein depletion strategies: centrifugation, acetone precipitation, methanol-chloroform precipitation, a commercial depletion kit (Minute ML-044), and perchloric acid (PerCA) precipitation, against five performance dimensions: total protein identification, high-abundance protein depletion efficiency, low-molecular-weight protein enrichment, unique protein identification, and cost per sample. PerCA emerged as the most efficient and cost-effective strategy at approximately one cent per sample, over 300 times less expensive than the commercial kit uniquely enriching low-molecular-weight proteins below 26 kDa (37.5% of unique identifications), identifying 93 proteins not detected by any other method, and uncovering 115 proteins not previously reported in any of four publicly available human milk proteome databases. These findings establish a validated, accessible framework that allows researchers to probe the milk proteome at

a depth not previously achievable with a single depletion approach, and are published in ACS Omega 2025.

Building directly on these methodological findings, the second study applies this validated framework to a clinical population that needs it most: mothers of very preterm infants (gestational age 22 to 31 weeks). By analyzing paired milk samples from eight mothers at Week 1 (postpartum days 7 to 10) and Week 4 (postpartum days 21 to 28) using two complementary preparation strategies, undepleted and PerCA-depleted, alongside elemental profiling of ten clinically relevant minerals by ICP-MS (PerkinElmer NexION 2000), we produced the first integrated proteomic and ionomic portrait of preterm milk maturation across early lactation. Across 1,043 total proteins identified, stage-specific proteomic shifts were observed in proteins involved in infection defense (ELANE and BPI, elevated at Week 4) and gut barrier formation (MUC5B, elevated at Week 1), alongside meaningful changes in clinically relevant minerals including iron, zinc, calcium, and phosphorus, findings with direct implications for NICU supplementation protocols and donor milk fortification strategies.

Together, these studies advance human milk science in two important and connected ways. Methodologically, they provide a practical, affordable, and validated tool for comprehensive milk proteomics accessible to researchers in any laboratory setting. Biologically, they offer the first molecular-level insight into how the protein and mineral composition of preterm milk dynamically evolves during the most critical weeks of neonatal life, a window during which nutritional decisions directly shape long-term outcomes. Ultimately, this work is a step toward understanding human milk not as a static

nutritional fluid but as a dynamic biological system, and toward improving clinical nutrition and care for the most vulnerable infants in the NICU.

LIST OF KEYWORDS

- Human milk proteomics
- Preterm human milk
- Mass spectrometry
- LC-MS/MS
- ICP-MS
- Protein depletion
- Perchloric acid precipitation
- Low-abundance proteins
- Label-free quantification
- Lactation dynamics
- Preterm birth
- Necrotizing enterocolitis
- Neonatal intensive care unit
- Bioactive proteins
- Mineral composition
- Bottom-up proteomics

LIST OF ABBREVIATIONS

Abbreviation	Full form
CK	Commercial Kit
DDA	Data-Dependent Acquisition
FDR	False Discovery Rate
GO	Gene Ontology
HAP	High-Abundance Protein
HCD	Higher-Energy Collisional Dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid
HM	Human Milk
HMBANA	Human Milk Banking Association of North America
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
IRB	Institutional Review Board
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAP	Low-Abundance Protein
LC-MS/MS	Liquid Chromatography Tandem Mass Spectrometry
LMWP	Low-Molecular-Weight Protein
NEC	Necrotizing Enterocolitis
NICU	Neonatal Intensive Care Unit
PCA	Principal Component Analysis
PerCA	Perchloric Acid
PLS-DA	Partial Least Squares Discriminant Analysis
PPI	Protein-Protein Interaction
PSM	Peptide Spectral Match
PTM	Post-Translational Modification
SIgA	Secretory Immunoglobulin A
TCRD	Target Central Resource Database
TMT	Tandem Mass Tags
UHPLC	Ultra-High Performance Liquid Chromatography
XIC	Extracted Ion Chromatogram

CHAPTER 1: INTRODUCTION TO HUMAN MILK PROTEOMICS AND MASS SPECTROMETRY

1.1 Human Milk as a Biologically Active Fluid

Human milk is far more than a source of calories; it is a sophisticated, living biological system. Shaped by millions of years of evolution, it delivers a level of molecular complexity that no manufactured substitute has been able to replicate. The World Health Organization and the American Academy of Pediatrics both advocate for exclusive breastfeeding for the first six months of life, a position rooted in decades of evidence showing that human milk protects infants against respiratory infections, gastrointestinal illness, and long-term metabolic diseases, including obesity and type 2 diabetes (1,2). These benefits are not incidental; they reflect the extraordinary diversity of biologically active molecules that human milk delivers to the infant gut during the precise developmental window when the immune system, microbiome, and organ systems are being established for the first time.

Beyond the fats, lactose, and macronutrients that have been quantified and studied for decades, human milk carries hundreds of bioactive proteins, enzymes, hormones, immune factors, and growth regulators that function as a molecular shield for the developing infant. The most abundant of these have been characterized in considerable depth. α-Lactalbumin is not only a major whey protein but also a key regulator of lactose synthase activity in the mammary gland and generates bioactive peptides with antimicrobial properties upon digestion in the infant gut (3). Lactoferrin is an iron-binding glycoprotein present at particularly high concentrations in colostrum (~7 g/L) that declines as lactation matures; it exerts broad-spectrum antimicrobial activity against bacteria,

viruses, and fungi through iron sequestration, disruption of microbial membranes, and inhibition of pathogen adhesion to host epithelium (4). Secretory immunoglobulin A (SIgA) is the most abundant antibody in human milk and functions as a critical mucosal immune agent coating the surfaces of the infant intestine to neutralize pathogens, prevent bacterial adhesion, and shape the composition of the developing gut microbiota (5). The casein proteins: β-casein, κ-casein, and αs1-casein, deliver calcium to the infant and generate biologically active peptides upon proteolytic digestion that have opioid-like, immunomodulatory, and antimicrobial properties (3).

These abundant proteins are important, biologically meaningful, and extensively studied. But they are not the whole story. Beneath them lies a vast and largely unexplored layer of low-abundance proteins whose functions span early brain development, antiviral defense, immune regulation, gut maturation, and maternal-infant signaling (6). These proteins exist at concentrations that are orders of magnitude lower than those of the dominant species and have received limited systematic attention, largely because studying them is technically difficult. That difficulty and what can be done about it are at the heart of this dissertation.

In addition to proteins, human milk contains an array of minerals and trace elements that are essential for infant growth, skeletal development, and organ function. Calcium and phosphorus drive bone mineral accretion; iron and zinc are required for neurological development, enzymatic activity, and immune competence; and copper, manganese, and selenium contribute to antioxidant defense and metabolic regulation (3). The bioavailability of these minerals in human milk, mediated in part by binding proteins such as lactoferrin for iron, is generally higher than in formula, contributing to the clinical

advantages of breastfeeding. Despite their recognized importance, how mineral concentrations change systematically across the stages of lactation, particularly in the context of preterm birth, has not been comprehensively characterized using modern high-sensitivity analytical methods. This gap is directly addressed by the second study in this dissertation.

1.2 The Analytical Challenge: Dynamic Range in Human Milk Proteomics

The problem of dynamic range is not unique to human milk; it is a defining hurdle in the proteomics of any biological fluid. In human blood plasma, albumin and the immunoglobulins collectively account for more than 90% of the total protein mass, while biologically informative proteins are present at concentrations spanning more than 10 orders of magnitude below them (7). In human milk, casein and whey proteins play structurally similar roles. During LC-MS/MS analysis, the mass spectrometer operates under data-dependent acquisition, meaning it selects the most abundant peptide ions in each survey scan for fragmentation and identification. When those abundant ions come overwhelmingly from a handful of dominant proteins, the instrument spends the majority of its measurement time characterizing species the researcher already knows well, and the rarer, potentially more informative proteins are simply not sampled. The result is that the hidden layer of the milk proteome, the proteins most likely to reveal new biology, remains largely invisible without deliberate intervention.

To overcome this, researchers use protein depletion strategies to remove or reduce the abundant background before analysis, essentially clearing the noise so that the quiet biological signals can be heard. Several strategies have been developed and adapted from plasma and serum proteomics for application to milk. High-speed

centrifugation exploits differences in density and sedimentation to remove casein micelles, the large, abundant protein aggregates that form the structural core of casein in milk. Organic solvent-based precipitation with acetone, methanol, chloroform, or combinations thereof induces protein aggregation by disrupting the solvation shells around proteins, thereby driving them out of solution. Acid precipitation using perchloric acid (PerCA) or trichloroacetic acid works through a different mechanism: the acid disrupts the ionic interactions and hydrogen bonds that maintain protein tertiary structure, causing large, folded abundant proteins to precipitate while smaller or less structured proteins remain in solution (8). Commercial affinity-based kits use selective binding surfaces or precipitation chemistries specifically designed to capture the most abundant milk proteins while allowing the remaining proteome to pass through for downstream analysis (9).

Each of these approaches carries trade-offs. Methods that most aggressively remove abundant proteins may co-deplete lower-abundance species, reducing total identification counts. Methods that preserve the greatest diversity of proteins may not sufficiently clear the dominant background to meaningfully improve the detection of the rare proteins of interest. The question of which approach works best for human milk specifically, and for what analytical objective, had never been answered through a systematic, controlled comparison prior to the work described in this dissertation. Despite the growing body of literature on human milk proteomics, most published studies employed a single depletion method without evaluating alternative methods side by side (10). Chapter 2 fills that gap directly.

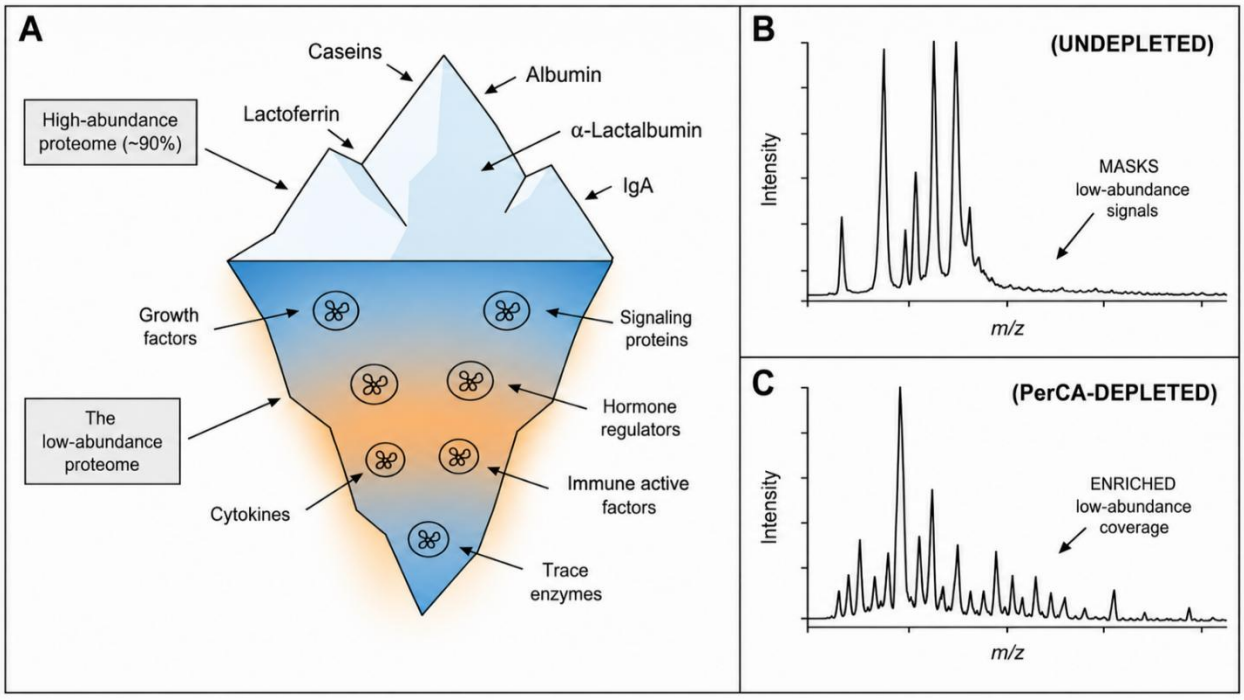


Figure 1-1. The dynamic range challenge in human milk proteomics and the effect of abundant protein depletion on LC-MS/MS detection. (A) Iceberg schematic illustrating the human milk proteome, where a small number of highly abundant proteins, including caseins, lactoferrin, albumin, α-lactalbumin, and IgA, account for approximately 90% of the total protein mass and mask detection of the low-abundance proteome during standard LC-MS/MS analysis. The submerged low-abundance proteome contains biologically important proteins involved in immune regulation, signaling, growth, and enzymatic activity. (B) Representative undepleted LC-MS/MS spectrum showing dominance of high-abundance protein signals, which suppresses the detection of low-abundance species. (C) Representative LC-MS/MS spectrum following perchloric acid (PerCA) depletion, demonstrating selective removal of abundant proteins and enhanced coverage of low-abundance proteins.

1.3 Mass Spectrometry as a Bioanalytical Tool

Mass spectrometry has become the gold standard for large-scale protein identification and quantification in complex biological matrices. Its combination of sensitivity, speed, molecular coverage, and the ability to identify and quantify thousands of proteins simultaneously in a single analytical run makes it uniquely suited to the study of biological fluids as complex as human milk (11). The bottom-up proteomics workflow used throughout this dissertation begins with protein extraction from the biological

sample, followed by reduction of disulfide bonds with a reducing agent, alkylation of free thiols to prevent disulfide re-formation, and enzymatic digestion of proteins into peptides using trypsin, a serine protease that cleaves specifically after lysine and arginine residues, producing peptides of a predictable and MS-amenable size range. The resulting peptide mixture is then separated by reversed-phase liquid chromatography, in which peptides are retained on a hydrophobic stationary phase and eluted by a gradient of increasing organic solvent concentration. Separated peptides are introduced into the mass spectrometer via electrospray ionization, which transfers peptide ions from solution into the gas phase. In both studies presented in this dissertation, peptide separation was performed using an EasySpray analytical column (3 µm, 75 µm x 15 cm, Thermo Fisher Scientific), which integrates the electrospray emitter directly into the column outlet, eliminating post-column dead volume and minimizing peak broadening and analyte loss during transfer of peptides to the mass spectrometer, an important consideration for the detection of low-abundance peptides where signal losses at any stage of the analytical path can reduce sensitivity below the detection threshold.

Inside the instrument, two levels of mass analysis occur. At the MS1 level, the instrument performs a full survey scan, recording the mass-to-charge ratio (m/z) and intensity of all ionized peptide precursors eluting from the LC column at that moment. In data-dependent acquisition mode, the instrument then selects the most abundant precursor ions, typically the top ten to twenty, and isolates them one at a time for fragmentation by higher-energy collisional dissociation. The resulting fragment ion spectra, recorded at the MS2 level, contain sequence-specific b- and y-ions that collectively encode the amino acid sequence of the peptide. These spectra are then

computationally searched against a protein sequence database; in this work, the UniProt reviewed human proteome to identify the peptide sequence and, by inference, the parent protein from which it was derived (11).

The Q Exactive HF-X hybrid mass spectrometer used in both studies presented in this dissertation represents a current benchmark for bottom-up proteomics. It combines a quadrupole mass filter for high-selectivity precursor isolation, a higher-energy collisional dissociation cell for efficient peptide fragmentation, and an Orbitrap mass analyzer for measurement of fragment ions with resolving power exceeding 60,000 at m/z 200 and mass accuracy below 5 parts per million. These characteristics allow confident identification and quantification of thousands of proteins per LC-MS/MS run, even in complex matrices with high dynamic range. Label-free quantification, the approach used throughout this dissertation, estimates relative protein abundance from the area under the extracted ion chromatogram of selected peptides, providing a sensitive, reproducible, and cost-effective basis for comparing protein levels across samples without the need for chemical isotope labeling strategies such as TMT or iTRAQ (12). Label-free quantification was selected over chemical labeling approaches such as tandem mass tags (TMT) or isobaric tags for relative and absolute quantitation (iTRAQ) for several reasons specific to the design of both studies. First, chemical labeling reagents represent a significant per-sample cost, particularly for experiments involving large numbers of samples across multiple depletion conditions and time points. Second, TMT-based workflows are limited in the number of samples that can be multiplexed in a single analytical run, typically 16 to 18 channels, which constrains experimental design when many conditions must be compared simultaneously. Third, co-isolation interference, a well-recognized artifact of

isobaric labeling, artificially compresses measured fold-changes between conditions and reduces sensitivity for detecting true differences in low-abundance proteins, precisely the proteins that both studies in this dissertation were designed to enrich and characterize. Fourth, label-free quantification via extracted ion chromatogram area integration places no upper limit on the number of samples compared, requires no additional preparation beyond standard digestion and cleanup, and is free from co-isolation artifacts, making it the most appropriate choice for discovery-phase proteomics studies of this design (12).

Inductively coupled plasma mass spectrometry (ICP-MS) is the complementary analytical platform used in the second study to profile the mineral composition of the same preterm milk samples analyzed by LC-MS/MS. Unlike LC-MS/MS, which measures molecular ions derived from peptides, ICP-MS operates by introducing liquid samples into a high-temperature argon plasma at approximately 6,000 to 8,000 K. At this temperature, every element in the sample is simultaneously atomized and ionized. The resulting atomic ions are separated and detected according to their mass-to-charge ratio, enabling simultaneous quantification of dozens of elements in a single analytical run. The PerkinElmer NexION 2000 ICP-MS instrument used in this work provides a dynamic range spanning six to eight orders of magnitude with detection limits in the parts-per-trillion range, making it capable of accurately quantifying both major minerals like calcium and phosphorus and trace elements like zinc and copper in the small sample volumes available from individual preterm milk collections (13,14). By combining label-free LC-MS/MS proteomics and ICP-MS ionomics on the same samples, we can build a more complete molecular portrait of preterm milk than either platform alone could provide, capturing both protein signals and the mineral context in which those proteins operate.

1.4 Preterm Birth and the Clinical Importance of Human Milk in the NICU

The most critical application of the technology described in this dissertation is in the neonatal intensive care unit. Preterm birth, defined as delivery before 37 completed weeks of gestational age, affects approximately one in ten pregnancies globally and is the leading cause of neonatal mortality worldwide (15). The clinical spectrum of prematurity is broad. Infants born at 34 to 36 weeks, classified as late preterm, generally survive with relatively limited intervention. Infants born at 28 to 32 weeks, the very preterm group most relevant to the clinical work in Chapter 3, require prolonged NICU admission and face significant risks of morbidity across multiple organ systems. Infants born below 28 weeks, at the extreme of viability, face the greatest risks of mortality and long-term disability, including cerebral palsy, cognitive impairment, chronic lung disease, and severe retinal disease (16).

For these infants, the mother's own human milk is the single most powerful evidence-based clinical intervention available from a nutritional standpoint. Studies consistently demonstrate that preterm infants fed human milk have significantly lower rates of necrotizing enterocolitis (NEC), the most feared and devastating gastrointestinal complication of prematurity, with mortality rates of 20 to 30% in affected infants (17), and higher human milk exposure is associated with a significantly lower risk of NEC or death in extremely low birth weight infants (18). Every 10% increase in the proportion of total enteral intake provided as human milk is associated with a measurable reduction in the risk of NEC or death (18). Human milk feeding is also associated with lower rates of late-onset sepsis, faster achievement of full enteral feeding, better tolerance of enteral nutrition, and significantly improved neurodevelopmental outcomes at 18 to 24 months corrected gestational age (16). When a mother's own milk is unavailable or insufficient, a

common occurrence in mothers who deliver extremely preterm, pasteurized donor human milk from an accredited milk bank is strongly preferred over preterm formula, as it retains many of the protective bioactive proteins and immunological components even after Holder pasteurization at 62.5°C for 30 minutes (16).

What makes this clinically complex is that preterm milk is not simply a dilute or deficient version of term milk; it is compositionally distinct in ways that appear to reflect a biological adaptation to the specific vulnerabilities of the preterm infant. In the first days and weeks after preterm delivery, milk contains higher concentrations of total protein, secretory IgA, lactoferrin, and multiple cytokines, including interleukin-6, interleukin-8, and tumor necrosis factor-α, as well as elevated levels of growth factors, including epidermal growth factor and insulin-like growth factor-1 (3). Endogenous proteolytic activity is also greater in preterm milk, generating elevated concentrations of bioactive peptide fragments. These advantages are not static; however, both preterm and term milk undergo rapid compositional maturation in the early weeks of lactation, with protein content declining substantially from colostrum to transitional and then mature milk (19). The trajectory of this maturation at the level of the full proteome, which proteins change, in which direction, and at what rate, has not been comprehensively characterized for preterm milk specifically. A recent paired study of 40 mothers found significant proteomic differences between milk collected in early and later lactation, but noted that most prior investigations had focused on selected proteins rather than the broader proteome, leaving the full picture incomplete (20). That gap is directly addressed in Chapter 3 of this dissertation.

The mineral composition of preterm milk adds yet another layer of clinical urgency. Preterm infants have high requirements for calcium and phosphorus to support bone mineral accretion, and metabolic bone disease, resulting from inadequate mineral delivery during a period of rapid skeletal growth, is a recognized complication of very preterm birth. Iron deficiency is also prevalent, given the limited body iron stores accumulated during a shortened gestation and the high demand for iron during rapid brain development in the early weeks of extrauterine life. Zinc and copper are required for immune function and connective tissue formation. ICP-MS analysis of preterm milk has previously shown that its mineral content changes over the course of lactation, with zinc and copper declining when compared at equivalent postmenstrual ages, and that preterm milk may contain lower concentrations of certain key minerals than term milk at corresponding lactation stages (13). These findings have direct clinical implications for designing and timing supplementation protocols in the NICU, yet quantitative, element-specific characterization of preterm milk mineral dynamics during the early weeks of lactation, the window most relevant to clinical decision-making, remains limited.

1.5 Scope and Organization of This Dissertation

The research described in this dissertation was conducted at the University of Oklahoma, primarily within the Mass Spectrometry, Proteomics and Metabolomics Core Facility. The preterm milk samples analyzed in the second study were collected and provided through the clinical biorepository maintained by Dr. Hala Chaaban's laboratory at the University of Oklahoma Health Campus. Mass spectrometry, in its LC-MS/MS form for protein analysis and its ICP-MS form for elemental analysis, serves as the central

analytical platform throughout, and both studies share the same institutional setting, instrumentation, and analytical philosophy.

The research follows a logical progression from methodological foundation to clinical application, with each chapter building directly on the one that precedes it.

Chapter 2 presents the first systematic, head-to-head comparison of five abundant protein depletion strategies for human milk proteomics. By evaluating centrifugation, acetone precipitation, methanol-chloroform precipitation, a commercial depletion kit, and perchloric acid (PerCA) precipitation across five performance dimensions, total protein identification yield, high-abundance protein depletion efficiency, low-molecular-weight protein enrichment, unique protein identification, and cost per sample, this chapter validates PerCA precipitation as the most cost-effective and efficient approach for accessing the low-abundance and low-molecular-weight proteome of human milk. This work was published in ACS Omega in 2025 (10).

Chapter 3 builds directly on that methodological foundation and applies the validated approach to a study with direct clinical relevance. By integrating label-free LC-MS/MS proteomics with ICP-MS ionomics on paired milk samples from mothers of very preterm infants collected at one week and four weeks postpartum, this chapter characterizes how the proteome and ionome of preterm milk change across the early weeks of lactation, producing the first integrated molecular portrait of this clinically important fluid over time.

Chapter 4 synthesizes the findings from both studies, discusses their implications for neonatal nutrition research and clinical practice, acknowledges the limitations of the work, and outlines the future of personalized human milk science, a field in which

understanding precisely what is in milk, how it changes, and what that means for the infant represents one of the most important unanswered questions in neonatal medicine.

The work begins in Chapter 2 with a question that had to be answered before any biological study of human milk could be designed with confidence: among the available strategies for removing abundant proteins from human milk, which one works best, and for what purpose?

CHAPTER 2: ENRICHMENT OF LOW-ABUNDANCE AND LOW-MOLECULAR-WEIGHT PROTEINS IN HUMAN MILK USING MASS SPECTROMETRY-BASED PROTEOMICS

Reprinted (adapted) with permission from Singh A, Peng Z, Shah SM, Salkini M, Mallah AF, Ali AF, Kurella SR, Huynh AN, Saleemi SM, Khan A, Mesiya S, Ali A, Samour A, Chaaban H, Yang Z, Ahsan N. Enrichment of Low-Abundance and Low-Molecular-Weight Proteins in Human Milk Using Mass Spectrometry-Based Proteomics. ACS Omega. 2025;10(37):42507–42517. doi: 10.1021/acsomega.5c03776. Copyright 2025 American Chemical Society.

Permission was obtained through the ACS RightsLink service. As the first author and a co-author of this work, reproduction rights in a doctoral dissertation are granted automatically by the American Chemical Society under its standard author reuse policy.

License: https://creativecommons.org/licenses/by-nc-nd/4.0/

RightsLink record: https://marketplace.copyright.com/rs-ui-web/mp/search?type=all&q=10.1021%2Facsomega.5c03776

Author Contributions

The following contributions reflect the roles of each author in the work described in this chapter, which has been reproduced from the following open-access publication. **Amit Singh** designed and executed all aspects of sample preparation including protein extraction, reduction, alkylation, trypsin/Lys-C in-solution digestion, C18 Sep-Pak desalting, and SpeedVac drying across all six depletion conditions and 30 total LC-MS/MS runs. Performed all proteomics data acquisition on the Q Exactive HF-X mass spectrometer coupled to the Dionex UltiMate 3000 UHPLC system. Conducted all bioinformatics analyses including database searches in Proteome Discoverer 2.4 using Sequest HT, KEGG pathway enrichment analysis using ShinyGO 0.81, protein-protein interaction network analysis using STRING, Venn diagram construction using InteractiVenn, and upset plot generation using SRplot. Generated all figures and tables presented in this chapter. Wrote the primary draft of the manuscript and led all revisions through peer review and publication. **Zongkai Peng** assisted with sample collection and processing and contributed to manuscript writing, review, and editing. **Salmaan M. Shah** assisted with sample collection and processing and contributed to manuscript review and editing.

The following undergraduate research assistants, **Malek Salkini, Abdullah F. Mallah, Aleena F. Ali, Simran R. Kurella, Annelise N. Huynh, Saaim M. Saleemi, Ayan Khan, Sohaib Mesiya, Akbar Ali, and Amir Samour**, assisted with sample collection and processing, specifically the execution of protein depletion protocols, under the direct supervision of Amit Singh. All depletion protocols were developed and optimized by Amit Singh prior to undergraduate involvement. All experimental contributions from undergraduate assistants were verified and quality-controlled by Amit Singh before downstream processing and analysis.

Hala Chaaban contributed to study conceptualization, provided clinical expertise on human milk biology and neonatal nutrition, contributed to manuscript review and editing, and secured funding for the research. **Zhibo Yang** contributed to study conceptualization, supervised the analytical chemistry and mass spectrometry methodology, provided laboratory resources and instrumentation access, contributed to manuscript review and editing, and secured funding for the research. **Nagib Ahsan** contributed to study conceptualization, provided oversight of proteomics analysis and bioinformatics, co-wrote the primary manuscript draft, and contributed to manuscript writing, review, and editing.

2.1 Abstract

Human milk is a uniquely complex biological fluid, rich in bioactive components such as proteins, antibodies, enzymes, and hormones that are critical for immunity, development, and disease prevention. However, studying these trace proteins requires specialized depletion methods to counteract the dominance of abundant proteins. In this study, we systematically evaluated five different protein depletion methods, centrifugation, organic solvent-based approaches, acid precipitation, and a commercial kit (CK), to assess their efficacy in enriching low-abundance proteins (LAPs) for LC-MS/MS analysis. The comparative analysis of different depletion methods selectively enriched the sequencing depth of LAPs, highlighting the necessity of method selection in milk proteomics research. Our results further revealed that perchloric acid (PerCA) precipitation was the most effective method for identifying unique low-molecular-weight proteins (LMWPs), and optimization of the abundant protein depletion strategy could greatly increase (>10%) the extraction of LMWPs from milk samples. Additionally, several new proteins were identified in our investigation compared with publicly available milk proteome data. Thus, expanding the list of human milk proteomes enriched the dataset and offered valuable insights into the complex biological functions of human milk and its impact on neonatal research.

2.2 Introduction

As discussed in Chapter 1 (Section 1.2), the dynamic range of the human milk proteome poses a fundamental analytical challenge that requires targeted depletion strategies prior to LC-MS/MS analysis. Human milk, often considered the gold standard for infant nutrition, not only supplies essential macronutrients but also serves numerous vital roles in the overall well-being of the infant (21–23). In addition to its unique biological fluid composition, its bioactive components, including many low-abundance proteins, are linked to lowering the risks of allergies, diabetes, obesity, and immune-related disorders, as well as supporting developmental and regulatory functions, emphasizing its extensive benefits beyond immediate nutrition (24–27). On the other hand, various processing methods like pasteurization, ultrahigh-temperature treatment, and homogenization can lead to the degradation or loss of these beneficial bioactive compounds, which significantly diminishes their potential health benefits (28,29). Furthermore, the solubility and physical state of proteins in milk may be changed by pasteurization-induced structural changes like protein denaturation and aggregation (30). These modifications may affect the protein interaction with various protein extraction buffers containing different solvents, acids, or binding surfaces, which could have an effect on milk proteome profiling.

The protein composition of human milk is highly heterogeneous, reflecting a diverse range of bioactivities and physical properties (31,32). Although the total protein content is relatively low, at approximately 1%, it is dominated by whey proteins (~70%), including α-lactalbumin, lactoferrin, and secretory IgA (SIgA), and casein proteins such as β-casein, lysozyme, serum albumin, osteopontin, lactadherin, and κ-casein (3,33). While these abundant proteins are essential for infant nutrition and immunity, their

dominance often masks the detection of many low-abundance proteins, which may hold biological significance (34–36).

Therefore, profiling the milk proteome is considered to be one of the fundamental steps in neonatal research. Numerous milk proteomics studies have been previously performed using different techniques, including electrophoresis, Western blot, chromatography, and mass spectrometry (MS). Due to its high sensitivity, broad molecular coverage, and flexibility to couple with separation techniques, MS has become the major technique for proteomics studies, including those of mammalian milks. In addition to traditional bottom-up proteomics techniques, top-down methods have been used to sequence intact proteins, allowing for studies of sequence variants and post-translational modifications (PTMs) (37,38).

Proteomics studies of farm animal milk have been previously carried out, demonstrating the significance and challenges in analyzing bioactive peptides and low-abundance proteins (39). Various protein depletion and extraction techniques have been developed to selectively remove high-abundance proteins from milk, such as whey and casein, making low-abundance proteins more accessible for analysis using advanced techniques such as liquid chromatography-mass spectrometry (LC-MS). A survey of earlier studies on milk proteome analysis revealed that various major protein depletion and precipitation methodologies have been successfully used to enhance the identification of low-abundance proteins (Figure 2-1A and Table S2-1). These precipitation methods can predominantly be classified into three categories: physical/physicochemical (54%), organic-solvent-based (41%), and acid-based (5%) (Figure 2-1A). Within the physical/physicochemical category, centrifugation accounts for

94% of methods, with isoelectric precipitation making up the remaining 6%. In the organic solvent-based group, acetone-based methods constitute 28%, methanol/chloroform-based approaches 12%, and other methods, primarily ethanol-based, account for 60%. Among acid-based techniques, 84% utilize trichloroacetic acid, while 16% employ PerCA (Figure S2-1 and Table S2-1).

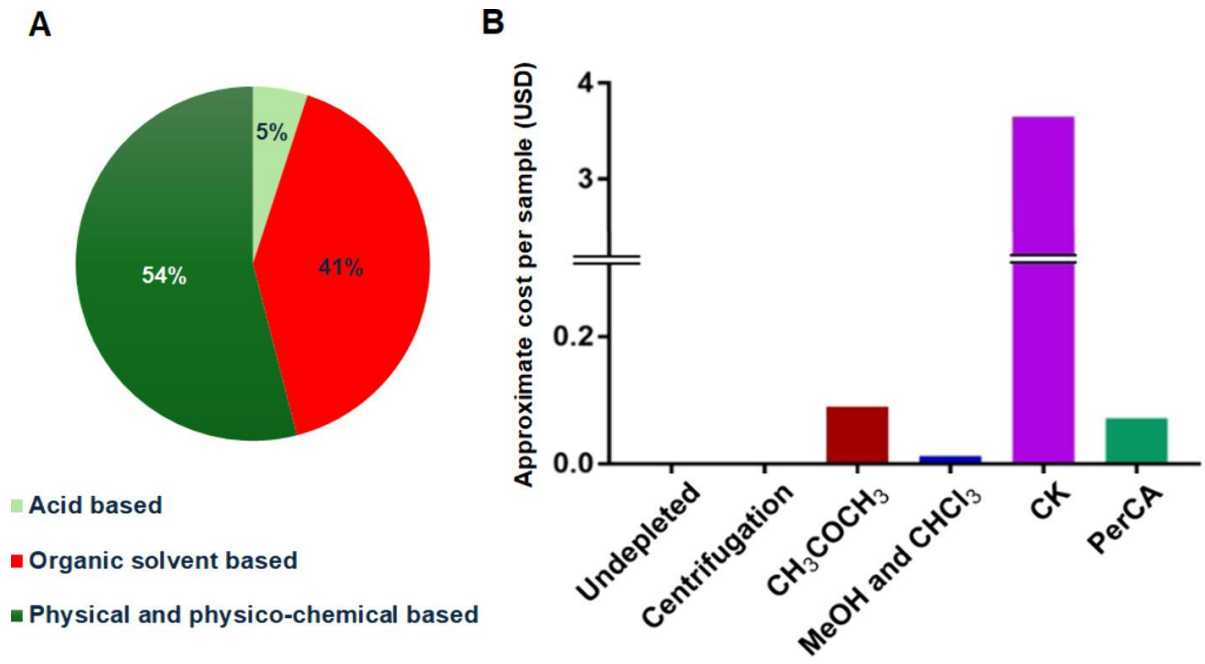


Figure 2-1. Distribution of protein depletion methods used in milk proteomics. (A) A pie chart illustrates the percentage of methods used for protein depletion in milk. The most employed approach is physical and physicochemical methods (54%), followed by organic solvent-based methods (41%) and acid-based methods (5%). (B) Bar chart comparing the cost per sample for processing using each precipitation method, highlighting the differences in expenses between approaches and providing insight into the cost-effectiveness of each technique in milk proteomics research.

The sequencing depth of low-abundance proteins from a range of biofluids, such as urine, serum, and plasma, can be significantly increased using various commercial kits (40–42). However, to the best of our knowledge, there are not many commercial kits available for depleting the major proteins from milk samples. Furthermore, employing

various commercial kits to deplete the major proteins costs between USD 1 and USD 100, which is a significant barrier for large-scale sample analysis, as well as for many other countries throughout the world (9). To establish cost-effective methods for depleting abundant proteins from human milk, it is necessary to perform comprehensive studies with different depletion methods based on various mechanisms, allowing for improved deep proteomics studies.

Because the efficiency of various depletion techniques in enriching trace proteins varies, it is necessary to evaluate different methods and determine the most suitable techniques for studying the milk proteome. In this study, we employed five different methods: centrifugation, acetone precipitation, methanol-chloroform precipitation, depletion using a commercial kit (CK), and PerCA precipitation to deplete abundant proteins (Figure 2-2). Among the evaluated methods, the CK is the most expensive at approximately USD 3.65 per sample, followed by acetone-based depletion (USD 0.10), PerCA treatment (USD 0.01), methanol-chloroform-based depletion (USD 0.012), and the undepleted and centrifugation methods, which were considered as the reference (zero USD per sample) (Figure 2-1B). The above costs were obtained from vendors (for commercial kits) or estimated based on the materials' price and amount (for solvent or acid) of each sample treatment. This parameter provides important insight into designing and conducting proteomic-based experiments (Figure 2-2). Through this comparative study, we aim to determine the most effective method for isolating, profiling, and analyzing low-abundance and low-molecular-weight proteins (LMWPs) in human milk. This comprehensive evaluation provides valuable insights into the functional roles of these trace proteins in immunity, disease pathways, and infant health (43). Overall, our findings

emphasize the potential of human milk as a diagnostic tool, offering critical information about maternal health and facilitating advancements in both nutritional science and disease research.

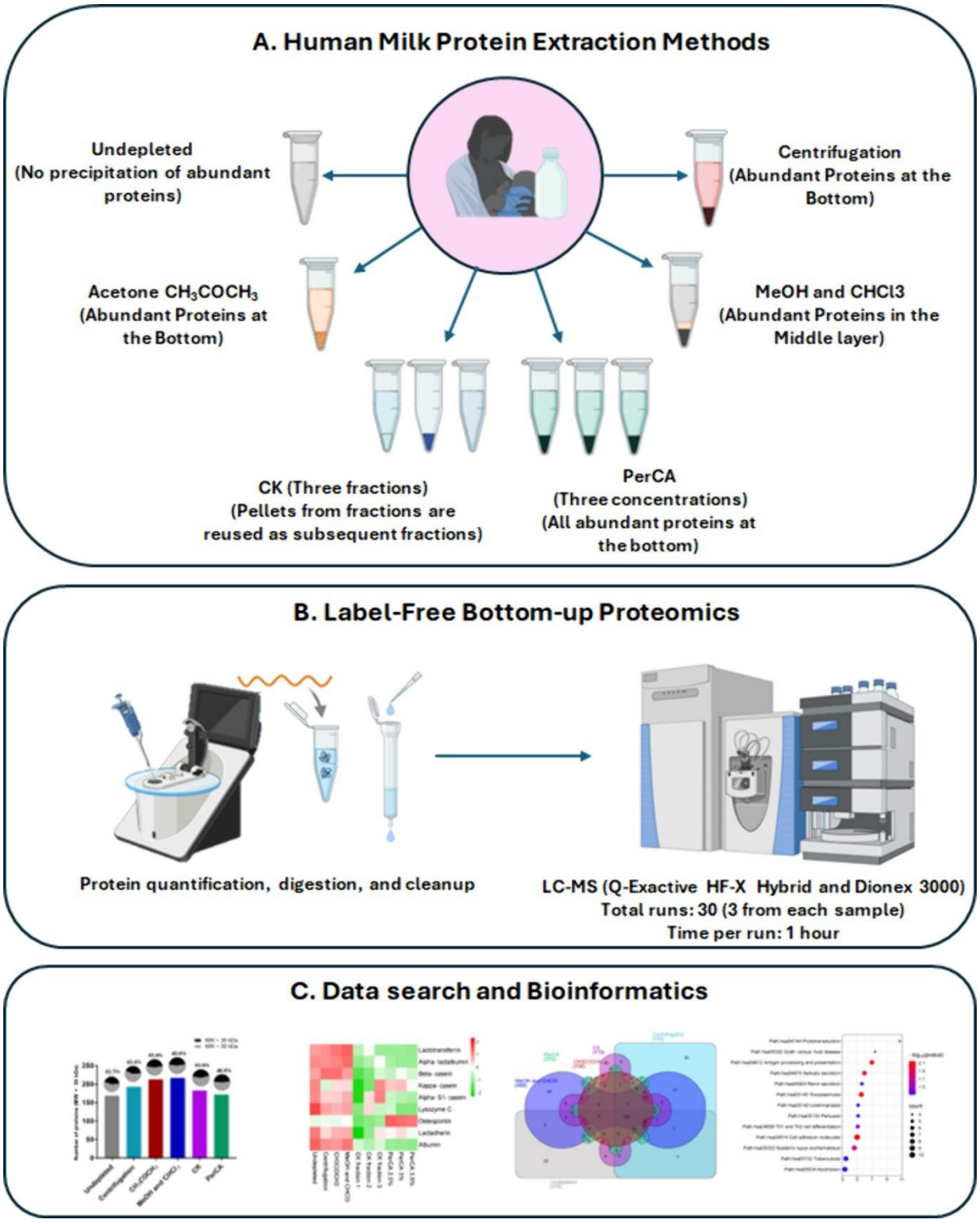


Figure 2-2. Schematic representation of the experimental workflow and data analysis for proteomics of human milk samples obtained from a milk bank using six methods. (A) Proteins were analyzed from whole, undepleted milk and after the depletion of major proteins using centrifugation, acetone precipitation, methanol–chloroform precipitation, a commercial kit (CK), and perchloric acid precipitation. (B) Workflow of bottom-up proteomics. Proteins were quantified, digested, cleaned, and analyzed by using LC-MS. (C) Data analysis and bioinformatic studies of the LC-MS output. The figure was created in BioRender (https://BioRender.com/ba37fbc).

2.3 Materials and Methods

2.3.1 Collection of Human Milk Samples

Human milk samples were collected from the Oklahoma Mothers Milk Bank (OKC, OK). Samples were stored at −80°C until further processing. Before protein extraction, the sample was thawed at 4°C and homogenized to ensure uniformity. It is important to note that the Oklahoma Mothers' Milk Bank gathers donor milk with informed consent as part of standard banking practices. Since samples were de-identified and provided for research use, our study was deemed exempt from IRB review under institutional guidelines. The human milk donor used in our study was obtained prior to pasteurization. All donors were screened according to the Human Milk Banking Association of North America (HMBANA) guidelines, which include completion of detailed medical and lifestyle history, blood testing for HIV-1 and -2, HTLV, hepatitis B and C, and syphilis as well as physician clearance. Donors were excluded if they used tobacco, recreational drugs, or incompatible medications or had recent blood transfusions or organ transplants.

2.3.2 Protein Depletion Methods

A total of five different abundant protein depletion methods, which are based on different working mechanisms and have been used to deplete abundant proteins from other types of samples, were used for comparative analysis (9,42). These are as follows: high-speed centrifugation, acetone precipitation, methanol and chloroform precipitation,

commercial kit (CK), and perchloric acid (PerCA) precipitation. High-speed centrifugation can effectively remove abundant milk proteins (e.g., casein and whey proteins) as a precipitated pellet (44,45); the CK dissolves HAPs and then precipitates the low-abundance proteins; and organic solvents can induce protein aggregation, resulting in precipitation (8,42,46). Thaw-collected human milk samples were thawed on ice until they are fully liquid, and 100 µL of the milk sample was aliquoted into a low-binding Eppendorf tube. Throughout this process, the samples must remain cold to prevent any potential protein degradation. For each method, three replications were used. Figure 2-2 represents the experimental details.

Processing of Undepleted Milk (Control). A milk sample (100 µL) was mixed with an equal volume (100 µL) of 8 M urea in a 1.5 mL low-binding tube. The mixture was vortexed thoroughly to ensure proper mixing. Subsequently, 400 µL of 20 mM HEPES buffer was added to the tube, and the solution was centrifuged at 16,000 × g for 2 min at room temperature. After centrifugation, the sample was stored at −80°C until further processing and protein quantification using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, IL, USA).

Centrifugation. The thawed milk samples (100 µL) were centrifuged at 1,500 × g for 10 min at 4°C to separate the lipid layer from the aqueous phase, known as the milk serum (47). The aqueous phase was carefully transferred to a new low-binding tube. To further purify the milk serum, it was centrifuged at 17,000 × g for 10 min at 4°C, incorporating minor modifications to the previously established protocol (30). After centrifugation, the clear supernatant was collected, which is the purified milk serum used for protein extraction.

Acetone Precipitation. Milk samples (100 µL) were mixed with 4× (400 µL) of cold acetone (−20°C), vortexed thoroughly, and stored at −20°C for 2 h. The resulting protein pellet was centrifuged at 14,000 × g for 10 min at 4°C, and then the supernatant was discarded. Another 400 µL of cold acetone (−20°C) was added to the protein pellet, vortexed thoroughly, and centrifuged again at 14,000 × g for 10 min at 4°C. The supernatant was discarded. The protein pellets were air-dried and lysed with 100 µL of 8 M urea and subjected to quantification using the NanoDrop One spectrophotometer.

Methanol and Chloroform Precipitation. Methanol-chloroform precipitation was performed as described previously (48). Briefly, 400 µL of methanol was added to 100 µL of the milk protein sample, vortexed thoroughly, and then centrifuged at 9,000 × g for 30 s. Next, 200 µL of chloroform was added, vortexed well, and centrifuged at 9,000 × g for another 30 s. Then, 300 µL of water was added, vortexed again, and centrifuged at 10,000 × g for 15 min at room temperature. The top aqueous layer was carefully removed and discarded, leaving behind the interphase layer that contains the protein. Methanol (400 µL) was added to the remaining chloroform layer, vortexed vigorously, centrifuged at 10,000 × g for 5 min at room temperature, and then the supernatant was discarded while retaining the protein pellet. The protein pellets were air-dried, lysed with 100 µL of 8 M urea, and subjected to quantification using the NanoDrop One spectrophotometer.

Commercial Kit (CK). A commercial kit (Minute Organic Solvent-Free Milk Lipid Depletion Kit, Cat. No. ML-044, Invent Biotechnologies, Inc.), which separates the lipid compounds from milk, was used to extract milk proteins. The procedure was carried out in compliance with the manufacturer's protocol. Briefly, whole milk (100 µL) was mixed with an equal volume of CK buffer A (100 µL) in a 1.5 mL low-bind tube, homogenized by

pipetting up and down 10–20 times, and incubated on ice for 10 min. The sample was centrifuged at 16,000 × g for 5 min at room temperature, and the supernatant, which represented fraction 1 (CK Fraction 1), was carefully transferred to a new low-binding tube. The remaining pellet was resuspended in 100 µL of PBS buffer and 200 µL of CK buffer B (the CK buffer B bottle was shaken vigorously before adding), yielding a final volume of 300 µL. After vortexing for 20 s and incubating for 2–3 min at room temperature, the sample was centrifuged again at 16,000 × g for 5 min, resulting in three distinct layers: a lipid phase (top), an aqueous phase (middle), and a pellet (bottom). The lipid phase was removed carefully with Kimtech KimWipes to prevent contamination. The aqueous phase (200–250 µL), representing fraction 2 (CK Fraction 2), was collected, mixed with an equal volume of 8 M urea, vortexed, and stabilized with double the volume of HEPES buffer before protein quantification. The pellet was resuspended in 100 µL of 8 M urea, vortexed, treated with 200 µL of HEPES buffer, and quantified as fraction 3 (CK Fraction 3). All three CK fractions (CK1–CK3) underwent in-solution trypsin digestion, and each fraction underwent a separate LC-MS/MS analysis.

Perchloric Acid (PerCA) Precipitation. The PerCA precipitation depletion method was adopted using previously described methods (8,49). In this investigation, we employed three different final concentrations of PerCA (2.5%, 3.0%, and 3.5%) to precipitate proteins from the milk samples, even though there is no specified optimal concentration for milk protein extraction. Briefly, equal volumes of milk (100 µL) and PerCA (100 µL) were added to a low-binding Eppendorf tube and mixed vigorously, and the mixture was kept on ice for 15 min. Afterward, samples were centrifuged at 16,000 × g for 10 min. The supernatant (100 µL) was collected carefully, avoiding the top lipid layer.

This supernatant was further lysed with an equal volume of 8 M urea and subjected to protein quantification. The supernatant obtained from each PerCA concentration was separately analyzed by LC-MS/MS, and the proteomics data were pooled together for comparison.

2.3.3 Sample Preparation, LC-MS/MS, and Data Analysis

For in-solution peptide digestion, a total of 100 µg of protein was taken from each sample. Prior to adding Trypsin/Lys-C (catalog no. V5071, Promega, WI, USA) to the samples, 1.0 µg of yeast enolase (CAS 9014-08-8, Sigma-Aldrich, Inc., St. Louis, MO, USA) was added as the internal standard. In-solution digestion was performed as described previously (9). Peptides were desalted with a Waters C18 Sep-Pak cartridge (cat. no. WAT023590, Waters, Milford, MA, USA) and subsequently dried using a Speed-vac (SVC 100H, Savant) for 8 h. Each sample was then resuspended in 100 µL of mobile phase A (water with 0.1% FA). An equal amount (2 µg) of sample was injected into the LC-MS/MS analysis.

The LC-MS/MS analysis was conducted according to our previously published protocols (9). This analysis utilized a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific, CA, USA) connected to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The resuspended tryptic peptides were separated using an EasySpray (3 µm, 75 µm × 15 cm, ES900, Thermo Fisher Scientific, Germany) analytical column. Peptide separation took place over a total run time of 60 min with an analytical gradient of ACN from 5% to 35% over the first 30 min at a flow rate of 350 nL/min, followed by a rapid ramp up to 95% ACN for the wash step, and then equilibrated back to 5% ACN for the final 10 min.

MS RAW files were searched against the UniProt reviewed human (TaxID: UP000005640) protein database using the Sequest algorithm within Proteome Discoverer version 2.4 (Thermo Fisher Scientific, San Jose, CA). The Sequest database search was conducted utilizing the following parameters: specific cleavage by the trypsin enzyme, allowance for two potential missed cleavages, a mass tolerance of 10 ppm for precursor ions, and 0.02 Da mass tolerance for fragment ions. The search parameters included the dynamic modification of methionine oxidation (+15.9949 Da) and the static modification of carbamidomethylation (+57.0215 Da) on cysteine. The peptide assignments derived from the database search were filtered to achieve a false discovery rate (FDR) of 1%. The process of label-free quantitation across the samples employed the Minora algorithm alongside the available bioinformatics tools within Proteome Discoverer. A statistical significance threshold was set, where a 1.5-fold increase or decrease in abundance with a p-value of less than 0.05 was deemed significant.

2.3.4 Bioinformatics

All functional and pathway analyses were performed using the ShinyGO 0.81 bioinformatics platform (50). The Venn diagrams were created using an online platform InteractiVenn (51). Heat maps and an abstract graph were created using BioRender (https://www.biorender.com/). An Upset plot was generated using SRplot (https://www.bioinformatics.com.cn/en), a free online platform for data analysis and visualization. The protein-protein interaction analysis was performed using STRING (https://string-db.org/) (52). All bar plots were generated using GraphPad Prism version 7.0 for Windows, GraphPad Software, Boston, Massachusetts, USA

(https://www.graphpad.com). The other components were made using Microsoft PowerPoint and Excel 365.

2.4 Results and Discussion

Since most biological samples are dominated by a small number of extremely abundant proteins, which significantly limit the detection level of low-abundance proteins, it is imperative to remove the abundant proteins from samples (53). Commercial kits are definitely very useful in accelerating large-scale research based on their method of depleting abundant proteins. Furthermore, it has already been demonstrated that these commercial kits have the selectivity to enrich a certain protein group because of their mode of action (9). To the best of our knowledge, no comprehensive analysis had been conducted comparing the commercial and non-commercial depletion methods, despite the fact that a number of articles have been published on profiling the human milk proteome (54–56). Herein, we examined the efficacy of five depletion techniques, wherein one commercial (CK) and four non-commercial, such as centrifugation, acetone precipitation, methanol-chloroform precipitation, and PerCA treatment, were used to enrich low-abundant proteins. The efficacy assessment was based on several important parameters, including the number of identified proteins and peptides, cost per sample, enrichment of low-abundance proteins, and biological insights of these methods related to identifying proteins associated with infant and maternal health, immunity, and disease-related pathways.

2.4.1 Proteome Landscape of Human Milk Identified by Different Depletion Methodologies

As anticipated, there were considerable differences in the overall quantity of peptides and proteins found by using each approach (Figure 2-3A). Among the precipitation techniques, methanol-chloroform precipitation yielded a total of 2,668 peptides corresponding to 468 protein groups. On the other hand, the PerCA acid-based approach yielded a total of 1,597 peptides corresponding to 370 proteins (Table S2-2 and Table S2-3). It is also important to note that we tested three distinct PerCA concentrations (2.5%, 3.0%, and 3.5%) to deplete the main proteins. Our findings further showed that the percentage of PerCA also affects the key proteins' depletion (Figure S2-2 and Table S2-2). Overall, the efficiency of each precipitation method in identifying protein-peptides is as follows: methanol-chloroform > acetone > centrifugation > CK > PerCA (Table S2-2 and Table S2-3).

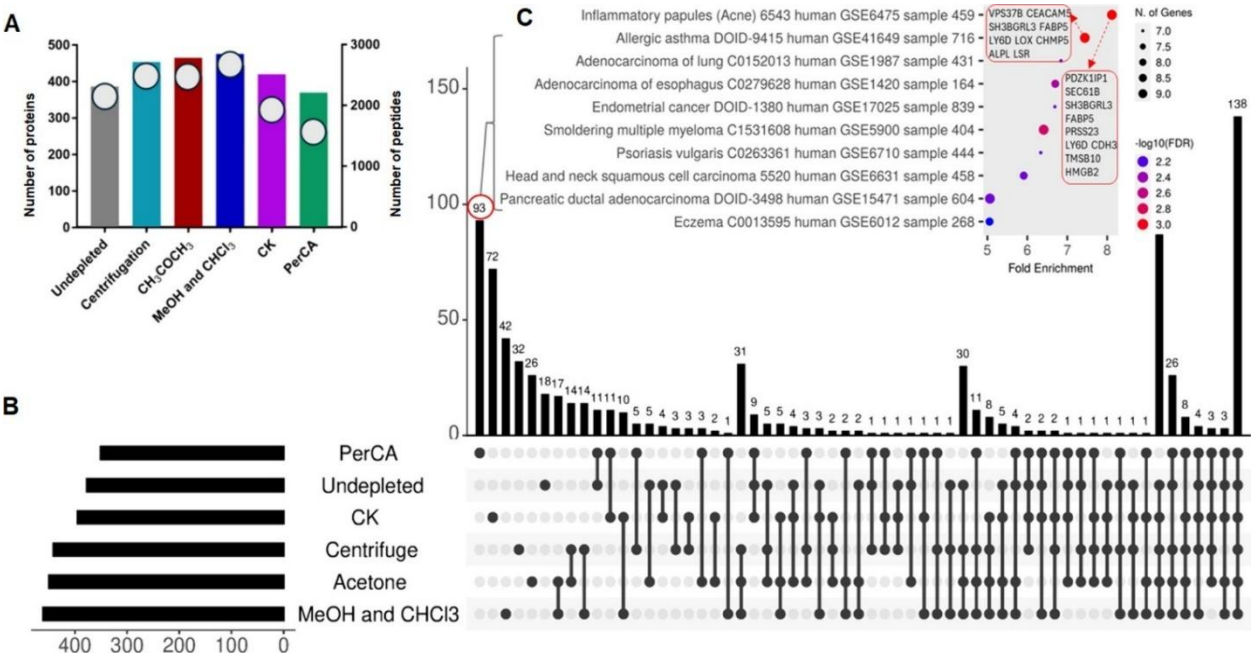


Figure 2-3. Comparison of protein and peptide identification across different precipitation (depletion) methods. (A) Bar graph showing the total number of proteins and peptides identified using each depletion method, including pooled proteomics data from the three

fractions of the commercial kit (CK) and perchloric acid at 2.5, 3, and 3.5% concentrations. (B) Upset plot displaying the overlap and unique proteins identified across all methods, highlighting that perchloric acid (pooled across all concentrations) resulted in the highest number of unique proteins, followed by CK (pooled across all fractions), while undepleted samples identified the fewest unique proteins. (C) KEGG pathway analysis of the PerCA-specific 93 unique protein sets.

A pairwise comparison of each depletion method against the undepleted milk sample further revealed that employment of methanol-chloroform (31.53%) and PerCA (31.03%) led to the detection of more unique proteins than the other methods (28.7–29.5%) (Figure S2-3). Similarly, the common proteins (%) between undepleted milk and PerCA-depleted samples were also the lowest (36.3%), while the overlapped proteins (%) between other depletion techniques and undepleted samples ranged from 48.12% to 57.12% (Figure S2-3). A comprehensive comparison of method-specific identified proteins with an undepleted sample shows that PerCA identified the highest number (93) of unique proteins, followed by CK (72), methanol-chloroform (42), centrifugation (32), and acetone (26) (Figure 2-3B). A disease perturbation pathway analysis of these 93 unique proteins identified by PerCA methods further showed that a number of potential marker proteins associated with inflammatory papules, allergic asthma, and cancers were exclusively found using this specific depletion method (Figure 2-3C and Table S2-4). These results revealed that while PerCA detects fewer proteins compared to other depletion methods, a higher percentage of them are distinct, suggesting that it is useful in enhancing the identification of low-abundance proteins and/or potential biomarkers for various diseases (8,49). Ultimately, these findings emphasize the importance of selecting appropriate depletion techniques to achieve a more comprehensive understanding of milk's complex protein landscape.

2.4.2 Enrichment of Low-Molecular-Weight Proteins (LMWPs)

LMWPs and bioactive peptides found in human milk are crucial for helping infants with digestion, supporting their immune system, and protecting them from infections (57,58). However, studying these proteins can be tricky because they exist in small quantities and are often overshadowed by more abundant proteins like β-casein and α-lactalbumin (59). While several commercial kits are available for depleting major serum or plasma proteins, their use in large-scale projects or low-resource settings is limited due to their incapability to be utilized with milk samples (9,60). This limitation necessitates the selection of ideal methods that are both cost-effective and efficient for the enrichment of LMWPs in milk. Thus, optimization of various techniques could enhance the accessibility of LMWP analysis in various neonatal research and diagnostic applications.

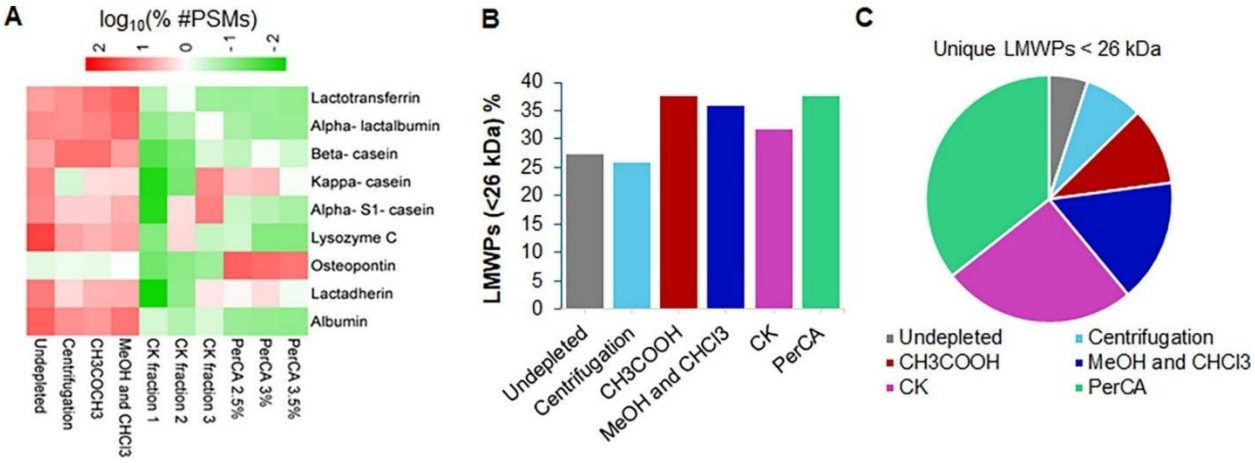


Figure 2-4. Efficacy evaluation of major protein depletion methodologies in human milk samples. (A) A heat map illustrates the efficacy of each depletion method in precipitating the nine most abundant proteins found in human milk. (B) A bar diagram displays the percentage of unique LMWPs identified by using each depletion method. (C) A pie chart shows the number of unique LMWPs identified by using each depletion method.

We further demonstrate the depletion efficiency of each method by evaluating the abundance of the nine most abundant proteins in human milk, including lactotransferrin, α-lactalbumin, β-casein, κ-casein, αs1-casein, lysozyme C, osteopontin, lactadherin, and

albumin (Table S2-2). Among the tested methods, we found that PerCA and CK exhibited the highest depletion efficiency by reducing the levels of these nine abundant proteins compared to other methods (Figure 2-4A).

Our findings also showed a direct correlation between the decrease in the abundance of milk proteins and the discovery of LMWPs. The depletion procedures used in this work significantly boost the enrichment of unique LMWPs by more than 10% (Figure 2-4B), with PerCA-based approaches demonstrating the greatest results (37.5%) of unique LMWP identification when compared to other techniques (Figure 2-4C). Thus, PerCA not only enriches LMWPs but also provides deeper insights into proteome complexity by selectively enriching unique, low-abundance LMWPs, making it a valuable tool for studying post-translational modifications using top-down proteomics (61,62).

2.4.3 Comparative Study with Various Milk Proteome Datasets

The proteome sequencing depth of biological fluids could be increased by using novel protein extraction techniques, which always aid in the discovery of new protein sets [24]. The human milk proteome has recently been profiled by a number of research teams under various experimental and biological conditions, most notably by the use of some kind of depletion approach (54,55,63–66).

To evaluate the efficacy of the various depletion techniques in identifying low-abundance proteins, we compared our proteome dataset with a number of recently published studies that identified a larger number of proteins than our proteome dataset (Figure 2-5A and Table S2-5). A Venn diagram analysis illustrating the overlap of proteins among the study groups revealed that 115 unique proteins were found exclusively in our study (Figure 2-5A and Table S2-5). This implies that techniques for reducing the

abundance level of major proteins can significantly improve the detection limit of low-abundance proteins, hence enhancing the depth of rare protein sequencing in milk samples.

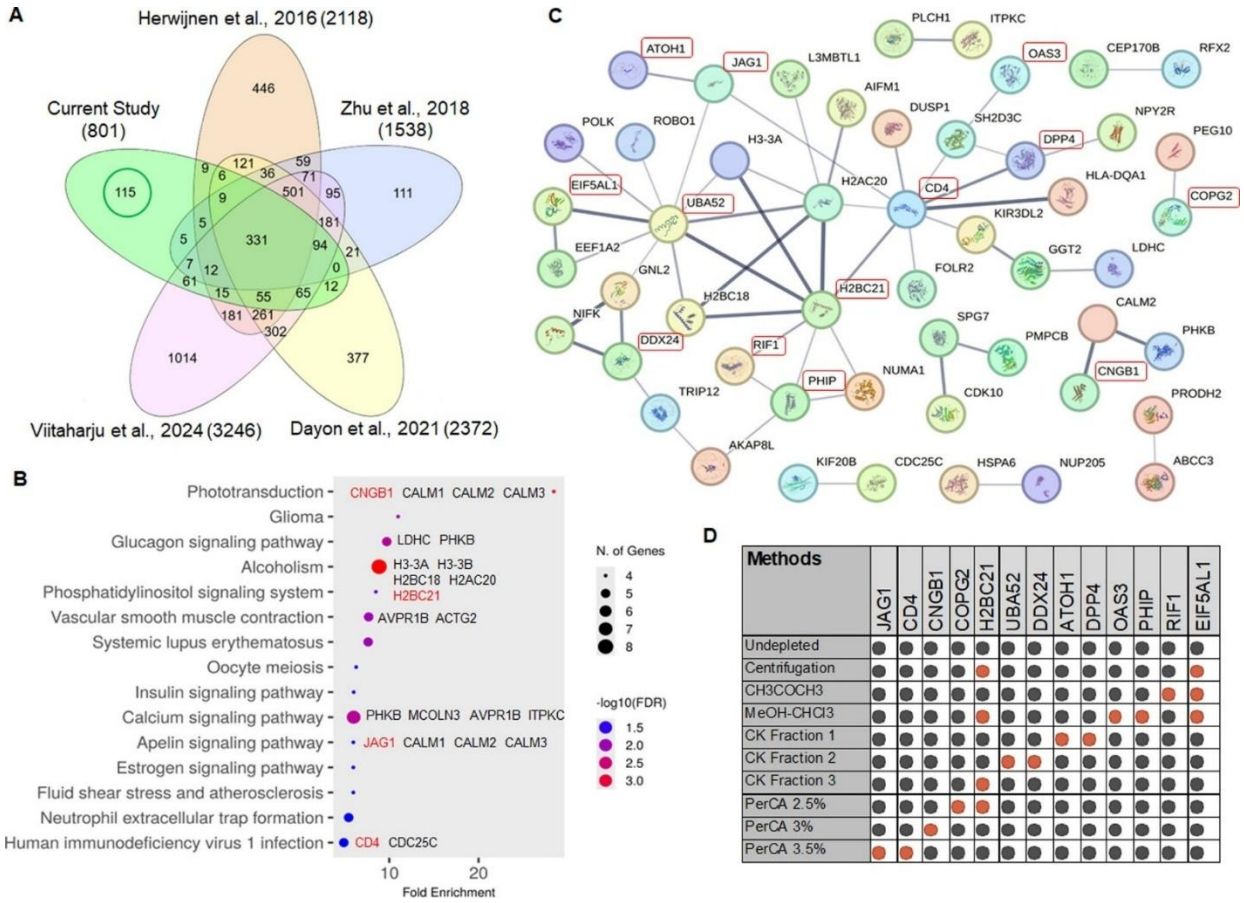


Figure 2-5. Comprehensive analysis of unique proteins identified in our study compared to previous proteomic studies on human milk over the past decade. (A) A Venn diagram compares the unique proteins identified in our study with those reported in previous studies, including Herwijnen et al. (2016), (54) Zhu et al. (2018), (64) Dayon et al. (2021), (55) and Viitaharju et al., (2024). (65) (B) KEGG pathway analysis shows the functional roles of the unique protein sets identified in the current study (green circle). (C) Protein–protein interaction map of the unique protein sets identified in this study. (D) Example of method-specific target protein (red box in panel (C)) sets that showed a strong association with other proteins in the protein–protein network and are related to various neonatal developmental processes. Proteins present and absent in a specific method are indicated by red- and black-filled circles, respectively.

To better understand the possible functional diversity of these unique protein sets, a KEGG pathway analysis was performed (Figure 2-5B). The identification of

phototransduction pathways raises intriguing possibilities about the potential involvement of human milk proteins in early visual development (67). It is also important to note that these distinct protein sets were further shown to have a substantial connection through protein-protein interaction (PPI) network analysis (Figure 2-5C). Our results further demonstrated a strong association with a range of proteins, including JAG1, CD4, CNGB1, COPG2, H2BC21, UBA52, DDX24, ATOH1, DPP4, OAS3, PHIP, RIF1, and EIF5AL1, which were specifically found in this investigation and displayed method-specific enrichment (Figure 2-5D). H2BC21, an LMWP belonging to the histone gene family, has been associated with immune infiltration, tumor growth, and transcriptional regulation in malignancies like gastric cancer and glioma. Furthermore, the presence of H2BC21 in amniotic fluid suggests a potential immunomodulatory role during fetal development, consistent with studies showing that amniotic fluid contains innate immune proteins and inhibits proinflammatory signals like TLR4 to protect against infection and inflammation (68–72). UBA52 is also an LMWP; it plays a dual role as a ubiquitin-ribosomal fusion protein, contributing to both protein synthesis and cellular stress responses, and has been implicated in neurodegenerative disorders, DNA damage repair, and cell cycle regulation (73–75). CD4, a crucial immune coreceptor, is essential for T-cell activation and signaling, and its detection in milk highlights the potential immunomodulatory roles of specific low-abundance proteins in supporting neonatal adaptive immunity (76). Thus, functional analysis and protein-protein interaction studies revealed that these unique proteins play essential roles in brain development and cognitive function, immune system and defense mechanisms, overall infant growth and protection, and maternal health and well-being (77–83).

Table 2-1. Pros and Cons of Different Protein Depletion Methods

criteria	undepleted	centrifugation	CH_3COCH_3	MeOH and $CHCl_3$	CK	PerCA
cost/sample	–	–	+	+	+++	+
abundant protein removal	–	+	+	+	+++	+++
protein/peptide identification	++	+++	+++	+++	++	++
unique protein identification	+	++	++	++	+++	+++
protein MW < 26 kDa	+	+	+	++	+	+++
sample processing time	+	+	+	++	+++	++
throughput	+++	+++	+++	++	+	++

2.4.4 Comparison Among Different Abundant Protein Depletion Methods

The key highlights of our study are summarized in Table 2-1, comparing the advantages and limitations of different protein depletion methods used in profiling human milk samples. The table evaluates various methods based on cost per sample, efficiency in removing abundant proteins, overall protein and peptide identification, identification of unique proteins, and the percentage of LMWPs (≤26 kDa). Although the cost of CK is the highest, it performed well for the efficient depletion of abundant proteins and detection of unique proteins. Particularly, PerCA demonstrates superior efficiency in depleting high-abundance proteins compared with methanol-chloroform precipitation, allowing for the detection of a higher proportion of LMWPs. Precipitation by organic solvents and centrifugation showed excellent performance in identifying more proteins. Nevertheless, the overall performance of the PerCA precipitation method was excellent according to all evaluation criteria.

2.5 Limitations of the Study

We recognize that the protein extraction technique alone may not be the only factor that contributes to the enhancement of protein sequencing depth, especially in the case of identifying low-abundance proteins in human milk samples. Donor variability and sample size (47,54), LC-MS/MS equipment (84,85), LC-MS/MS methodologies (55,65), and chromatographic column performance (86) are some important aspects that may have a big impact on protein identifications, especially for low-abundance proteins. These factors were incorporated to provide a more impartial analysis of the data and to emphasize the significance of standardized procedures in milk proteomics studies.

2.6 Conclusions

This study illustrates that abundant protein depletion methods significantly influence the characterization of the human milk proteome. Five different depletion methods with various mechanisms were evaluated for their cost, efficiency of depleting abundant proteins, total number of protein and peptide identifications, number of identified unique proteins, and capability of detecting LMWPs. Among these methods, PerCA stands out as a cost-effective alternative to other methods, making it a more accessible choice for research applications. Our results indicate that PerCA is an effective and practical approach for refining the LMWP human milk proteome, depending on study objectives. Overall, our research expands the human milk proteome landscape by identifying several previously unknown proteins. This comparative analysis provides insights into the trade-offs among all factors, guiding the determination of suitable methods for effective proteomics studies of human milk.

The findings presented in this chapter establish perchloric acid (PerCA) precipitation as the most cost-effective and efficient strategy for enriching unique and low-

molecular-weight proteins from human milk. This validated methodology forms the analytical backbone of the study described in Chapter 3, where both undepleted and PerCA-depleted sample preparations are applied in parallel to characterize how the proteome and mineral composition of preterm human milk evolve across the early weeks of lactation, a clinically meaningful biological question that this methodological foundation made it possible to answer with confidence.

2.7 Associated Content

All LC-MS/MS RAW files and results files can be found at the MassIVE database (https://massive.ucsd.edu) under accession MSV000097294.

CHAPTER 3: DYNAMIC PROTEOMIC AND IONOMIC LANDSCAPE OF PRETERM HUMAN MILK ACROSS EARLY LACTATION

Author Contributions

This chapter presents work from a manuscript in preparation for peer-reviewed publication. The following contributions reflect the roles of each author. **Amit Singh** conceptualized and designed the integrated proteomics and ionomics study in consultation with the corresponding authors. Performed all sample preparation for LC-MS/MS analysis including protein extraction, reduction, alkylation, trypsin/Lys-C in-solution digestion, C18 Sep-Pak desalting, and SpeedVac drying for both undepleted and PerCA-depleted conditions across all 18 samples. Performed all ICP-MS sample preparation, including mixed-acid digestion of milk samples. Performed all proteomics data acquisition on the Q Exactive HF-X mass spectrometer coupled to the Dionex UltiMate 3000 UHPLC system. Conducted all ICP-MS elemental analysis on the PerkinElmer NexION 2000. Conducted all bioinformatics analyses including database searches in Proteome Discoverer 2.4 using Sequest HT, multivariate analyses including PCA and PLS-DA, differential abundance analysis, disease association and therapeutic relevance classification, KEGG pathway enrichment, protein-protein interaction network analysis, and elemental data processing and statistical analysis. Generated all figures and tables presented in this chapter. Wrote the primary draft of the manuscript.

Kelsey B. Hicks assisted with clinical sample collection and coordination, contributed to the clinical interpretation of findings and their implications for neonatal nutrition and NICU practice, and contributed to manuscript writing, review, and editing.

Hala Chaaban conceptualized the study, provided clinical expertise on preterm infant nutrition and NICU care, provided access to the preterm human milk biorepository at the University of Oklahoma Health Sciences Campus from which all samples were obtained, supervised clinical aspects of the research, and contributed to manuscript review, editing, and funding acquisition.

Nagib Ahsan conceptualized the study, provided intellectual direction for the proteomics and ionomics experimental design and analytical strategy, supervised bioinformatics and data interpretation, and contributed to manuscript writing, review, and editing.

3.1 Abstract

Human milk (HM) is the optimal source of nutrition for preterm infants, providing antimicrobial, immunologic, metabolic, and neurodevelopmental benefits that significantly improve clinical outcomes in the neonatal intensive care unit (NICU). Despite its recognized importance, the comprehensive characterization of the preterm human milk proteome and how it evolves across the early weeks of lactation remains incomplete. In this study, we applied label-free LC-MS/MS proteomics and inductively coupled plasma mass spectrometry (ICP-MS) to paired milk samples collected at one week and four weeks postpartum from eight mothers of very preterm infants (gestational age 22–31 weeks). Two complementary sample preparation strategies were employed in parallel: undepleted milk (control) and perchloric acid (PerCA)-depleted milk. Across all samples, a total of 1,043 proteins were identified, 878 in undepleted and 648 in PerCA-depleted preparations, with 462 proteins shared between the two workflows. Multivariate analyses including partial least squares discriminant analysis (PLS-DA) and principal component analysis (PCA) demonstrated clear separation of samples by lactation stage in both datasets. Differential abundance analysis identified 49 proteins upregulated and 21 downregulated at four weeks versus one week in undepleted samples, and 29 upregulated and 17 downregulated in PerCA-depleted samples. Stage-specific proteins in the PerCA dataset, including ELANE, BPI, and MINDY1 elevated at four weeks, and MUC5B, COX5A, and LY6E elevated at one week, were characterized for disease associations and therapeutic relevance. Cross-study comparison with three recently published human milk proteomics datasets identified 153 proteins unique to this study. Complementary ICP-MS elemental profiling quantified changes in ten elements, iron, phosphorus, magnesium, calcium, potassium, sodium, zinc, copper, manganese, and

mercury, across the two lactation stages. Together, these findings provide the first integrated proteomic and ionomic characterization of preterm human milk across early lactation, establishing a molecular foundation for understanding how the composition of this clinically important fluid evolves to meet the changing needs of the preterm infant.

3.2 Introduction

Human milk is the optimal source of nutrition for infants, with antimicrobial, immunologic, metabolic, and neurodevelopmental effects providing both short- and long-term clinical benefits (1). Preterm infants derive particular benefit from human milk feeding (16,87), with documented decreases in the rates of necrotizing enterocolitis, late-onset sepsis (18), chronic lung disease (88), and retinopathy of prematurity (89), as well as improved neurodevelopmental outcomes (90). These clinical effects are mediated in part by the complex and biologically active composition of human milk, which changes dynamically over the course of lactation (3). Human milk produced by mothers of preterm infants differs from that of mothers delivering at term, with higher macronutrient density (3), increased concentrations of key growth and immune factors (91), and greater endogenous protein degradation (92).

Given the biological complexity and dynamic composition of human milk, comprehensive analytical approaches are needed to characterize its protein components. Untargeted proteomics enables large-scale characterization of protein composition, allowing identification of hundreds to thousands of proteins within complex biological samples (93). This approach is well-suited to the study of human milk, given its diverse bioactive components, and has been increasingly applied in recent years.

Prior proteomic studies have largely focused on the milk of mothers who delivered at term, demonstrating that milk composition is highly personalized (94) and influenced by maternal factors including BMI (47,65), gestational diabetes (95,96), COVID-19 infection (97), and even geographical region (98,99). In addition to maternal characteristics, these studies consistently demonstrate substantial variation across lactation stages, with colostrum enriched in total protein and immunologic factors, while mature milk shows increased representation of pathways related to lipid biosynthesis and protein metabolism (6,100).

In contrast, relatively few proteomic studies have examined milk produced following preterm delivery, despite its particular importance for this population. Early work focused on peptidomic differences between preterm and term milk across stages of lactation (92). More recently, a proteomic and transcriptomic analysis of extracellular vesicles isolated from preterm and term milk compared colostrum and mature milk, identifying stage-specific differences in proteins and miRNAs associated with immune modulation, metabolic processes, and developmental signaling (101). However, comprehensive characterization of the broader preterm milk proteome, particularly its temporal evolution during the early weeks of lactation, remains incomplete.

A major challenge in proteomic analysis of biological fluids is the large dynamic range of protein concentrations. As discussed in Chapter 2, a small number of highly abundant proteins dominate the human milk proteome and can obscure the detection of lower-abundance proteins during mass spectrometry-based analysis (7). Strategies that remove or reduce these abundant proteins have therefore been widely used to improve detection sensitivity and expand proteome coverage (11). Chapter 2 of this dissertation

evaluated five different techniques for depleting abundant proteins in human milk and identified perchloric acid (PerCA) precipitation as the most cost-effective and efficient method (10). Building on this finding, the objective of the current study was to characterize the low-abundance proteome of preterm human milk at Week 1 and Week 4 of lactation using complementary abundant-protein depletion strategies, and to identify proteins and pathways that change across this critical window of early lactation. Complementary ICP-MS elemental profiling of the same samples was performed to characterize changes in mineral composition across the same lactation window, providing the first integrated proteomic and ionomic portrait of preterm milk maturation

3.3 Materials and Methods

3.3.1 Study Design and Ethical Approval

Preterm human milk samples were selected from an IRB-approved prospective biorepository with linked maternal and infant clinical data at the University of Oklahoma Health Center. Mothers provided prospective consent for the collection of human milk for future research purposes. The current proteomic and ionomic analysis was conducted under a separate IRB-approved secondary analysis protocol (IRB #: 16838).

3.3.2 Sample Collection and Storage

Milk samples were obtained from the hospital milk laboratory, where mothers had provided bags containing a full expression of milk. Research staff collected 1 mL aliquots from each milk bag, and the date of milk expression recorded on the bag label was used to assign the appropriate lactation time point. Aliquots were initially stored at −20°C in the milk laboratory prior to transfer, then stored at −80°C in the research laboratory until analysis.

3.3.3 Study Population

Mothers who delivered at 22–31 weeks' gestation and had stored milk samples available at both collection time points were included. Selection was limited to paired samples, as the study design required within-mother comparisons across lactation time points. Week 1 was defined as postpartum days 7–10, and Week 4 was defined as postpartum days 21–28. These time points were selected to capture early transitional milk and more mature milk within the preterm lactation period. For each mother, multiple milk expressions within each defined postpartum day range were selected and analyzed as independent samples. Samples were selected and tracked at the infant level, as expressed milk in the clinical laboratory was stored by infant. This pilot study included 8 mothers, one of whom had a twin gestation, yielding samples from 9 infants. Twin-derived samples were retained as independent observations, given the exploratory nature of this analysis.

3.3.4 Sample Preparation Strategy for Protein Profiling

To investigate proteomic differences during early lactation and evaluate the effectiveness of perchloric acid precipitation for enriching low-abundance proteins, human milk samples collected at two postpartum time points (1 week and 4 weeks after delivery) were analyzed using two sample preparation approaches: undepleted milk (control) and perchloric acid (PerCA) precipitation. The undepleted samples represent the native milk protein composition, whereas PerCA precipitation selectively removes abundant proteins through acid-induced precipitation, as validated in Chapter 2.

Prior to processing, milk samples were thawed on ice until fully liquid, and 100 µL of milk was aliquoted into low-binding microcentrifuge tubes. Throughout the procedure, samples were maintained at low temperature to minimize potential protein degradation.

In total, 18 samples were analyzed across the two postpartum groups and two preparation methods. The overall experimental workflow is illustrated in Figure 3-1.

3.3.5 Preparation of Undepleted Milk Samples (Control)

For preparation of undepleted milk samples, 100 µL of whole milk was transferred to a 1.5 mL low-binding microcentrifuge tube and mixed with an equal volume (100 µL) of 8 M urea. During the procedure, samples were kept cold to minimize potential protein degradation. The mixture was vortexed thoroughly to ensure complete homogenization. Subsequently, 400 µL of 20 mM HEPES buffer was added, followed by centrifugation at 16,000 × g for 2 min at room temperature. The resulting solution was stored at −80°C until further analysis and protein quantification, which were performed using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, IL, USA).

3.3.6 Preparation of Perchloric Acid-Treated Milk Samples (Depleted)

A 5% perchloric acid (PerCA) working solution was prepared by diluting stock perchloric acid (60%). For protein precipitation, 100 µL of the prepared PerCA solution was added to 100 µL of human milk in a low-binding microcentrifuge tube and mixed thoroughly. The mixture was incubated on ice for 15 min to allow protein precipitation. Following incubation, samples were centrifuged at 16,000 × g. After centrifugation, 100 µL of the clear supernatant was carefully collected without disturbing the top lipid layer. Protein concentration was then determined, and 8 M urea was subsequently added to the supernatant for further sample processing.

3.3.7 Milk Protein Extraction and Sample Preparation for Label-Free Proteomics

Protein concentration was measured at 280 nm using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, IL, USA). Measurements were performed

on three biological replicates for each milk sample. For downstream proteomic analysis, 100 µg of protein was taken from each sample, and 1.0 µg of yeast enolase (CAS 9014-08-8, Sigma-Aldrich, St. Louis, MO, USA) was added as an internal standard. In-solution digestion was performed using Trypsin/Lys-C (catalog no. V5071, Promega, WI, USA). The resulting peptides were desalted using a Waters C18 Sep-Pak cartridge (cat. no. WAT023590, Waters, Milford, MA, USA) and subsequently dried using a SpeedVac concentrator (SVC 100H, Savant) for approximately 8 h. Each sample was then resuspended in 100 µL of mobile phase A (water containing 0.1% formic acid). Two micrograms of peptides from each sample was injected for LC-MS/MS analysis.

3.3.8 LC-MS/MS Analysis and Data Processing

The LC-MS/MS analysis was performed according to our previously published protocols (10). This analysis utilized a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific, CA, USA) coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). The resuspended tryptic peptides were separated using an EasySpray analytical column (3 µm, 75 µm × 15 cm, ES900, Thermo Fisher Scientific, Germany). Peptide separation was carried out over a total run time of 60 min, with an analytical gradient of acetonitrile (ACN) from 5% to 35% over the first 30 min at a flow rate of 350 nL/min, followed by a rapid ramp to 95% ACN for column washing, then re-equilibration to 5% ACN for the final 10 min.

MS RAW files were searched against the UniProt reviewed human protein database (TaxID: UP000005640) using the Sequest algorithm within Proteome Discoverer version 2.4 (Thermo Fisher Scientific, San Jose, CA, USA). Database search parameters included: specific cleavage by trypsin, two allowed missed cleavages, 10 ppm

precursor ion mass tolerance, and 0.02 Da fragment ion mass tolerance. Dynamic modification of methionine oxidation (+15.9949 Da) and static modification of carbamidomethylation (+57.0215 Da) on cysteine were included. Peptide identifications were filtered to a 1% false discovery rate (FDR). Label-free quantification across samples was performed using the Minora algorithm in Proteome Discoverer. A 1.5-fold increase or decrease in protein abundance with a p-value < 0.05 was considered statistically significant.

3.3.9 Acid Digestion of Human Milk for ICP-MS Analysis

Human milk samples were digested using a mixed-acid procedure prior to elemental analysis by inductively coupled plasma mass spectrometry (ICP-MS). Briefly, 100 µL of human milk was transferred into a pre-labeled 1.5 mL microcentrifuge tube with a punctured lid to allow venting during heating. To each tube, 100 µL of concentrated HCl and 400 µL of concentrated HNO_3 were added, giving a total digestion volume of 600 µL. The tubes were gently mixed and incubated in a water bath at 65–70°C for 1 h. After digestion, samples were allowed to cool, and 125 µL of the digest was transferred into a 15 mL Falcon tube containing 4.875 mL Milli-Q water, resulting in a final diluted volume of 5.0 mL. Diluted samples were mixed thoroughly and stored at 4°C until ICP-MS analysis.

3.3.10 ICP-MS Elemental Analysis

Elemental analysis was performed using a PerkinElmer NexION 2000 ICP-MS. Calibration was carried out using a multi-element ICP-MS Calibration Standard 3 (PerkinElmer), which contained Al, As, Ba, Be, Bi, Cd, Ca, Cs, Cr, Co, Cu, Ga, In, Fe, Pb, Li, Mg, Mn, Ni, K, Rb, Se, Ag, Na, Sr, Tl, U, V, and Zn at 10 µg/mL in 5% HNO_3, together

with a separate Hg standard at 10 µg/mL in 5% HNO_3. Because phosphorus was not included in the multi-element standard, a separate 1,000 µg/mL phosphorus standard (Inorganic Ventures, part no. CGP1-30ML) was used for phosphorus calibration. Instrument response was monitored as counts per second (cps) for the selected isotopes, including As-75, Cd-111, Pb-208, Se-82, Fe-57, P-31, Mg-24, Ca-43, K-39, Na-23, Zn-66, Cu-63, Mn-55, Hg-202, and Y-89. Wash and reagent control samples were included throughout the sequence to monitor carryover and background signal. Elemental concentrations were reported in micrograms per milliliter (µg/mL).

3.3.11 Bioinformatics and Data Visualization

Data processing and visualization were performed using a combination of statistical software and online bioinformatics platforms. Bar diagrams, pie charts, and scatter plots were generated using Microsoft Excel. Pearson correlation analysis and scatter matrix plots were produced using Python (v3.12.12) with the pandas, NumPy, and Matplotlib libraries (102). PCA, heat maps, and volcano plots were generated using the SRplot online platform (103). Gene Ontology (GO) enrichment analysis and functional annotation were performed using ShinyGO version 0.85 (50), with additional GO visualizations generated using SRplot. Venn diagrams were created using InteractiVenn (51) and Venn Diagram Plotter. Protein-protein interaction networks were generated using the NDEx web tool (104–106). Drug target classification and protein family categorization, including TClin and TBio classifications, were performed using the NIH Pharos platform and the Target Central Resource Database (TCRD) (107). Sankey plots

were generated using SankeyMATIC. The graphical abstract was created using BioRender (https://www.biorender.com).

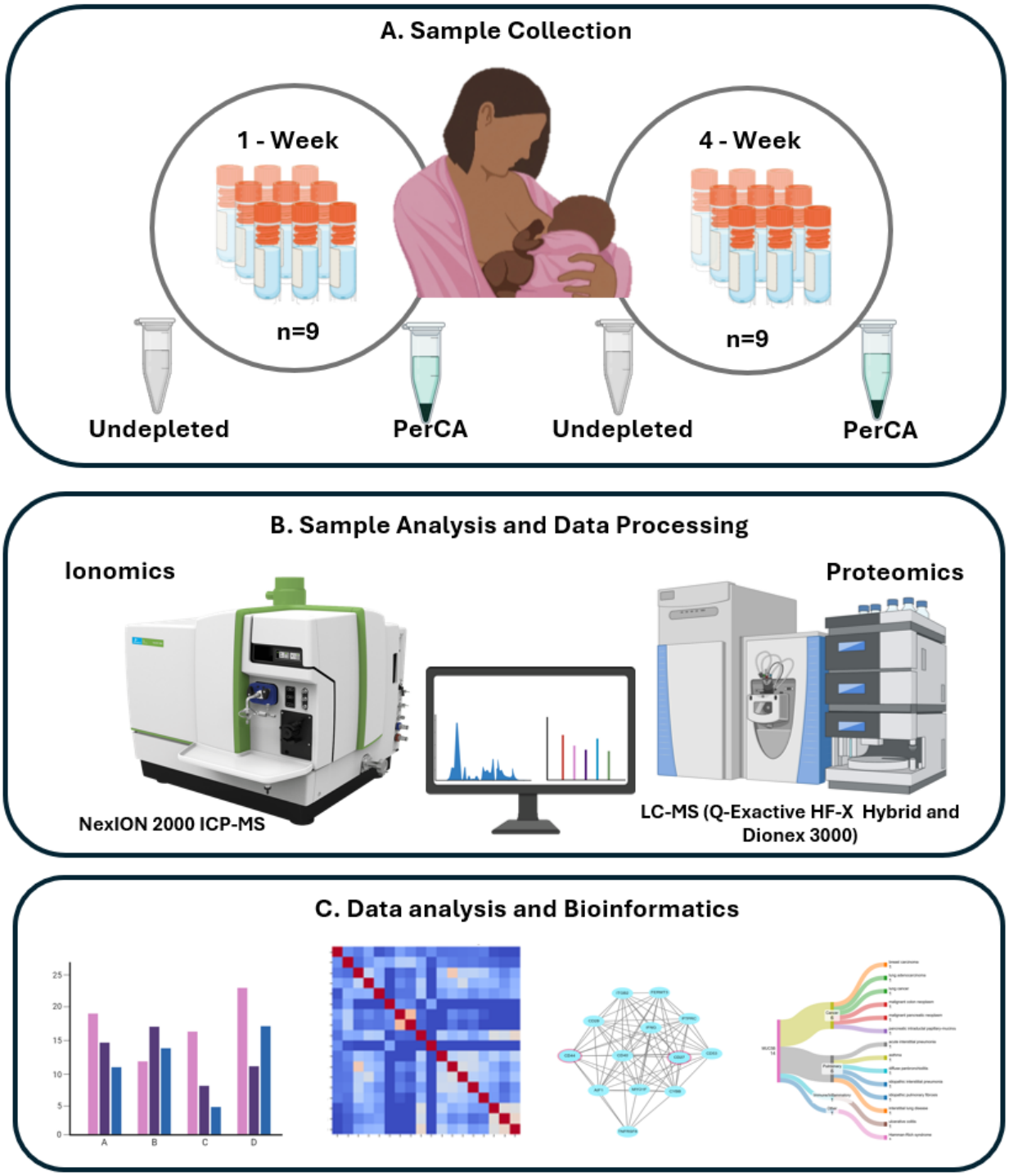


Figure 3-1. Experimental workflow for integrated proteomic and ionomic analysis of preterm human milk across early lactation stages. The upper panel illustrates the proteomics workflow including perchloric acid (PerCA) depletion, protein denaturation,

reduction and alkylation, enzymatic digestion using trypsin, peptide separation by Dionex UltiMate 3000 UHPLC, and protein identification and quantitation by Q Exactive HF-X mass spectrometry. The lower panel illustrates the ionomics workflow including acid digestion of human milk using HCl and HNO3, dilution with Milli-Q water, and elemental quantification by PerkinElmer NexION 2000 ICP-MS. Both workflows were applied to paired milk samples collected at one week and four weeks postpartum, enabling integrated analysis of the preterm milk proteome and ionome.

3.4 Results

3.4.1 Study Population

Table 3-1. Maternal and Infant Clinical Characteristics.

Mothers (n=8)	**Mean (SD) or N (%)**
Age	29.3 (5.3)
Gravida	
Primigravida	2 (25.0)
Multigravida	6 (75.0)
Para	
Primiparous	5 (62.5)
Multiparous	3 (37.5)
Race	
White	4 (50.0)
Other race/not reported[a]	4 (50.0)
Mode of Delivery	
Vaginal	4 (50.0)
Cesarean	4 (50.0)
Multiple birth	1 (12.5)
Comorbidities	
Diabetes	1 (12.5)
Pre-eclampsia	2 (25.0)
Obesity	3 (37.5)
Infants (n = 9)	
Gestational Age (weeks)	26.4 (2.8)
Birthweight	992.2 (301.7)
Female Sex	3 (33.3)

[a] *Racial categories with n < 4 are combined per institutional data security policies*

Eight mothers representing nine infants, including one twin gestation, were enrolled in this pilot study. Maternal and infant demographic and clinical characteristics are summarized in Table 3-1. Mothers had a mean age of 29.3 years (SD 5.3) and were predominantly multigravida (75.0%), though the majority were primiparous (62.5%). Mode of delivery was evenly distributed between vaginal and cesarean birth. Comorbidities were common in this high-risk preterm cohort, with obesity present in 37.5% of mothers, pre-eclampsia in 25.0%, and diabetes in 12.5%. Infants were born at a mean gestational age of 26.4 weeks (SD 2.8) with a mean birthweight of 992.2 g (SD 301.7), and one-third were female (33.3%)

3.4.2 Global Proteome Coverage Across Lactation Stages

To characterize the human milk proteome and evaluate the impact of PerCA depletion, LC-MS/MS analysis was performed on undepleted and PerCA-treated samples collected at one week and four weeks postpartum. Across all samples, a total of 1,043 proteins were identified. Of these, 878 proteins were detected in undepleted samples, whereas 648 proteins were identified following PerCA depletion (Figure 3-2A). Complete protein identification and abundance data for undepleted samples are provided in Supplementary Data 1.

Comparison of protein identifications between the two workflows revealed both shared and method-specific protein populations. A total of 462 proteins were common to both undepleted and PerCA-depleted datasets, while 405 proteins were uniquely detected in undepleted samples and 176 proteins were uniquely identified following PerCA depletion (Figure 3-2B). The 176 proteins uniquely identified following PerCA depletion were further subjected to pathway enrichment and network-based analyses to

evaluate their biological associations (Figures S3-2 through S3-7). Detailed lists of unique and shared proteins, along with GO enrichment analysis of undepleted-unique, PerCA-unique, and common protein sets, are provided in Supplementary Data 2.

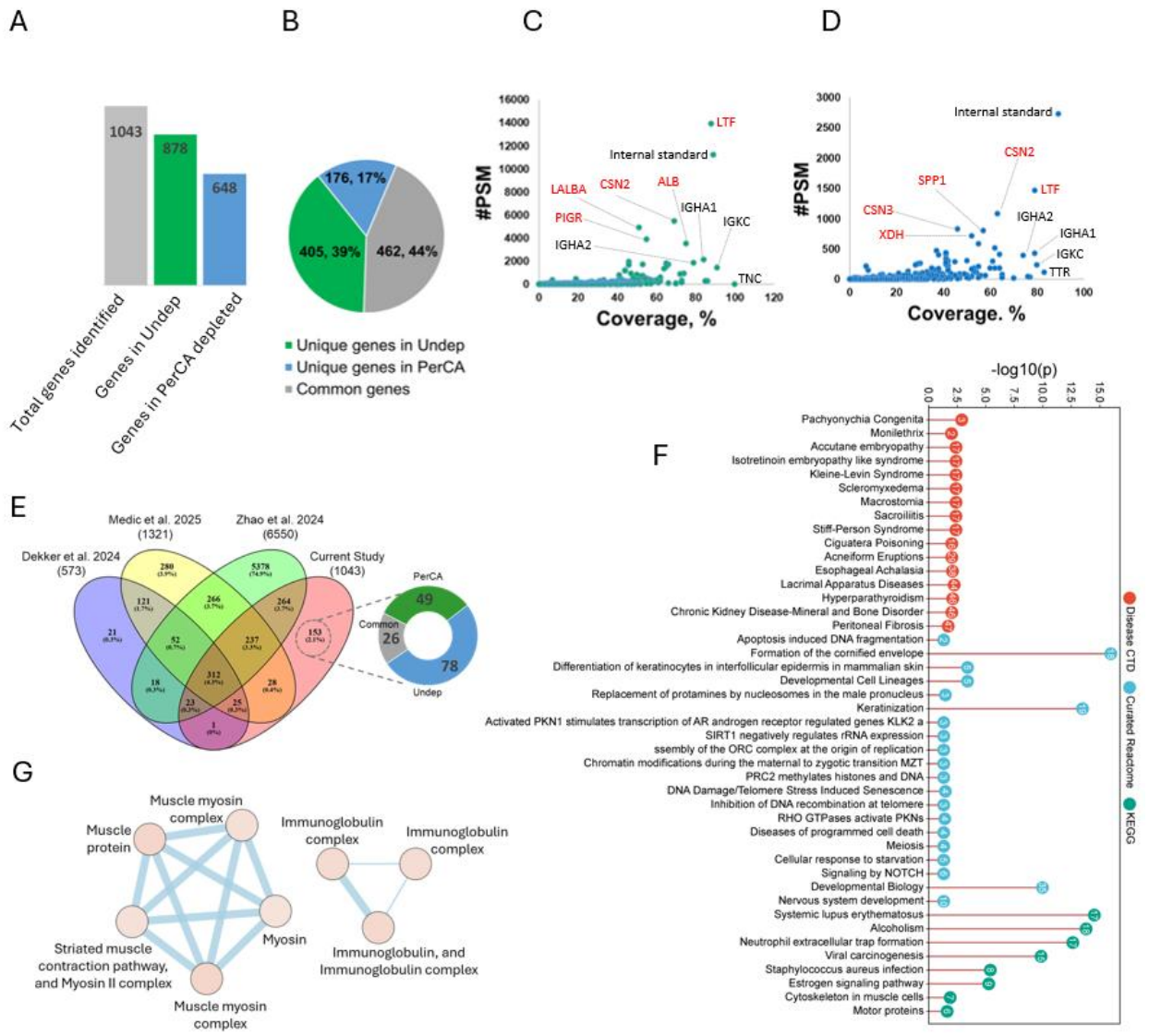


Figure 3-2. Comparative proteomic coverage, dataset characteristics, and biological interpretation of undepleted and PerCA-depleted human milk proteomes. (A) Total number of proteins identified across all samples (1,043), with comparison between undepleted (878 proteins) and PerCA-depleted datasets (648 proteins). (B) Distribution of shared and unique proteins between workflows, showing 462 proteins common to both datasets, 405 uniquely identified in undepleted samples, and 176 uniquely identified following PerCA depletion. (C) Relationship between protein sequence coverage and

peptide spectral matches (PSMs) in the undepleted dataset. (D) Sequence coverage versus PSM distribution in the PerCA-depleted dataset. (E) Comparative analysis of the current dataset with previously published human milk proteomics studies including Dekker et al. (2024), Medic et al. (2025), and Zhao et al. (2024). (F) Integrated pathway enrichment analysis of the 153 proteins uniquely identified in the current study, combining KEGG, Reactome, and disease ontology annotations, revealing enrichment of immune response, epithelial differentiation, host defense, and cell death-related pathways. (G) Protein-protein interaction network of the same 153 unique proteins, illustrating functional clustering into biologically connected modules including muscle/myosin-associated complexes and immunoglobulin-related pathways.

To assess dataset characteristics, the relationship between protein sequence coverage and peptide spectral matches (PSMs) was examined. In the undepleted dataset, proteins exhibited a broad distribution of PSM counts across sequence coverage values, including several proteins with high PSM abundance reflecting the dominant whey and casein proteins (Figure 3-2C). In contrast, the PerCA-depleted dataset showed a more compressed distribution of PSM values across sequence coverage (Figure 3-2D), reflecting the removal of the most abundant species and more even representation of the remaining proteins, a desirable characteristic for discovery-oriented proteomics. Underlying protein sequence coverage and PSM data are provided in Supplementary Data 3.

Comparative analysis with three previously published human milk proteomics datasets, Dekker et al. (2024) (108), Medic et al. (2025) (109), and Zhao et al. (2024) (110), demonstrated both overlap and study-specific protein identifications (Figure 3-2E). Protein lists used for cross-study comparison are provided in Supplementary Data 4. Within the current dataset, 153 proteins were uniquely identified, of which 49 were exclusively detected in PerCA-treated samples, 78 were unique to undepleted samples, and 26 were shared between both preparation methods within this subset. Functional enrichment analysis of these 153 proteins revealed associations with multiple biological

pathways including immune response, epithelial differentiation, host defense, and cell death-related pathways, derived from KEGG, Reactome, and disease ontology databases (Figure 3-2F). Protein-protein interaction analysis of the same 153 proteins showed clustering into functional modules, including muscle/myosin-associated complexes and immunoglobulin-related pathways (Figure 3-2G).

3.4.3 Multivariate and Quantitative Proteomic Profiling Across Lactation Stages

To investigate proteomic variation across early lactation stages, multivariate analyses were performed on both undepleted and PerCA-depleted datasets. Partial least squares discriminant analysis (PLS-DA) demonstrated clear separation between one-week and four-week samples in both workflows (Figure 3-3A–B), confirming that lactation stage is a primary driver of proteome variation. This separation was consistent with principal component analysis (PCA), which also showed clustering of samples according to lactation stage (Figure S3-1A–B).

To further evaluate sample relationships and data consistency, correlation-based analyses were performed. Pearson correlation matrices showed strong intra-group similarity and clustering of biological replicates within each lactation stage in both undepleted and PerCA datasets (Figure 3-3E–F). Scatter matrix analysis supported these observations, showing consistent pairwise relationships in protein abundance across samples (Figure S3-1C–F). Hierarchical clustering of protein abundance revealed distinct patterns between lactation stages, with heatmap visualization of both undepleted and PerCA-depleted datasets showing clear separation of samples according to one-week and four-week groups (Figure 3-3C–D).

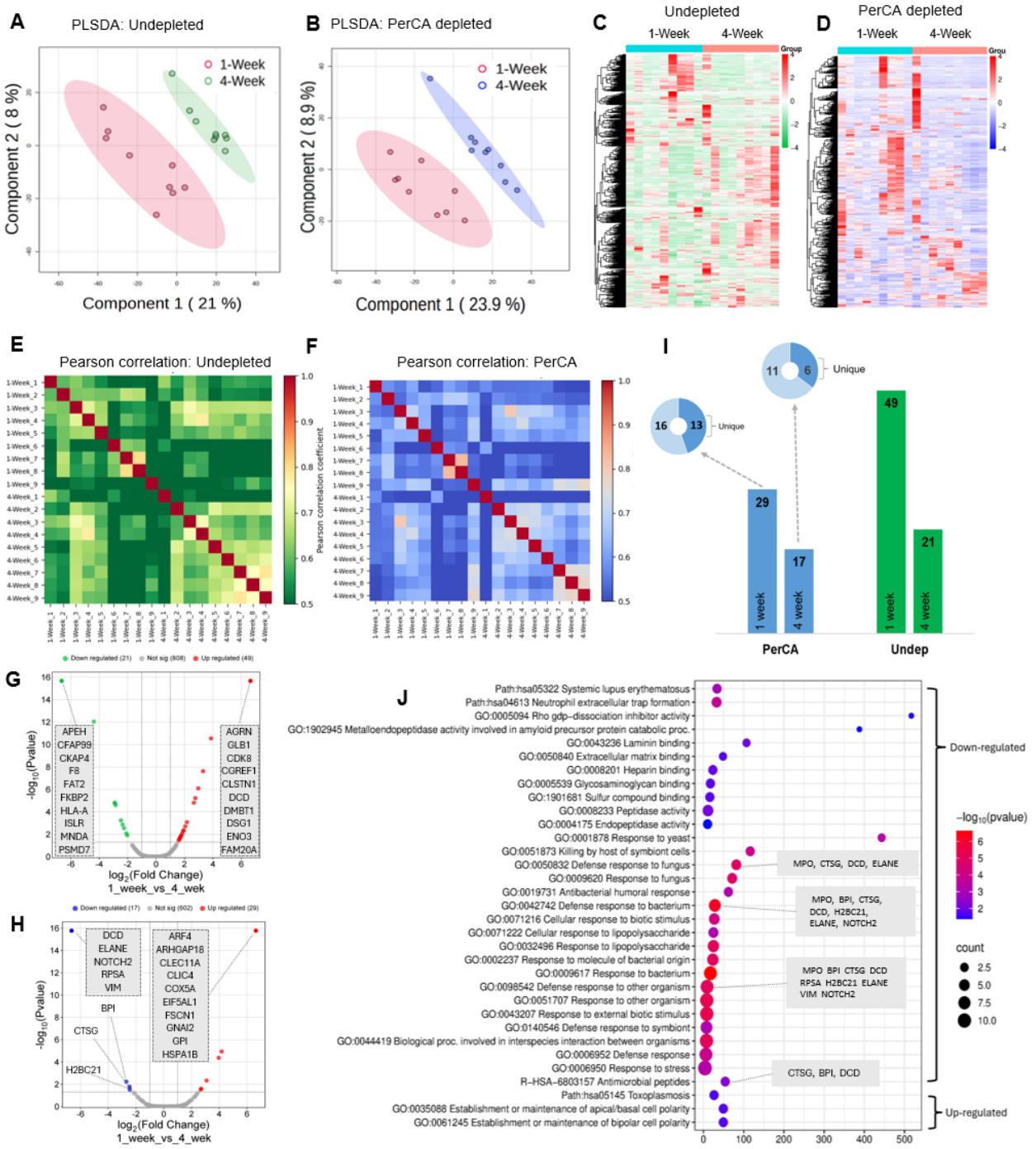


Figure 3-3. Multivariate analysis, correlation, and differential protein abundance in undepleted and PerCA-depleted human milk samples. (A) PLS-DA of undepleted samples showing distribution of one-week and four-week groups. (B) PLS-DA of PerCA-depleted samples showing distribution of one-week and four-week groups. (C) Heatmap with hierarchical clustering of protein abundance in undepleted samples. (D) Heatmap with hierarchical clustering of protein abundance in PerCA-depleted samples. (E) Pearson correlation matrix of undepleted samples. (F) Pearson correlation matrix of PerCA-depleted samples. (G) Volcano plot of undepleted samples comparing one-week versus four-week groups, with upregulated proteins (red, n=49), downregulated proteins

(green, n=21), and non-significant proteins (gray, n=808). (H) Volcano plot of PerCA-depleted samples comparing one-week versus four-week groups, with upregulated proteins (red, n=29), downregulated proteins (blue, n=17), and non-significant proteins (gray, n=602). (I) Bar plot showing the number of upregulated proteins at one-week and four-week stages in undepleted and PerCA-depleted datasets, with inset donut plots indicating uniquely regulated proteins. (J) Bubble plot of pathway enrichment analysis generated from significantly regulated proteins in the PerCA-depleted dataset, with bubble size representing protein count and color indicating log10 p-value.

Differential protein abundance analysis was performed to quantify stage-specific changes in protein composition. In the undepleted dataset, 49 proteins were significantly upregulated and 21 proteins were significantly downregulated at four weeks relative to one week, while 808 proteins were not significantly changed (Figure 3-3G). In the PerCA-depleted dataset, 29 proteins were significantly upregulated and 17 proteins were significantly downregulated, with 602 proteins showing no significant change (Figure 3-3H). Comparison of regulated proteins between workflows revealed limited overlap (Figure 3-3I), suggesting that the two preparation methods capture largely complementary portions of the regulated proteome, an important argument for applying both strategies in parallel. Complete lists of upregulated, downregulated, and uniquely expressed proteins are provided in Supplementary Data 3. Inset donut plots in Figure 3-3I highlight subsets of proteins uniquely regulated within each lactation stage in both PerCA and undepleted datasets; these uniquely regulated proteins are summarized in Table S3-1. Heatmaps generated after applying a 50% detection filter are shown in Figures S3-8B–C. Pathway enrichment analysis of differentially regulated proteins in the PerCA dataset revealed multiple enriched functional categories from integrated KEGG, GO, and Reactome annotations (Figure 3-3J).

3.4.4 Disease Association and Therapeutic Relevance of Stage-Specific Proteins

To further examine stage-specific protein expression in the PerCA-depleted dataset, selected proteins uniquely upregulated at each lactation stage were evaluated for their abundance patterns and disease associations (Figure 3-4). Bar plot analysis confirmed distinct stage-dependent abundance profiles for representative proteins. ELANE (neutrophil elastase), BPI (bactericidal/permeability-increasing protein), and MINDY1 (deubiquitinase MINDY-1) exhibited higher abundance in four-week samples, whereas MUC5B (mucin 5B), COX5A (cytochrome c oxidase subunit 5A), and LY6E (lymphocyte antigen 6E) showed higher abundance in one-week samples (Figure 3-4A). These patterns are consistent with the differential regulation identified in the PerCA dataset (Figure 3-3I; Table S3-1).

To characterize the biological context of these stage-specific proteins, disease association mapping was performed using Sankey diagrams. For proteins upregulated at four weeks, ELANE was mapped to disease categories using the TClin classification (Figure 3-4B), showing associations with cancer, immune/inflammatory, pulmonary, and other disease groups. ELANE is a serine protease produced by neutrophils with well-characterized roles in innate immune defense, including degradation of bacterial virulence factors and promotion of neutrophil extracellular trap formation. Its classification as a TClin drug target indicates the existence of approved small-molecule modulators; known compounds targeting ELANE are compiled in Table S3-2, which summarizes drugs,

primary mechanisms, and reported clinical or therapeutic uses, including sivelestat (a direct neutrophil elastase inhibitor), boceprevir, telaprevir, bortezomib, and quercetin.

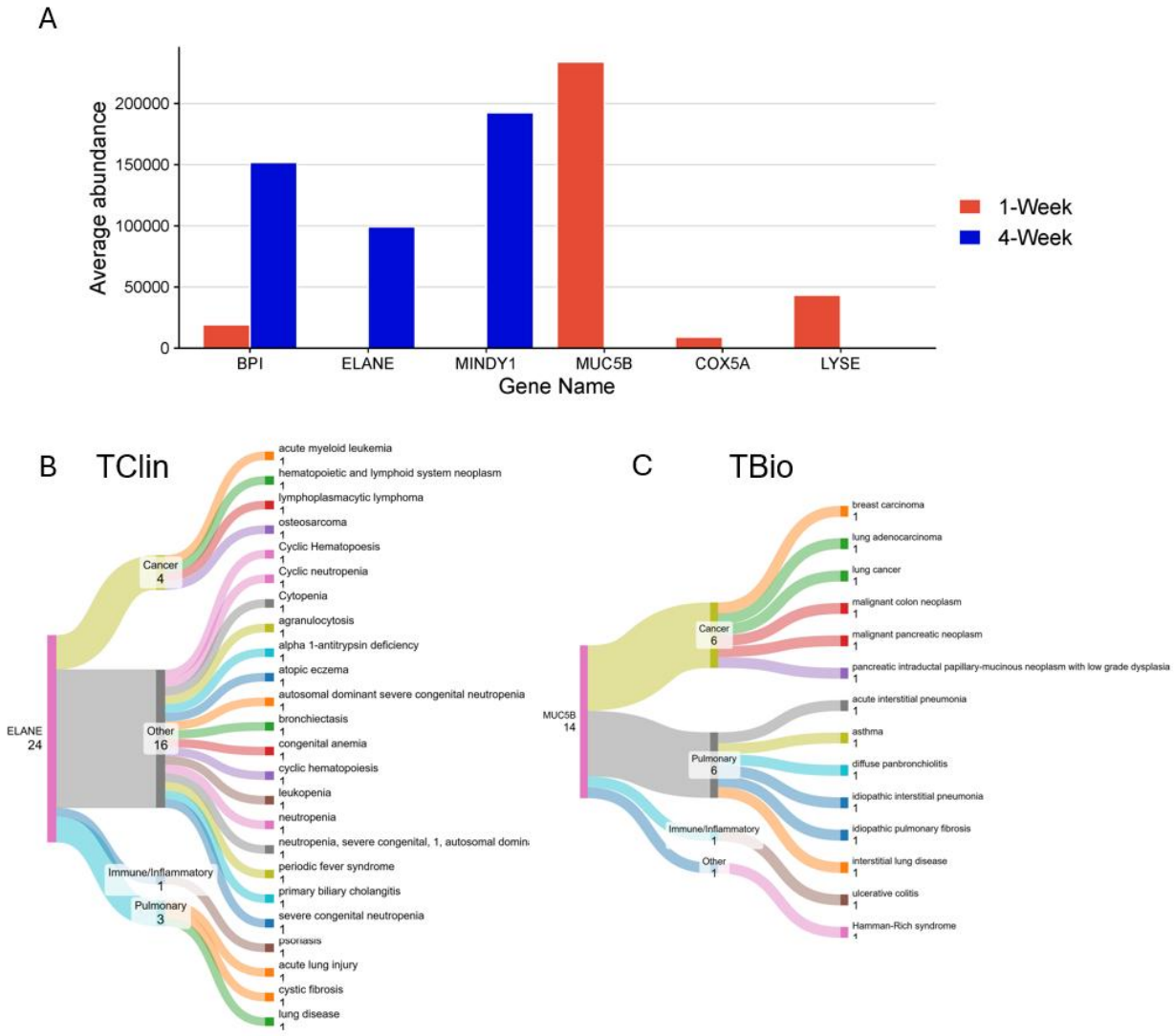


Figure 3-4. Stage-specific upregulated proteins and disease association analysis in the PerCA-depleted dataset. (A) Bar plot showing average abundance of selected proteins (BPI, ELANE, MINDY1, MUC5B, COX5A, and LY6E) in one-week (red) and four-week (blue) samples. (B) Sankey diagram illustrating disease associations of ELANE, a protein upregulated at four weeks and classified as a TClin target, with connections to cancer, immune/inflammatory, pulmonary, and other disease categories. (C) Sankey diagram illustrating disease associations of MUC5B, a protein upregulated at one week and classified as a TBio target, with connections to cancer, pulmonary, immune/inflammatory, and other disease categories.

For proteins upregulated at one week, MUC5B was analyzed using the TBio classification (Figure 3-4C), demonstrating associations across cancer, pulmonary, immune/inflammatory, and additional disease categories. MUC5B is a large secreted gel-forming mucin that forms the structural backbone of mucus layers in the gastrointestinal and respiratory tracts. Its enrichment in one-week preterm milk is consistent with the physiological priority of establishing gut barrier integrity in the immediate post-delivery period, when the extremely preterm intestine is most vulnerable to bacterial translocation and inflammatory injury. LY6E, a GPI-anchored protein with documented antiviral properties, and COX5A, a mitochondrial respiratory chain component involved in oxidative phosphorylation and cellular energy metabolism, were also elevated at one week, suggesting roles in early antiviral defense and metabolic support respectively. Together, these analyses highlight stage-specific protein signatures in the PerCA-depleted dataset and their mapping to established disease-associated categories, with selected proteins further connected to known therapeutic targets.

3.4.5 Elemental Variation Between One-Week and Four-Week Milk Samples

To evaluate elemental variation in human milk across lactation stages, ICP-MS analysis was performed on samples collected at one week and four weeks postpartum. Multivariate analysis of elemental profiles using PCA demonstrated separation between samples based on lactation stage (Figure 3-5A), confirming that the mineral composition of preterm milk undergoes systematic change across this early lactation window. Global patterns of elemental abundance were further visualized using a circular heatmap representation (Figure 3-5B), showing relative intensity differences across all measured elements between one-week and four-week samples.

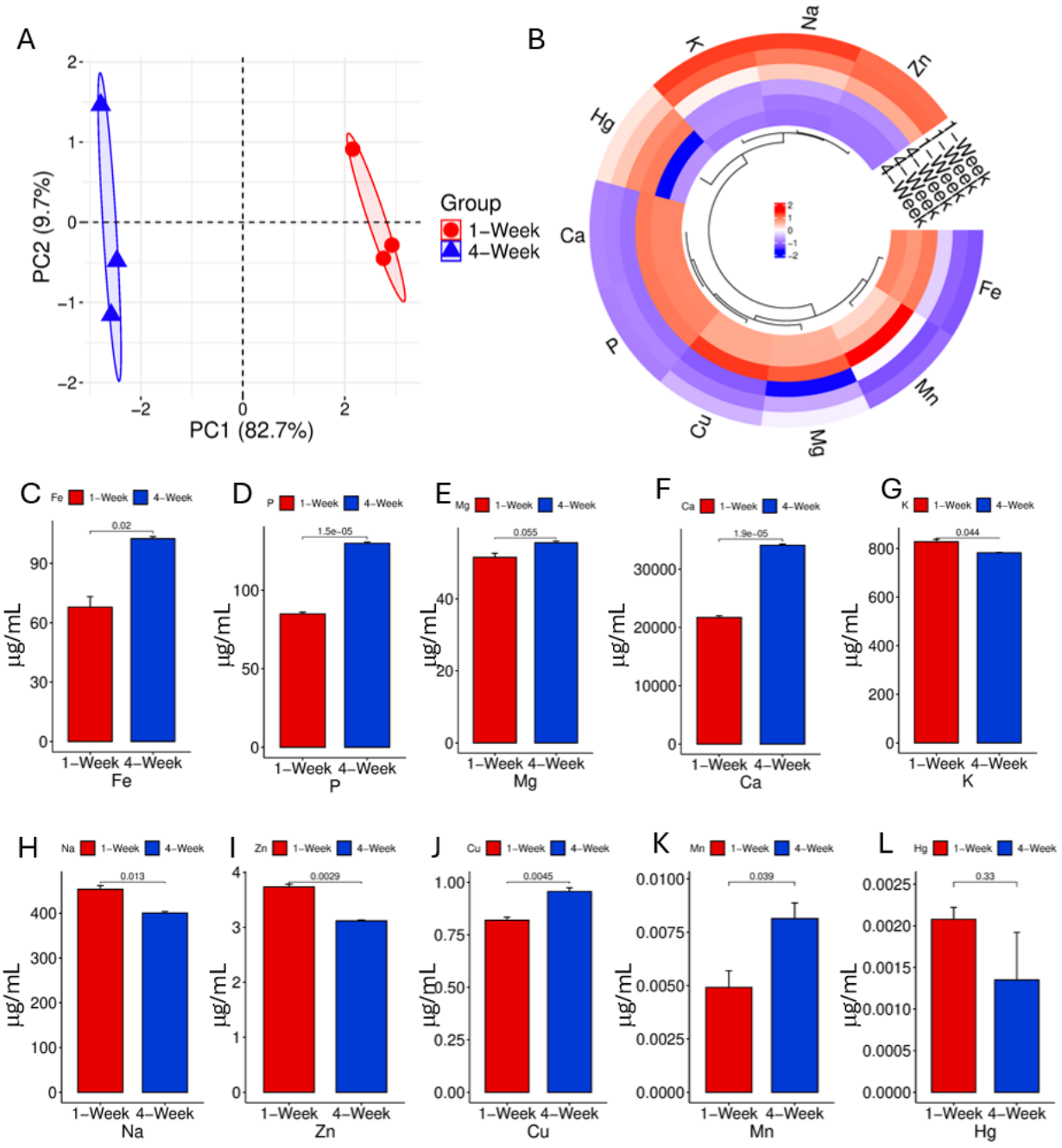


Figure 3-5. Elemental ionomics analysis of preterm human milk samples across lactation stages by ICP-MS. (A) PCA score plot showing distribution of one-week (red) and four-week (blue) samples based on elemental profiles. (B) Circular heatmap displaying relative abundance patterns of measured elements across samples with color scale representing normalized values. (C through L) Bar plots showing abundance of individual elements in one-week (red) and four-week (blue) samples: (C) Fe, (D) P, (E) Mg, (F) Ca, (G) K, (H) Na, (I) Zn, (J) Cu, (K) Mn, and (L) Hg. Error bars represent variability across samples and

p-values are indicated above each comparison. Elemental concentrations are reported in micrograms per milliliter (µg/mL).

Element-specific comparisons were performed to quantify differences in elemental composition between the two lactation stages. Bar plot analysis showed variation in multiple elements, including Fe, P, Mg, Ca, K, Na, Zn, Cu, Mn, and Hg (Figures 3-5C–L). Elemental concentrations were quantified in micrograms per milliliter (µg/mL) and compared across biological replicates for each group. Detailed elemental concentration values are provided in Supplementary Data 5.

Changes in iron (Fe) are of particular clinical relevance for preterm infants, who are at high risk of iron deficiency due to limited gestational iron accretion and high demand during rapid postnatal brain development. Changes in calcium (Ca) and phosphorus (P) have direct implications for supplementation strategies aimed at preventing metabolic bone disease, a recognized complication of very preterm birth. Zinc (Zn) and copper (Cu) are essential cofactors for immune function and enzymatic activity, and their temporal changes may reflect evolving nutritional requirements as the preterm infant matures. Mercury (Hg) was detectable at low levels in some samples; while concentrations were within ranges reported in prior human milk studies, its presence in donated preterm milk underscores the importance of ongoing donor screening and elemental monitoring in clinical milk bank practices.

3.5 Discussion

This study provides the first integrated proteomic and ionomic characterization of preterm human milk across early lactation, leveraging two complementary sample preparation strategies, undepleted and PerCA-depleted, to capture both the abundant

and low-abundance proteomes simultaneously. The use of paired samples from the same mothers at two postpartum time points enabled within-mother comparisons that control for the substantial inter-individual variability known to exist in human milk composition, and the integration of ICP-MS elemental analysis on the same samples allowed mineral dynamics to be characterized in direct parallel with protein dynamics.

The total of 1,043 proteins identified across all conditions represents a comprehensive characterization of the preterm milk proteome at this stage of lactation. The complementarity of the two preparation strategies is evident from the limited overlap in differentially regulated proteins between workflows, the undepleted dataset reveals changes in the abundant protein layer, while the PerCA dataset exposes the low-abundance proteome that would otherwise be masked. This observation directly validates the parallel strategy and supports the use of multiple depletion approaches when the goal is comprehensive proteome characterization rather than maximum total protein identification alone.

The identification of 153 proteins unique to this study relative to three recently published datasets expands the known preterm milk proteome and points toward biological components not previously catalogued. The functional clustering of these proteins into immunoglobulin-related and myosin/muscle-related interaction modules suggests that they represent genuine, organized biological entities rather than analytical artifacts. Future functional validation of selected proteins from this unique subset would be a logical next step.

The stage-specific protein patterns identified in the PerCA-depleted dataset are biologically plausible and clinically interpretable. The enrichment of ELANE and BPI at

four weeks is consistent with an evolving innate immune landscape in which the antimicrobial defense capacity of milk transitions from the immediate post-delivery immune surge characteristic of colostrum toward a more sustained, constitutively expressed innate immune profile. The enrichment of MUC5B at one week aligns with the established biological priority of establishing mucosal barrier function in the extremely preterm gut at the earliest possible opportunity after birth. LY6E elevation at one week is consistent with its known antiviral protective role, which may be most critical in the period of greatest immune vulnerability immediately following extremely preterm birth.

The ICP-MS findings provide a quantitative foundation for understanding mineral maturation in preterm milk that complements the proteomic data and carries independent clinical implications. The observed temporal changes in iron, zinc, calcium, and phosphorus across the first four weeks of lactation have direct relevance to NICU supplementation protocols that are currently based on generalized estimates rather than time-specific empirical data from preterm milk.

3.6 Conclusions

This chapter presented the first integrated proteomic and ionomic characterization of preterm human milk across early lactation. The application of parallel undepleted and PerCA-depleted sample preparations to paired samples from eight mothers of very preterm infants yielded a total of 1,043 protein identifications, clear lactation-stage-dependent proteome separation by multivariate analysis, and stage-specific protein signatures with disease associations and therapeutic target relevance. Proteins including ELANE, BPI, MUC5B, and LY6E carry functions directly relevant to innate immunity, antimicrobial defense, gut barrier integrity, and antiviral protection in the preterm infant,

functions of direct clinical importance during the NICU stay. Complementary ICP-MS analysis demonstrated parallel changes in mineral composition with clinical implications for NICU supplementation practices. Together with the methodological framework established in Chapter 2, these findings advance our understanding of preterm human milk as a dynamic biological system and establish a foundation for the personalized, evidence-based approach to human milk science discussed in Chapter 4.

3.7 Data Availability Statement

All LC-MS/MS RAW files and results files can be found at the MassIVE database (https://massive.ucsd.edu) under the following accessions: PerCA-depleted samples, MassIVE MSV000101302; Undepleted samples, MassIVE MSV000101301.

CHAPTER 4: CONCLUSIONS

4.1 Overview

The two studies presented in this dissertation were designed to work together. The first addressed a methodological problem that had to be solved before meaningful biological work could proceed, which protein depletion strategy best exposes the hidden proteome of human milk? The second used the answer to that question to pursue a clinically important biological goal, characterizing how the proteome and mineral composition of preterm human milk change across the early weeks of lactation. Looking back across both chapters, the progression from method to application is clear, and the findings from each study reinforce and extend the other in ways that are worth examining as a whole before discussing their implications and limitations, and before pointing toward where the work should go next.

4.2 Summary of Key Findings

4.2.1 Chapter 2: A Methodological Foundation for Human Milk Proteomics

The central finding of Chapter 2 was that the choice of abundant protein depletion strategy substantially shapes the protein landscape detected by LC-MS/MS in human milk, and that no single method is optimal for every analytical objective. The five methods evaluated, centrifugation, acetone precipitation, methanol-chloroform precipitation, commercial kit depletion, and perchloric acid (PerCA) precipitation, each performed differently across the five criteria assessed: total protein identification yield, high-abundance protein depletion efficiency, low-molecular-weight protein (LMWP) enrichment, unique protein identification, and cost per sample.

Methanol-chloroform precipitation identified the highest total number of proteins (468 protein groups, 2,668 peptides), making it the preferred choice when the objective is maximum proteome breadth. PerCA precipitation, by contrast, identified fewer proteins in total but outperformed all other methods for enriching unique and low-molecular-weight proteins. PerCA yielded the highest proportion of proteins below 26 kDa (37.5% of unique LMWPs), identified 93 proteins not detected by any other method, and achieved this at a cost of approximately one cent per sample, more than 300 times less expensive than the commercial kit alternative. Comparison of the combined dataset against four published human milk proteome databases identified 115 proteins not previously reported in the literature, including proteins with roles in neonatal brain development, innate immune defense, antiviral signaling, and maternal health.

These findings established PerCA as a validated, practical, and accessible depletion strategy for human milk proteomics. They also made a broader methodological point: in discovery-oriented proteomics of complex biological fluids, the method used to prepare the sample is not a neutral technical choice, it actively determines what biology is visible and what remains hidden. This insight directly shaped the experimental design of Chapter 3, where both undepleted and PerCA-depleted preparations were applied in parallel to ensure that neither the abundant nor the low-abundance proteome was systematically missed.

4.2.2 Chapter 3: Temporal Dynamics of the Preterm Milk Proteome and Ionome

The central finding of Chapter 3 was that the proteome and mineral composition of preterm human milk undergo systematic, biologically meaningful changes between one week and four weeks postpartum, and that these changes are detectable, reproducible,

and interpretable in functional terms. This study enrolled eight mothers of very preterm infants born at 22 to 31 weeks of gestational age and collected paired milk samples at the two time points, analyzing each sample by both label-free LC-MS/MS proteomics and ICP-MS elemental profiling.

Across all conditions, 1,043 proteins were identified in total. Multivariate analyses, PLS-DA and PCA, consistently separated one-week from four-week samples in both the undepleted and PerCA-depleted datasets, confirming that lactation stage is a primary driver of proteome variation. Differential abundance analysis identified 49 proteins upregulated and 21 downregulated at four weeks in undepleted samples, and 29 upregulated and 17 downregulated in the PerCA-depleted dataset. The limited overlap in regulated proteins between the two workflows confirmed that they sample largely non-overlapping portions of the regulated proteome, a direct validation of the parallel preparation strategy.

Stage-specific protein signatures in the PerCA-depleted dataset pointed toward biologically coherent transitions. ELANE, BPI, and MINDY1 were elevated at four weeks, suggesting an evolving innate immune and antimicrobial profile as lactation matures. MUC5B, LY6E, and COX5A were elevated at one week, pointing toward prioritization of mucosal barrier formation, antiviral defense, and mitochondrial energy support in the immediate post-delivery period. Cross-study comparison identified 153 proteins unique to this study, expanding the known preterm milk proteome. Disease association analysis using NIH Pharos and TCRD placed several of these proteins within established pharmacological and clinical frameworks, with ELANE identified as a TClin drug target with known therapeutic modulators.

ICP-MS analysis of ten elements, iron, phosphorus, magnesium, calcium, potassium, sodium, zinc, copper, manganese, and mercury, demonstrated PCA-based separation of one-week and four-week samples by elemental profile, confirming that the mineral composition of preterm milk also undergoes systematic temporal change. Individual element comparisons revealed quantifiable differences in biologically and clinically relevant minerals across the lactation window, with direct implications for NICU nutritional supplementation practices.

4.3 Integrated Interpretation

Taken together, these two studies tell a coherent story about human milk as a dynamic biological system, one whose complexity can be meaningfully accessed only when the right analytical tools are applied with the right methodological choices. Chapter 2 provided the lens. Chapter 3 used it to look at something that mattered clinically.

What emerges from Chapter 3 is a picture of preterm milk that is not static but actively remodeled across the early weeks of lactation in response to, or in anticipation of, the changing developmental needs of the preterm infant. The early enrichment of MUC5B supports gut barrier formation at the moment of greatest intestinal vulnerability. The early presence of LY6E may reflect maximal antiviral vigilance in a period when the immune system is least mature. The later elevation of ELANE and BPI suggests a shift toward a more constitutively antimicrobial immune profile as the acute post-delivery immune surge transitions to a more sustained pattern. These are not random fluctuations; they are biologically organized changes in a fluid whose composition is, at some level, being tuned to the infant's trajectory through the NICU.

The ICP-MS data add a mineral dimension to this picture that proteomics alone cannot provide. The fact that both the protein and mineral compositions of preterm milk change across the same lactation window, and that these changes can now be quantified simultaneously in the same samples, creates the opportunity for genuinely integrated analyses in future work. Whether changes in certain mineral concentrations are mechanistically related to changes in specific proteins, for example, whether iron dynamics are linked to the abundance of iron-binding or iron-transport proteins, is a question that the combined dataset is now positioned to address.

4.4 Implications for the Field

4.4.1 Methodological Implications

The validation of PerCA precipitation as a high-performance, cost-effective depletion strategy for human milk has practical significance that extends beyond this dissertation. The cost advantage, more than 300-fold relative to commercial kit alternatives, is particularly meaningful for large-scale biorepository-based studies, where per-sample costs accumulate significantly, and for researchers in lower-resource settings or lower-income countries who study human milk for public health purposes. The finding that PerCA accesses a distinct protein population that is largely invisible to other methods also makes a methodological argument that will be relevant to any future human milk proteomics study: applying only a single depletion approach risks systematically missing an entire layer of the proteome.

The parallel preparation strategy demonstrated in Chapter 3, analyzing the same samples with both undepleted and PerCA-depleted workflows, provides a practical template for future comprehensive milk proteomics studies. The limited overlap in

differentially regulated proteins between the two datasets in Chapter 3 demonstrates that this approach is not redundant but genuinely additive, and that the investment in running both preparations is justified by the biological information it returns.

4.4.2 Clinical Implications

From a clinical perspective, the findings of Chapter 3 have implications at several levels. The characterization of stage-specific protein signatures in preterm milk and the identification of proteins with roles in gut barrier formation, innate immunity, and antiviral defense at specific lactation time points contribute to the scientific basis for prioritizing the mother's own milk at specific stages of the NICU stay. If certain protective proteins peak in the first week and then decline, this argues for maximizing early enteral milk feeding initiation even at very small volumes. If others increase over time, this may be relevant to how donor milk, which is typically collected at later lactation stages and therefore may not reflect the early protein profile, is used as a substitute for a mother's own milk in the NICU.

The ICP-MS data provide an empirical, time-specific foundation for mineral supplementation protocols. Current NICU guidelines for calcium, phosphorus, iron, and zinc supplementation in preterm infants are based largely on estimates of fetal accretion rates and generalized data on milk mineral content, without precise time-resolved data on how preterm milk mineral concentrations change across lactation. The quantitative elemental data generated in this dissertation, while preliminary, point toward a more individualized and evidence-based approach to mineral supplementation for preterm infants.

The identification of ELANE as a TClin drug target elevated in four-week preterm milk raises an intriguing question for future translational work: does neutrophil elastase

activity in human milk contribute to the antimicrobial or gut barrier-modulating functions of the fluid? If so, this could have implications for how pasteurization protocols are designed, since heat treatment can denature enzymatically active proteins, and for how fortification strategies are developed to supplement the protein content of donated milk.

4.5 Limitations

Several limitations of the work presented in this dissertation deserve explicit acknowledgment.

Sample size and population. The preterm lactation study in Chapter 3 was a pilot study with eight mothers, a sample size chosen to establish feasibility and generate preliminary biological insights rather than to draw definitive conclusions. The statistical power of this study is limited, particularly for detecting modest differences in individual protein or element concentrations, and the findings should be considered hypothesis-generating rather than confirmatory. Future studies with larger and more diverse cohorts will be needed to validate the patterns identified here.

Single institution. Both studies drew samples from a single institution and geographic region, the University of Oklahoma and the Oklahoma Mothers Milk Bank. The findings may not fully generalize to populations with different genetic backgrounds, dietary patterns, environmental exposures, or clinical practices. Multi-site studies will be needed to assess the generalizability of these findings.

Single-pooled donor in Chapter 2. The depletion comparison study used a single pooled donor milk sample from the milk bank, which controlled for inter-individual variability as a confounder in the method comparison, but does not reflect the diversity of milk compositions that would be encountered in a clinical or research study. Validation of

PerCA performance across multiple individual donors would strengthen the methodological conclusions.

ICP-MS speciation. ICP-MS measures total elemental concentrations but does not distinguish between different chemical species or binding forms of a given element. For example, the total iron concentration measured includes iron bound to lactoferrin, free ionic iron, iron associated with casein micelles, and iron in other complexes, all of which may have different bioavailabilities in the infant gut. Speciation analysis would provide a more complete and clinically actionable picture of mineral bioavailability.

Temporal resolution. The Chapter 3 study compared two time points, one week and four weeks postpartum. While this captures a meaningful lactation window, it does not resolve the kinetics of compositional change within that window, and it does not characterize what happens beyond four weeks, during the extended NICU stay that many extremely preterm infants experience. A higher-resolution longitudinal study with additional time points would provide a more complete picture of preterm milk maturation.

4.6 Closing Remarks

Human milk is not simply food. It is a dynamic, biologically intelligent system that has been shaped over millions of years of evolution to meet the precise developmental needs of the human infant, and it carries information about the mother's physiological state that may one day serve as a non-invasive window into maternal health. For the preterm infant, born before this system has had its full gestational term to prepare, the provision of the mother's own milk is arguably the most impactful clinical intervention available, one whose full protective power we are only beginning to understand at the molecular level.

The work described in this dissertation is a contribution toward that understanding. The methods are practical and accessible. The findings are grounded in real biology. The questions they raise are worth pursuing. Every protein identified, every mineral quantified, and every methodological advance that makes these measurements more accessible to researchers around the world is a step toward a more complete picture of what human milk does, and how it can be better used to protect the most vulnerable infants we care for.

REFERENCES

1. Meek JY, Noble L, Section on Breastfeeding. Policy Statement: Breastfeeding and the Use of Human Milk. Pediatrics. 2022 Jun 27;150(1):e2022057988. doi:10.1542/peds.2022-057988

2. Victora CG, Bahl R, Barros AJD, França GVA, Horton S, Krasevec J, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. The Lancet. 2016 Jan 30;387(10017):475–90. doi:10.1016/S0140-6736(15)01024-7 PubMed PMID: 26869575.

3. Ballard O, Morrow AL. Human Milk Composition: Nutrients and Bioactive Factors. Pediatr Clin North Am. 2013 Feb;60(1):49–74. doi:10.1016/j.pcl.2012.10.002 PubMed PMID: 23178060; PubMed Central PMCID: PMC3586783.

4. Kim JW, Lee JS, Choi YJ, Kim C. The Multifaceted Functions of Lactoferrin in Antimicrobial Defense and Inflammation. Biomolecules. 2025 Aug 16;15(8):1174. doi:10.3390/biom15081174 PubMed PMID: 40867618; PubMed Central PMCID: PMC12384211.

5. Donald K, Petersen C, Turvey SE, Finlay BB, Azad MB. Secretory IgA: Linking microbes, maternal health, and infant health through human milk. Cell Host & Microbe. 2022 May 11;30(5):650–9. doi:10.1016/j.chom.2022.02.005

6. Zhu J, Dingess KA. The Functional Power of the Human Milk Proteome. Nutrients. 2019 Aug 8;11(8):1834. doi:10.3390/nu11081834 PubMed PMID: 31398857; PubMed Central PMCID: PMC6723708.

7. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002 Nov;1(11):845–67. doi:10.1074/mcp.r200007-mcp200 PubMed PMID: 12488461.

8. Rice SJ, Belani CP. Characterization of effective, simple, and low-cost precipitation methods for depleting abundant plasma proteins to enhance the depth and breadth of plasma proteomics. Proteomics. 2024 Aug;24(15):e2400071. doi:10.1002/pmic.202400071 PubMed PMID: 38700387.

9. Ahsan N, Fornelli L, Najar FZ, Gamagedara S, Hossan MR, Rao RSP, et al. Proteomics evaluation of five economical commercial abundant protein depletion kits for enrichment of diseases-specific biomarkers from blood serum. Proteomics. 2023 Oct;23(20):e2300150. doi:10.1002/pmic.202300150 PubMed PMID: 37199141; PubMed Central PMCID: PMC11166006.

10. Singh A, Peng Z, Shah SM, Salkini M, Mallah AF, Ali AF, et al. Enrichment of Low-Abundance and Low-Molecular-Weight Proteins in Human Milk Using Mass Spectrometry-Based Proteomics. ACS Omega. 2025 Sep 23;10(37):42507–17.

doi:10.1021/acsomega.5c03776 PubMed PMID: 41018635; PubMed Central PMCID: PMC12461359.

11. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR. Protein analysis by shotgun/bottom-up proteomics. Chem Rev. 2013 Apr 10;113(4):2343–94. doi:10.1021/cr3003533 PubMed PMID: 23438204; PubMed Central PMCID: PMC3751594.

12. Megger DA, Bracht T, Meyer HE, Sitek B. Label-free quantification in clinical proteomics. Biochim Biophys Acta. 2013 Aug;1834(8):1581–90. doi:10.1016/j.bbapap.2013.04.001 PubMed PMID: 23567906.

13. Sabatier M, Garcia-Rodenas CL, Castro CAD, Kastenmayer P, Vigo M, Dubascoux S, et al. Longitudinal Changes of Mineral Concentrations in Preterm and Term Human Milk from Lactating Swiss Women. Nutrients. 2019 Aug 9;11(8):1855. doi:10.3390/nu11081855 PubMed PMID: 31405051; PubMed Central PMCID: PMC6723843.

14. Balaram V. Strategies to overcome interferences in elemental and isotopic geochemical analysis by quadrupole inductively coupled plasma mass spectrometry: A critical evaluation of the recent developments. Rapid Commun Mass Spectrom. 2021 May 30;35(10):e9065. doi:10.1002/rcm.9065 PubMed PMID: 33587758.

15. Organization WH. Born too soon: decade of action on preterm birth. World Health Organization; 2023. 122 p.

16. Underwood MA. Human milk for the premature infant. Pediatr Clin North Am. 2013 Feb;60(1):189–207. doi:10.1016/j.pcl.2012.09.008 PubMed PMID: 23178065; PubMed Central PMCID: PMC3508468.

17. Neu J, Walker WA. Necrotizing enterocolitis. N Engl J Med. 2011 Jan 20;364(3):255–64. doi:10.1056/NEJMra1005408 PubMed PMID: 21247316; PubMed Central PMCID: PMC3628622.

18. Meinzen-Derr J, Poindexter B, Wrage L, Morrow AL, Stoll B, Donovan EF. Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death. J Perinatol. 2009 Jan;29(1):57–62. doi:10.1038/jp.2008.117 PubMed PMID: 18716628; PubMed Central PMCID: PMC2801431.

19. Miketinas DC, Patterson MA, Bender TM, Kinnaman JN, Carter DL, Meredith NA, et al. Protein and Free Amino Acid Composition of Preterm Human Milk: A Systematic Review and Meta-analysis. Adv Nutr. 2025 Jun;16(6):100432. doi:10.1016/j.advnut.2025.100432 PubMed PMID: 40288492; PubMed Central PMCID: PMC12149428.

20. Mól N, Zasada M, Suski M, Zasada W, Kwinta P. Time-Dependent Differences in the Human Milk Proteome After Preterm Birth: A Paired Two-Stage Proteomic

Study. Nutrients. 2026 Mar 5;18(5):848. doi:10.3390/nu18050848 PubMed PMID: 41830017; PubMed Central PMCID: PMC12987234.

21. Grapov D, Lemay DG, Weber D, Phinney BS, Azulay Chertok IR, Gho DS, et al. The Human Colostrum Whey Proteome Is Altered in Gestational Diabetes Mellitus. J Proteome Res. 2015 Jan 2;14(1):512–20. doi:10.1021/pr500818d

22. Hahn WH, Bae SP, Lee H, Park JM, Park S, Lee J, et al. The impact of freeze-drying on the glycoproteomic profiles of human milk [Internet]. 2020 Aug. doi:10.5806/AST.2020.33.4.177

23. Newburg DS, Walker WA. Protection of the Neonate by the Innate Immune System of Developing Gut and of Human Milk. Pediatr Res. 2007 Jan;61(1):2–8. doi:10.1203/01.pdr.0000250274.68571.18

24. van Odijk J, Kull I, Borres MP, Brandtzaeg P, Edberg U, Hanson LÅ, et al. Breastfeeding and allergic disease: a multidisciplinary review of the literature (1966–2001) on the mode of early feeding in infancy and its impact on later atopic manifestations. Allergy. 2003;58(9):833–43. doi:10.1034/j.1398-9995.2003.00264.x

25. França EL, Calderon I de MP, Vieira EL, Morceli G, Honorio-França AC. Transfer of Maternal Immunity to Newborns of Diabetic Mothers. Journal of Immunology Research. 2012;2012(1):928187. doi:10.1155/2012/928187

26. Morceli G, França E, Magalhães V, Damasceno D, Calderon I, Honorio-França A. Diabetes induced immunological and biochemical changes in human colostrum. Acta Paediatrica. 2011;100(4):550–6. doi:10.1111/j.1651-2227.2010.02070.x

27. Aslebagh R, Channaveerappa D, Arcaro KF, Darie CC. Comparative two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of human milk to identify dysregulated proteins in breast cancer. ELECTROPHORESIS. 2018;39(14):1723–34. doi:10.1002/elps.201800025

28. Chang JC, Chen CH, Fang LJ, Tsai CR, Chang YC, Wang TM. Influence of Prolonged Storage Process, Pasteurization, and Heat Treatment on Biologically-active Human Milk Proteins. Pediatrics & Neonatology. 2013 Dec 1;54(6):360–6. doi:10.1016/j.pedneo.2013.03.018

29. Liang N, Koh J, Kim BJ, Ozturk G, Barile D, Dallas DC. Structural and functional changes of bioactive proteins in donor human milk treated by vat-pasteurization, retort sterilization, ultra-high-temperature sterilization, freeze-thawing and homogenization. Front Nutr. 2022;9:926814. doi:10.3389/fnut.2022.926814 PubMed PMID: 36185694; PubMed Central PMCID: PMC9521613.

30. Gharbi N, Marciniak A, Perreault V, Stone D, Fittipaldi N, Unger S, et al. The effect of pasteurization treatment and skimming conditions on human milk proteins. LWT. 2022 Dec 30;172:114184. doi:10.1016/j.lwt.2022.114184

31. Thesbjerg MN, Sundekilde UK, Poulsen NA, Larsen LB, Nielsen SDH. A novel proteomic approach for the identification and relative quantification of disulfide-bridges in the human milk proteome. J Proteomics. 2024 Jun 15;301:105194. doi:10.1016/j.jprot.2024.105194 PubMed PMID: 38723850.

32. Park YW, Nam MS. Bioactive Peptides in Milk and Dairy Products: A Review. Korean J Food Sci Anim Resour. 2015;35(6):831–40. doi:10.5851/kosfa.2015.35.6.831 PubMed PMID: 26877644; PubMed Central PMCID: PMC4726964.

33. Aksan A, Erdal I, Yalcin SS, Stein J, Samur G. Osteopontin Levels in Human Milk Are Related to Maternal Nutrition and Infant Health and Growth. Nutrients. 2021 Jul 31;13(8):2670. doi:10.3390/nu13082670 PubMed PMID: 34444830; PubMed Central PMCID: PMC8402120.

34. Polaskova V, Kapur A, Khan A, Molloy MP, Baker MS. High-abundance protein depletion: comparison of methods for human plasma biomarker discovery. Electrophoresis. 2010 Jan;31(3):471–82. doi:10.1002/elps.200900286 PubMed PMID: 20119956.

35. Zhang H, Yi EC, Li X jun, Mallick P, Kelly-Spratt KS, Masselon CD, et al. High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. Mol Cell Proteomics. 2005 Feb;4(2):144–55. doi:10.1074/mcp.M400090-MCP200 PubMed PMID: 15608340.

36. Roncada P, Stipetic LH, Bonizzi L, Burchmore RJS, Kennedy MW. Proteomics as a tool to explore human milk in health and disease. J Proteomics. 2013 Aug 2;88:47–57. doi:10.1016/j.jprot.2013.04.008 PubMed PMID: 23590890.

37. Auzino B, Miranda G, Henry C, Krupova Z, Martini M, Salari F, et al. Top-Down proteomics based on LC-MS combined with cDNA sequencing to characterize multiple proteoforms of Amiata donkey milk proteins. Food Res Int. 2022 Oct;160:111611. doi:10.1016/j.foodres.2022.111611 PubMed PMID: 36076434.

38. Vincent D, Mertens D, Rochfort S. Optimisation of Milk Protein Top-Down Sequencing Using In-Source Collision-Induced Dissociation in the Maxis Quadrupole Time-of-Flight Mass Spectrometer. Molecules. 2018 Oct 26;23(11):2777. doi:10.3390/molecules23112777 PubMed PMID: 30373172; PubMed Central PMCID: PMC6278275.

39. Roncada P, Piras C, Soggiu A, Turk R, Urbani A, Bonizzi L. Farm animal milk proteomics. J Proteomics. 2012 Jul 19;75(14):4259–74. doi:10.1016/j.jprot.2012.05.028 PubMed PMID: 22641156.

40. Bielicki J, Huch R, von Mandach U. Time-course of leptin levels in term and preterm human milk. Eur J Endocrinol. 2004 Aug;151(2):271–6. doi:10.1530/eje.0.1510271 PubMed PMID: 15296484.

41. Berven FS, Kroksveen AC, Berle M, Rajalahti T, Flikka K, Arneberg R, et al. Pre-analytical influence on the low molecular weight cerebrospinal fluid proteome. Proteomics Clin Appl. 2007 Jul;1(7):699–711. doi:10.1002/prca.200700126 PubMed PMID: 21136725.

42. Peng Z, Munige SW, Bhusal D, Yang IL, Yang Z, Ahsan N. A cross-species proteomic assessment of cost-effective platforms for depleting high-abundant proteins from blood serum. BMC Methods. 2025;2(1):21. doi:10.1186/s44330-025-00042-7 PubMed PMID: 40989151; PubMed Central PMCID: PMC12450807.

43. Mangé A, Bellet V, Tuaillon E, Van de Perre P, Solassol J. Comprehensive proteomic analysis of the human milk proteome: contribution of protein fractionation. J Chromatogr B Analyt Technol Biomed Life Sci. 2008 Dec 15;876(2):252–6. doi:10.1016/j.jchromb.2008.11.003 PubMed PMID: 19019742.

44. Vincent D, Ezernieks V, Elkins A, Nguyen N, Moate PJ, Cocks BG, et al. Milk Bottom-Up Proteomics: Method Optimization. Front Genet. 2015;6:360. doi:10.3389/fgene.2015.00360 PubMed PMID: 26793233; PubMed Central PMCID: PMC4707256.

45. Blans K, Hansen MS, Sørensen LV, Hvam ML, Howard KA, Möller A, et al. Pellet-free isolation of human and bovine milk extracellular vesicles by size-exclusion chromatography. J Extracell Vesicles. 2017;6(1):1294340. doi:10.1080/20013078.2017.1294340 PubMed PMID: 28386391; PubMed Central PMCID: PMC5373680.

46. Fernández C, Santos HM, Ruíz-Romero C, Blanco FJ, Capelo-Martínez JL. A comparison of depletion versus equalization for reducing high-abundance proteins in human serum. Electrophoresis. 2011 Nov;32(21):2966–74. doi:10.1002/elps.201100183 PubMed PMID: 21997478.

47. Dekker PM, Boeren S, van Goudoever JB, Vervoort JJM, Hettinga KA. Exploring Human Milk Dynamics: Interindividual Variation in Milk Proteome, Peptidome, and Metabolome. J Proteome Res. 2022 Apr 1;21(4):1002–16. doi:10.1021/acs.jproteome.1c00879 PubMed PMID: 35104145; PubMed Central PMCID: PMC8981310.

48. Puopolo T, Seeram NP, Liu C. Chloroform/Methanol Protein Extraction and In-solution Trypsin Digestion Protocol for Bottom-up Proteomics Analysis. Bio Protoc. 2024 Aug 20;14(16):e5055. doi:10.21769/BioProtoc.5055 PubMed PMID: 39210950; PubMed Central PMCID: PMC11349489.

49. Viode A, van Zalm P, Smolen KK, Fatou B, Stevenson D, Jha M, et al. A simple, time- and cost-effective, high-throughput depletion strategy for deep plasma proteomics. Sci Adv. 2023;9(13):eadf9717. doi:10.1126/sciadv.adf9717 PubMed PMID: 36989362; PubMed Central PMCID: PMC10058233.

50. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020 Apr 15;36(8):2628–9. doi:10.1093/bioinformatics/btz931

51. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics. 2015 May 22;16(1):169. doi:10.1186/s12859-015-0611-3 PubMed PMID: 25994840; PubMed Central PMCID: PMC4455604.

52. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2022 Nov 12;51(D1):D638–46. doi:10.1093/nar/gkac1000 PubMed PMID: 36370105; PubMed Central PMCID: PMC9825434.

53. da Costa JP, Santos PSM, Vitorino R, Rocha-Santos T, Duarte AC. How low can you go? A current perspective on low-abundance proteomics. TrAC Trends in Analytical Chemistry. 2017 Aug 1;93:171–82. doi:10.1016/j.trac.2017.05.014

54. van Herwijnen MJC, Zonneveld MI, Goerdayal S, Nolte-'t Hoen ENM, Garssen J, Stahl B, et al. Comprehensive Proteomic Analysis of Human Milk-derived Extracellular Vesicles Unveils a Novel Functional Proteome Distinct from Other Milk Components. Mol Cell Proteomics. 2016 Nov;15(11):3412–23. doi:10.1074/mcp.M116.060426 PubMed PMID: 27601599; PubMed Central PMCID: PMC5098039.

55. Dayon L, Macron C, Lahrichi S, Núñez Galindo A, Affolter M. Proteomics of Human Milk: Definition of a Discovery Workflow for Clinical Research Studies. J Proteome Res. 2021 May 7;20(5):2283–90. doi:10.1021/acs.jproteome.0c00816 PubMed PMID: 33769819.

56. Qu J, Zhang L, Yin L, Liu J, Sun Z, Zhou P. Changes in bioactive proteins and serum proteome of human milk under different frozen storage. Food Chem. 2021 Aug 1;352:129436. doi:10.1016/j.foodchem.2021.129436 PubMed PMID: 33691214.

57. Sugisawa H, Itou T, Saito M, Moritomo T, Miura Y, Sakai T. A low-molecular-weight fraction of bovine colostrum and milk enhances the oxidative burst activity of polymorphonuclear leukocytes. Vet Res Commun. 2003 Sep;27(6):453–61. doi:10.1023/a:1025737623792 PubMed PMID: 14582744.

58. Fox PF, Uniacke-Lowe T, McSweeney PLH, O'Mahony JA. Milk Proteins. In: Fox PF, Uniacke-Lowe T, McSweeney PLH, O'Mahony JA, editors. Dairy Chemistry and Biochemistry [Internet]. Cham: Springer International Publishing; 2015 [cited 2026 May 9]. p. 145–239. Available from: https://doi.org/10.1007/978-3-319-14892-2_4 doi:10.1007/978-3-319-14892-2_4

59. Stastna M. Advances in separation and identification of biologically important milk proteins and peptides. Electrophoresis. 2024 Jan;45(1–2):101–19. doi:10.1002/elps.202300084 PubMed PMID: 37289082.

60. Das L, Murthy V, Varma AK. Comprehensive Analysis of Low Molecular Weight Serum Proteome Enrichment for Mass Spectrometric Studies. ACS Omega. 2020 Nov 10;5(44):28877–88. doi:10.1021/acsomega.0c04568 PubMed PMID: 33195941; PubMed Central PMCID: PMC7659158.

61. Skinner OS, Haverland NA, Fornelli L, Melani RD, Do Vale LHF, Seckler HS, et al. Top-down characterization of endogenous protein complexes with native proteomics. Nat Chem Biol. 2018 Jan;14(1):36–41. doi:10.1038/nchembio.2515 PubMed PMID: 29131144; PubMed Central PMCID: PMC5726920.

62. Kline JT, Belford MW, Boeser CL, Huguet R, Fellers RT, Greer JB, et al. Orbitrap Mass Spectrometry and High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) Enable the in-Depth Analysis of Human Serum Proteoforms. J Proteome Res. 2023 Nov 3;22(11):3418–26. doi:10.1021/acs.jproteome.3c00488 PubMed PMID: 37774690; PubMed Central PMCID: PMC10629265.

63. Wang C, Lu Y, Hu J, Yang Y, Cheng J, Jiang S, et al. Comparative proteomics of human milk casein fraction collected from women of Korean and Han ethnic groups in China. Front Nutr. 2023;10:1078355. doi:10.3389/fnut.2023.1078355 PubMed PMID: 36755990; PubMed Central PMCID: PMC9901581.

64. Zhu J, Garrigues L, Van den Toorn H, Stahl B, Heck AJR. Discovery and Quantification of Nonhuman Proteins in Human Milk. J Proteome Res. 2019 Jan 4;18(1):225–38. doi:10.1021/acs.jproteome.8b00550 PubMed PMID: 30489082; PubMed Central PMCID: PMC6326037.

65. Viitaharju J, Polari L, Kauko O, Merilahti J, Rokka A, Toivola DM, et al. Improved breast milk proteome coverage by DIA based LC-MS/MS method. Proteomics. 2024 Jul;24(14):e2300340. doi:10.1002/pmic.202300340 PubMed PMID: 38873899.

66. Gao X, McMahon RJ, Woo JG, Davidson BS, Morrow AL, Zhang Q. Temporal changes in milk proteomes reveal developing milk functions. J Proteome Res. 2012 Jul 6;11(7):3897–907. doi:10.1021/pr3004002 PubMed PMID: 22676802.

67. Paditz E. Postnatal Development of the Circadian Rhythmicity of Human Pineal Melatonin Synthesis and Secretion (Systematic Review). Children (Basel). 2024 Sep 29;11(10):1197. doi:10.3390/children11101197 PubMed PMID: 39457162; PubMed Central PMCID: PMC11506472.

68. Espinoza J, Chaiworapongsa T, Romero R, Edwin S, Rathnasabapathy C, Gomez R, et al. Antimicrobial peptides in amniotic fluid: defensins, calprotectin and bacterial/permeability-increasing protein in patients with microbial invasion of the amniotic cavity, intra-amniotic inflammation, preterm labor and premature rupture of

membranes. J Matern Fetal Neonatal Med. 2003 Jan;13(1):2–21. doi:10.1080/jmf.13.1.2.21 PubMed PMID: 12710851.

69. Good M, Siggers RH, Sodhi CP, Afrazi A, Alkhudari F, Egan CE, et al. Amniotic fluid inhibits Toll-like receptor 4 signaling in the fetal and neonatal intestinal epithelium. Proc Natl Acad Sci U S A. 2012 Jul 10;109(28):11330–5. doi:10.1073/pnas.1200856109 PubMed PMID: 22733781; PubMed Central PMCID: PMC3396489.

70. Šket T, Ramuta TŽ, Starčič Erjavec M, Kreft ME. The Role of Innate Immune System in the Human Amniotic Membrane and Human Amniotic Fluid in Protection Against Intra-Amniotic Infections and Inflammation. Front Immunol. 2021;12:735324. doi:10.3389/fimmu.2021.735324 PubMed PMID: 34745106; PubMed Central PMCID: PMC8566738.

71. Hannan S, Hami I, Dey RK, Gupta SD. A Systematic Exploration of Key Candidate Genes and Pathways in the Biogenesis of Human Gastric Cancer: A Comprehensive Bioinformatics Investigation. Journal of Translational Gastroenterology. 2024 Mar 30;2(1):9–20. doi:10.14218/JTG.2023.00072

72. Jia J, Han Z, Wang X, Zheng X, Wang S, Cui Y. H2B gene family: A prognostic biomarker and correlates with immune infiltration in glioma. Front Oncol. 2022;12:966817. doi:10.3389/fonc.2022.966817 PubMed PMID: 36387186; PubMed Central PMCID: PMC9641242.

73. Tiwari S, Singh A, Gupta P, Singh S. UBA52 Is Crucial in HSP90 Ubiquitylation and Neurodegenerative Signaling during Early Phase of Parkinson's Disease. Cells. 2022 Nov 25;11(23):3770. doi:10.3390/cells11233770 PubMed PMID: 36497031; PubMed Central PMCID: PMC9738938.

74. Kobayashi M, Oshima S, Maeyashiki C, Nibe Y, Otsubo K, Matsuzawa Y, et al. The ubiquitin hybrid gene UBA52 regulates ubiquitination of ribosome and sustains embryonic development. Sci Rep. 2016 Nov 10;6:36780. doi:10.1038/srep36780 PubMed PMID: 27829658; PubMed Central PMCID: PMC5103194.

75. Lee SO, Kelliher JL, Song W, Tengler K, Sarkar A, Dray E, et al. UBA80 and UBA52 fine-tune RNF168-dependent histone ubiquitination and DNA repair. J Biol Chem. 2023 Aug;299(8):105043. doi:10.1016/j.jbc.2023.105043 PubMed PMID: 37451480; PubMed Central PMCID: PMC10413357.

76. Lisco A, Ye P, Wong CS, Pei L, Hsu AP, Mace EM, et al. Lost in Translation: Lack of CD4 Expression due to a Novel Genetic Defect. J Infect Dis. 2021 Feb 24;223(4):645–54. doi:10.1093/infdis/jiab025 PubMed PMID: 33471124; PubMed Central PMCID: PMC7904290.

77. Lei L, Zhang X, Yang X, Su Y, Liu H, Yang H, et al. A Genetic Model Reveals Biological Features of Neonatal CD4 Helper Cells Undergone Homeostasis in Mice.

Front Cell Dev Biol. 2021;9:659744. doi:10.3389/fcell.2021.659744 PubMed PMID: 33777965; PubMed Central PMCID: PMC7991801.

78. Wu SR, Butts JC, Caudill MS, Revelli JP, Dhindsa RS, Durham MA, et al. Atoh1 drives the heterogeneity of the pontine nuclei neurons and promotes their differentiation. Sci Adv. 2023 Jun 30;9(26):eadg1671. doi:10.1126/sciadv.adg1671 PubMed PMID: 37390208; PubMed Central PMCID: PMC10313176.

79. Elliott KL, Pavlinkova G, Chizhikov VV, Yamoah EN, Fritzsch B. Neurog1, Neurod1, and Atoh1 are essential for spiral ganglia, cochlear nuclei, and cochlear hair cell development. Fac Rev. 2021;10:47. doi:10.12703/r/10-47 PubMed PMID: 34131657; PubMed Central PMCID: PMC8170689.

80. Poudel PP, Bhattarai C, Ghosh A, Kalthur SG. Expression of ATOH1 gene and activated signaling pathways for the neurogenesis of cerebellar granule cells: A review. Neuroscience Research Notes. 2022 Apr 28;5(2):125–125. doi:10.31117/neuroscirn.v5i2.125

81. Wu J, You Q, Lyu R, Qian Y, Tao H, Zhang F, et al. Folate metabolism negatively regulates OAS-mediated antiviral innate immunity via ADAR3/endogenous dsRNA pathway. Metabolism. 2023 Jun;143:155526. doi:10.1016/j.metabol.2023.155526 PubMed PMID: 36822494.

82. Drakul M, Čolić M. Immunomodulatory activity of dipeptidyl peptidase-4 inhibitors in immune-related diseases. Eur J Immunol. 2023 Dec;53(12):e2250302. doi:10.1002/eji.202250302 PubMed PMID: 37732495.

83. Huang J, Liu X, Wei Y, Li X, Gao S, Dong L, et al. Emerging Role of Dipeptidyl Peptidase-4 in Autoimmune Disease. Front Immunol. 2022;13:830863. doi:10.3389/fimmu.2022.830863 PubMed PMID: 35309368; PubMed Central PMCID: PMC8931313.

84. Heil LR, Damoc E, Arrey TN, Pashkova A, Denisov E, Petzoldt J, et al. Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. J Proteome Res. 2023 Oct 6;22(10):3290–300. doi:10.1021/acs.jproteome.3c00357 PubMed PMID: 37683181; PubMed Central PMCID: PMC10563156.

85. Greguš M, Koller A, Ray S, Ivanov AR. Improved Data Acquisition Settings on Q Exactive HF-X and Fusion Lumos Tribrid Orbitrap-Based Mass Spectrometers for Proteomic Analysis of Limited Samples. J Proteome Res. 2024 Jun 7;23(6):2230–40. doi:10.1021/acs.jproteome.4c00181 PubMed PMID: 38690845; PubMed Central PMCID: PMC11165581.

86. Ahsan N, Belmont J, Chen Z, Clifton JG, Salomon AR. Highly reproducible improved label-free quantitative analysis of cellular phosphoproteome by optimization of LC-MS/MS gradient and analytical column construction. J Proteomics. 2017 Aug

8;165:69–74. doi:10.1016/j.jprot.2017.06.013 PubMed PMID: 28634120; PubMed Central PMCID: PMC5542054.

87. Parker MG, Greenberg LT, Edwards EM, Ehret D, Belfort MB, Horbar JD. National Trends in the Provision of Human Milk at Hospital Discharge Among Very Low-Birth-Weight Infants. JAMA Pediatr. 2019 Oct 1;173(10):961–8. doi:10.1001/jamapediatrics.2019.2645

88. Hair AB, Peluso AM, Hawthorne KM, Perez J, Smith DP, Khan JY, et al. Beyond Necrotizing Enterocolitis Prevention: Improving Outcomes with an Exclusive Human Milk–Based Diet. Breastfeeding Medicine. 2016 Mar 1;11(2):70–4. doi:10.1089/bfm.2015.0134

89. Okamoto T, Shirai M, Kokubo M, Takahashi S, Kajino M, Takase M, et al. Human milk reduces the risk of retinal detachment in extremely low-birthweight infants [Internet]. [cited 2026 May 11]. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1442-200X.2007.02483.x

90. Vohr BR, Poindexter BB, Dusick AM, McKinley LT, Higgins RD, Langer JC, et al. Persistent Beneficial Effects of Breast Milk Ingested in the Neonatal Intensive Care Unit on Outcomes of Extremely Low Birth Weight Infants at 30 Months of Age. Pediatrics. 2007 Oct 1;120(4):e953–9. doi:10.1542/peds.2006-3227

91. Chaaban H, Burge K, McElroy SJ. Evolutionary bridges: how factors present in amniotic fluid and human milk help mature the gut. J Perinatol. 2024 Nov;44(11):1552–9. doi:10.1038/s41372-024-02026-x

92. Dallas DC, Smink CJ, Robinson RC, Tian T, Guerrero A, Parker EA, et al. Endogenous human milk peptide release is greater after preterm birth than term birth. J Nutr. 2015 Mar;145(3):425–33. doi:10.3945/jn.114.203646 PubMed PMID: 25540406; PubMed Central PMCID: PMC4336528.

93. Shuken SR. An Introduction to Mass Spectrometry-Based Proteomics. J Proteome Res. 2023 Jul 7;22(7):2151–71. doi:10.1021/acs.jproteome.2c00838 PubMed PMID: 37260118.

94. Zhu J, Dingess KA, Mank M, Stahl B, Heck AJR. Personalized Profiling Reveals Donor- and Lactation-Specific Trends in the Human Milk Proteome and Peptidome. J Nutr. 2021 Apr 8;151(4):826–39. doi:10.1093/jn/nxaa445 PubMed PMID: 33693758; PubMed Central PMCID: PMC8030701.

95. Seitz T, Viitaharju J, Benchraka C, Merilahti J, Kalliomäki M, Polari L, et al. Breast Milk Proteome: Changes in the Different Stages of Lactation and Impacts of Gestational Diabetes Mellitus and Body Mass Index. Mol Nutr Food Res. 2025 Nov;69(22):e70225. doi:10.1002/mnfr.70225 PubMed PMID: 40891102; PubMed Central PMCID: PMC12643191.

96. Mu Q, Luo Y, Li W, Guo Y, Zeng X, Li J, et al. Temporal Dynamics of the Milk Proteome in Women with Gestational Diabetes Mellitus. J Proteome Res. 2026 Jan 2;25(1):131–44. doi:10.1021/acs.jproteome.5c00503 PubMed PMID: 41362056; PubMed Central PMCID: PMC12772122.

97. Guo Y, Li C, Tan M, Chen Y, Zhu S, Zhi C, et al. Dynamic Changes in Antibodies and Proteome in Breast Milk of Mothers Infected with Wild-Type SARS-CoV-2 and Omicron: A Longitudinal Study. Nutrients. 2025 Apr 21;17(8):1396. doi:10.3390/nu17081396 PubMed PMID: 40284260; PubMed Central PMCID: PMC12030011.

98. Cao J, Cui X, Lu H, Wang H, Ma W, Yue Z, et al. Regional and longitudinal dynamics of human milk protein components assessed by proteome analysis on a fast and robust micro-flow LC-MS/MS system. Food Chem. 2025 Feb 15;465(Pt 1):141981. doi:10.1016/j.foodchem.2024.141981 PubMed PMID: 39550967.

99. Yu W, Zhang X, Yu Y, Zhao D, Duan S, Szeto IMY, et al. Discovery of milk-derived antimicrobial peptides in human milk by DeepMAMP based on peptidomics technology and deep learning method. Food Chem. 2025 Nov 1;491:145174. doi:10.1016/j.foodchem.2025.145174 PubMed PMID: 40561758.

100. Zhang Y, Zhang X, Mi L, Li C, Zhang Y, Bi R, et al. Comparative Proteomic Analysis of Proteins in Breast Milk during Different Lactation Periods. Nutrients. 2022 Sep 3;14(17):3648. doi:10.3390/nu14173648 PubMed PMID: 36079904; PubMed Central PMCID: PMC9460426.

101. Picáns-Leis R, Vázquez-Mosquera ME, Pereira-Hernández M, Vizoso-González M, López-Valverde L, Barbosa-Gouveia S, et al. Characterization of the functional component in human milk and identification of the molecular mechanisms undergoing prematurity. Clin Nutr. 2025 Jan;44:178–92. doi:10.1016/j.clnu.2024.12.013 PubMed PMID: 39700709.

102. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020 Sep;585(7825):357–62. doi:10.1038/s41586-020-2649-2 PubMed PMID: 32939066; PubMed Central PMCID: PMC7759461.

103. Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, et al. SRplot: A free online platform for data visualization and graphing. PLoS One. 2023;18(11):e0294236. doi:10.1371/journal.pone.0294236 PubMed PMID: 37943830; PubMed Central PMCID: PMC10635526.

104. Pillich RT, Chen J, Churas C, Liu S, Ono K, Otasek D, et al. NDEx: Accessing Network Models and Streamlining Network Biology Workflows. Curr Protoc. 2021 Sep;1(9):e258. doi:10.1002/cpz1.258 PubMed PMID: 34570431; PubMed Central PMCID: PMC8544027.

105. Pillich RT, Chen J, Rynkov V, Welker D, Pratt D. NDEx: A Community Resource for Sharing and Publishing of Biological Networks. Methods Mol Biol. 2017;1558:271–301. doi:10.1007/978-1-4939-6783-4_13 PubMed PMID: 28150243.

106. Pratt D, Chen J, Pillich R, Rynkov V, Gary A, Demchak B, et al. NDEx 2.0: A Clearinghouse for Research on Cancer Pathways. Cancer Res. 2017 Nov 1;77(21):e58–61. doi:10.1158/0008-5472.CAN-17-0606 PubMed PMID: 29092941; PubMed Central PMCID: PMC5679399.

107. Sheils TK, Mathias SL, Kelleher KJ, Siramshetty VB, Nguyen DT, Bologa CG, et al. TCRD and Pharos 2021: mining the human proteome for disease biology. Nucleic Acids Res. 2021 Jan 8;49(D1):D1334–46. doi:10.1093/nar/gkaa993 PubMed PMID: 33156327; PubMed Central PMCID: PMC7778974.

108. Dekker PM, Boeren S, Saccenti E, Hettinga KA. Network analysis of the proteome and peptidome sheds light on human milk as a biological system. Sci Rep. 2024 Mar 30;14(1):7569. doi:10.1038/s41598-024-58127-2 PubMed PMID: 38555284; PubMed Central PMCID: PMC10981717.

109. Medić A, Milićević T, Khraibah A, Herceg Romanić S, Matek Sarić M, Li Y, et al. Total proteome and calcium-binding proteins from human breast milk: Exploring the impact of tobacco smoke exposure and environmental factors. Food Chem. 2025 Apr 30;472:142959. doi:10.1016/j.foodchem.2025.142959 PubMed PMID: 39842200.

110. Zhao H, Cao X, Quan Z, Liu X, Yue X. Novel insights into differences in the whey proteome of porcine and human colostrum and mature milk. Food Chem. 2025 Jun 30;478:143584. doi:10.1016/j.foodchem.2025.143584 PubMed PMID: 40049129.

APPENDIX A

Permission for reproduction of published material

The material presented in Chapter 2 of this dissertation has been reproduced from the following open-access publication:

Singh A, Peng Z, Shah SM, Salkini M, Mallah AF, Ali AF, Kurella SR, Huynh AN, Saleemi SM, Khan A, Mesiya S, Ali A, Samour A, Chaaban H, Yang Z, Ahsan N. Enrichment of Low-Abundance and Low-Molecular-Weight Proteins in Human Milk Using Mass Spectrometry-Based Proteomics. ACS Omega. 2025;10(37):42507–42517. doi:10.1021/acsomega.5c03776.

This article is published under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). Reproduction in this doctoral dissertation is permitted under the terms of this license, as dissertation use constitutes non-commercial reproduction with full attribution and no modification of the original content.

License: https://creativecommons.org/licenses/by-nc-nd/4.0/

RightsLink record: https://marketplace.copyright.com/rs-ui-web/mp/search?type=all&q=10.1021%2Facsomega.5c03776

Supplementary figures: Chapter 2

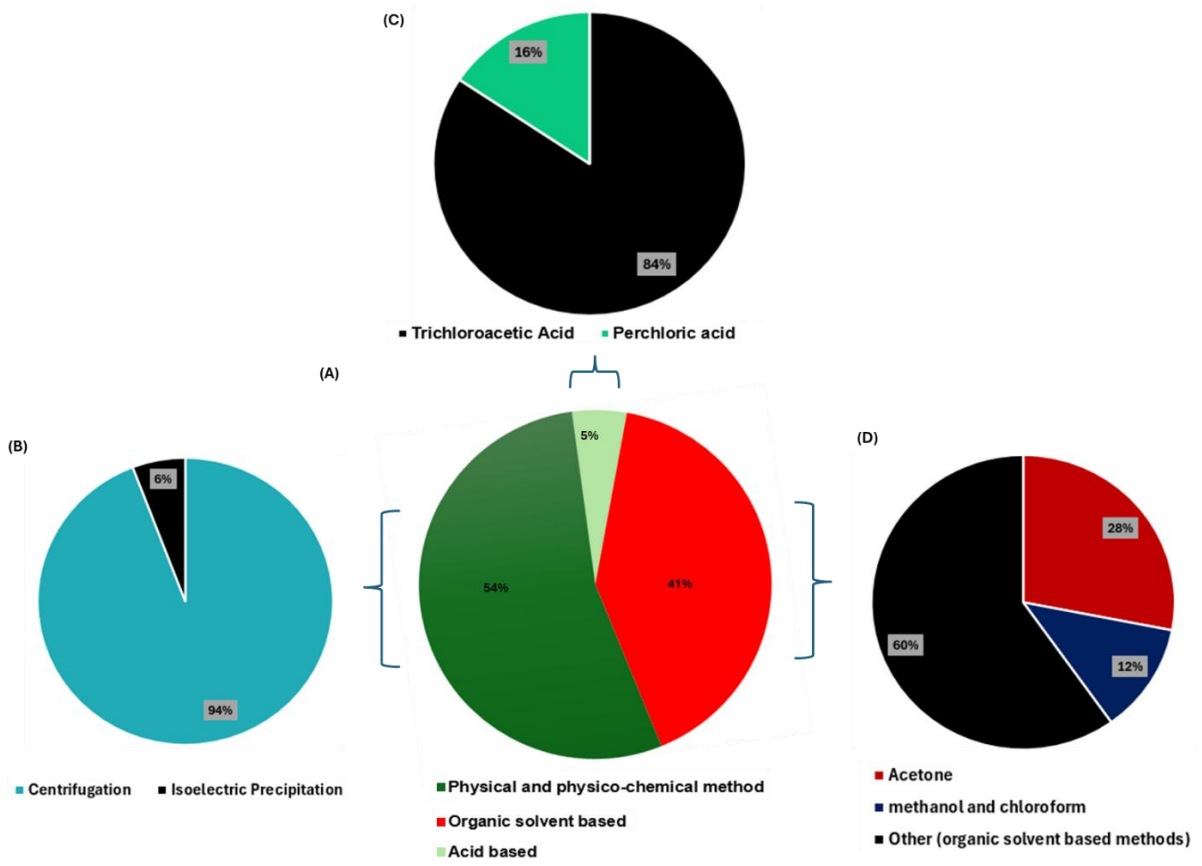


Figure S2-1. Current status of milk proteome depletion methods in published literature. A ScienceDirect literature search conducted on February 5, 2025 using keywords including acetone, centrifugation, methanol chloroform, perchloric acid, trichloroacetic acid, and milk proteomics revealed that physical and physicochemical methods are the most commonly used approaches for protein depletion in milk proteomics (54%), followed by organic solvent-based methods (41%) and acid-based methods (5%). Within the physical and physicochemical category, centrifugation accounts for 94% of methods with isoelectric precipitation accounting for the remaining 6%. Among organic solvent-based approaches, acetone precipitation accounts for 28%, methanol-chloroform for 12%, and other methods primarily ethanol-based for 60%. Acid-based precipitation is predominantly performed using trichloroacetic acid (84%) while perchloric acid is used in 16% of studies. Detailed search results are provided in Table S2-1.

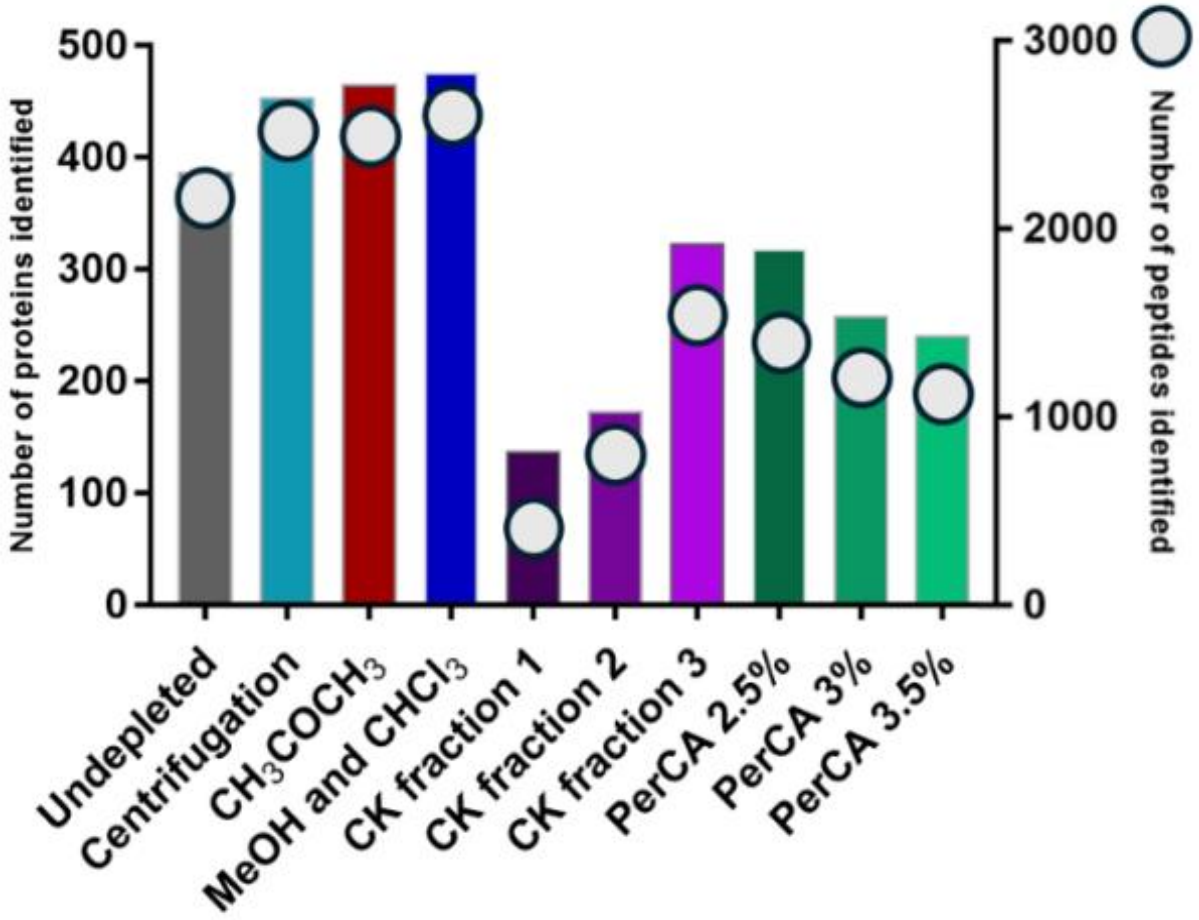


Figure S2-2. Protein and peptide identification at individual PerCA concentrations and CK fractions. The number of proteins and peptides identified using each individual condition including the three fractions of the commercial kit (CK-F1, CK-F2, CK-F3) and perchloric acid at three concentrations (PerCA 2.5%, PerCA 3.0%, and PerCA 3.5%). Results are shown separately for each fraction and concentration to illustrate the contribution of each individual condition to overall proteome coverage. Detailed protein and peptide lists are provided in Table S2-2 and Table S2-3, respectively.

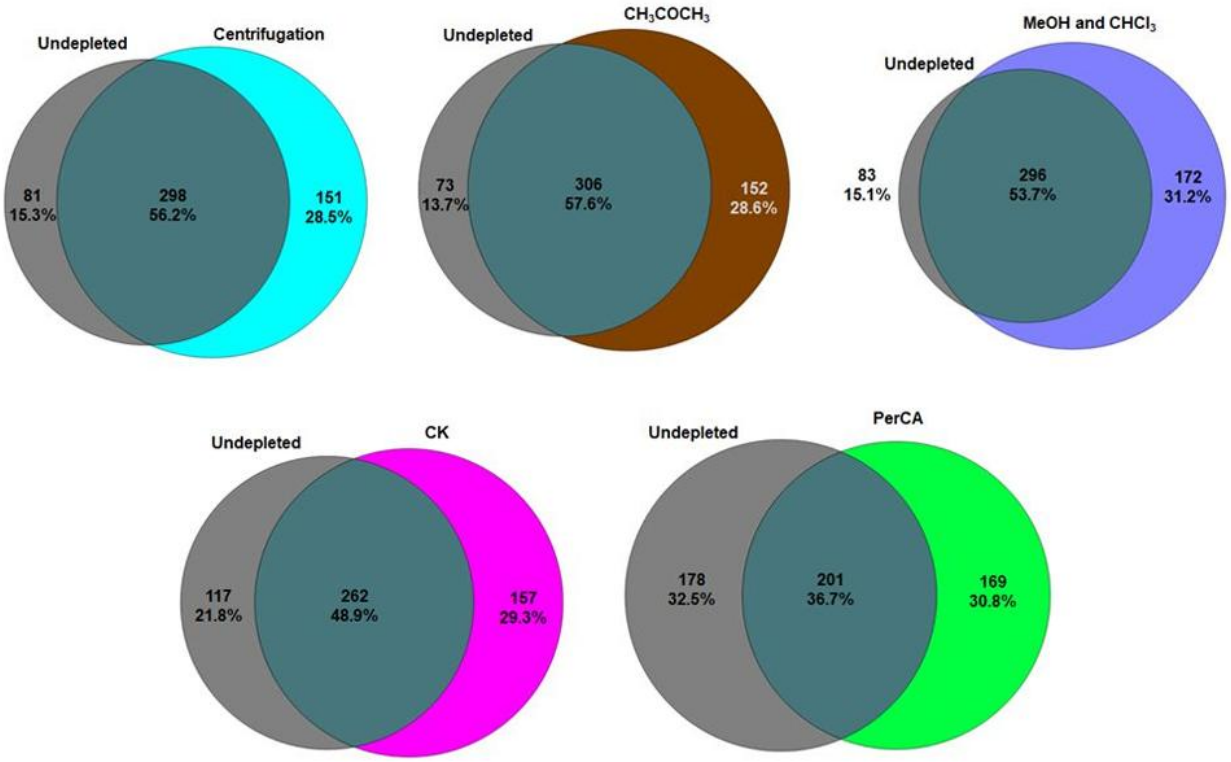


Figure S2-3. Pairwise milk proteome comparison obtained from different depletion methods. Venn diagrams illustrate the comparison of proteins identified by different methods against the control (undepleted milk).

Supplementary tables: Chapter 2

Table S2-1. Literature search results on milk proteome depletion methods. A ScienceDirect literature search was conducted on February 5, 2025, using keywords including acetone, centrifugation, methanol, chloroform, perchloric acid, depletion, and milk proteomics/proteome. Results represent the number of published studies retrieved for each search term and method category.

Method	Number of Results
Physical and Physicochemical Methods	15,336
Centrifugation	14,448
Isoelectric Precipitation	888
Organic Solvent-Based Methods	11,407
Acetone	3,199
Methanol and Chloroform	1,380
Other (primarily ethanol-based)	6,828
Acid-Based Methods	1,335
Trichloroacetic Acid	1,126
Perchloric Acid	209
Total	28,078

Table S2-2. Complete list of proteins identified by each depletion method using LC-MS/MS analysis. Full data available at:
https://pubs.acs.org/doi/suppl/10.1021/acsomega.5c03776/suppl_file/ao5c03776_si_003.xlsx

Table S2-3. Complete list of peptides identified by each depletion method using LC-MS/MS analysis. Full data available at:

https://pubs.acs.org/doi/suppl/10.1021/acsomega.5c03776/suppl_file/ao5c03776_si_004.xlsx

Table S2-4. List of unique proteins identified exclusively by the PerCA depletion method. Full data available at:

https://pubs.acs.org/doi/suppl/10.1021/acsomega.5c03776/suppl_file/ao5c03776_si_005.xlsx

Table S2-5. Comparative milk proteome analysis against recently published datasets. Full data available at:
https://pubs.acs.org/doi/suppl/10.1021/acsomega.5c03776/suppl_file/ao5c03776_si_006.xlsx

APPENDIX B

Supplementary figures: Chapter 3

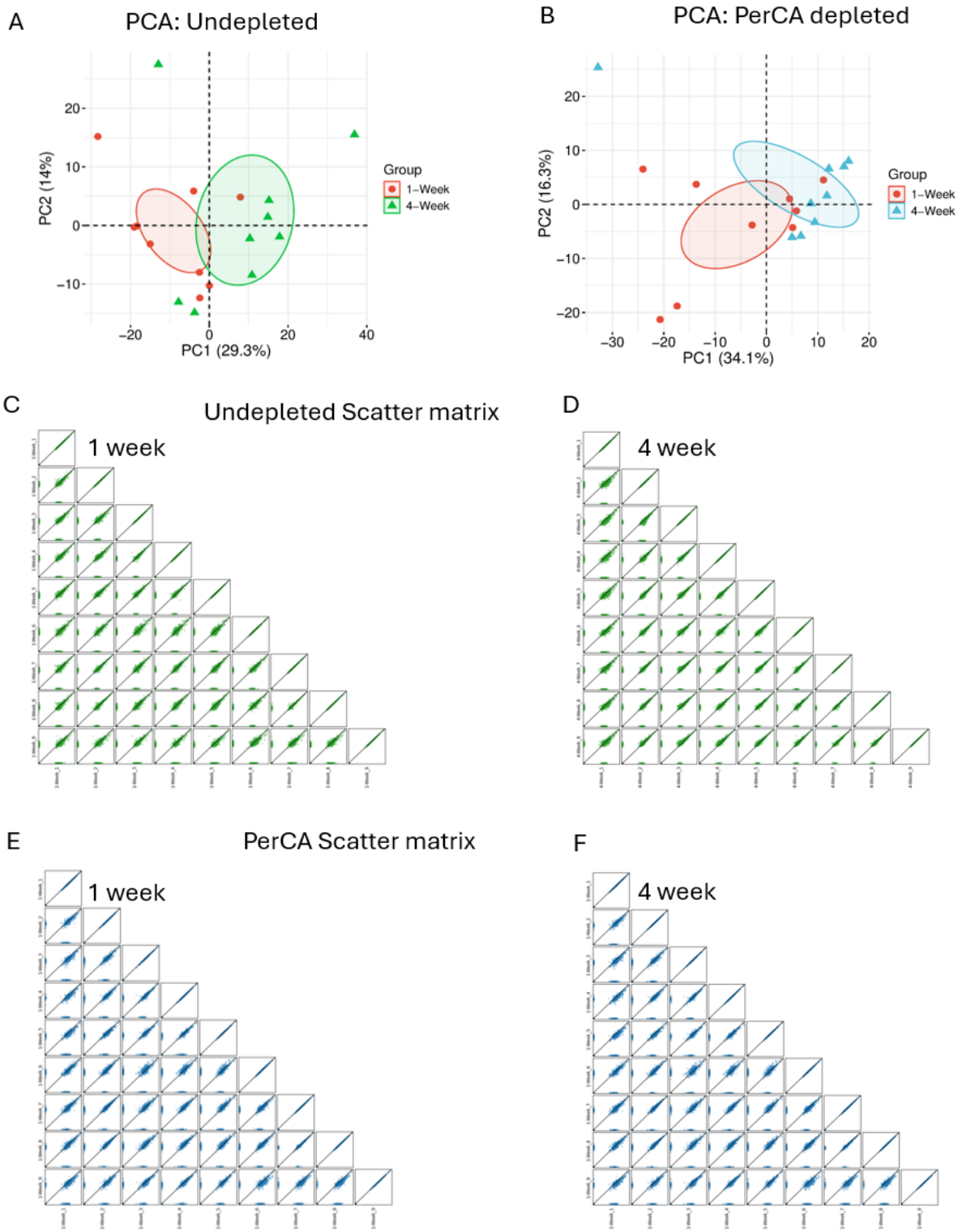


Figure S3-1. Principal component analysis (PCA) and scatter matrix analysis of protein abundance in undepleted and PerCA-depleted datasets. (A) PCA score plot of undepleted samples showing clustering of one-week and four-week groups based on global protein abundance profiles. (B) PCA score plot of PerCA-depleted samples showing corresponding lactation stage separation. (C and D) Scatter matrix plots illustrating pairwise relationships in protein abundance between individual undepleted samples at one week and four weeks. (E and F) Scatter matrix plots illustrating pairwise relationships in protein abundance between individual PerCA-depleted samples at one week and four weeks. Consistent clustering across both multivariate approaches confirms that lactation stage is the primary source of proteome variation in both sample preparation workflows.

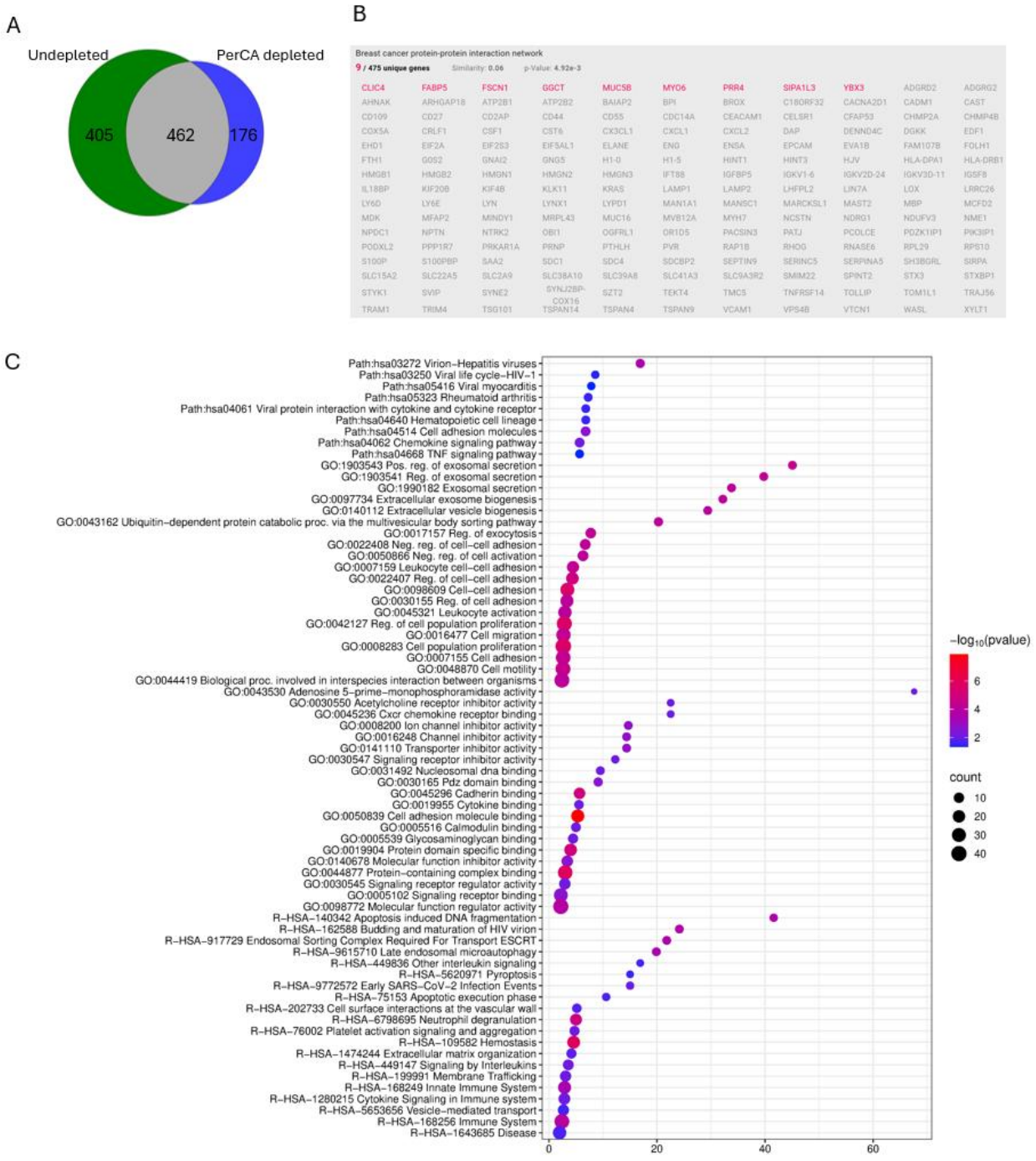


Figure S3-2. Overview of proteins uniquely identified in PerCA-depleted samples. (A) Venn diagram showing the 176 proteins exclusively detected in PerCA-depleted samples relative to the undepleted dataset. (B) Gene Ontology enrichment analysis of the 176 PerCA-unique proteins showing enriched biological process, molecular function, and cellular component terms. (C) Functional classification of PerCA-unique proteins by biological category, highlighting representation across immune defense, metabolic regulation, and structural functions.

A

WP2877 - Vitamin D receptor pathway - Homo sapiens

7 / 184 unique genes Similarity: **0.05** p-Value: **3.72e-1**

ATP2B1	CEACAM1	CST6	G0S2	HLA-DRB1	IGFBP5	PTHLH	ADGRD2	ADGRG2
AHNAK	ARHGAP18	ATP2B2	BAIAP2	BPI	BROX	C18ORF32	CACNA2D1	CADM1
CAST	CD109	CD27	CD2AP	CD44	CD55	CDC14A	CELSR1	CFAP53
CHMP2A	CHMP4B	CLIC4	COX5A	CRLF1	CSF1	CX3CL1	CXCL1	CXCL2
DAP	DENND4C	DGKK	EDF1	EHD1	EIF2A	EIF2S3	EIF5AL1	ELANE
ENG	ENSA	EPCAM	EVA1B	FABP5	FAM107B	FOLH1	FSCN1	FTH1
GGCT	GNAI2	GNG5	H1-0	H1-5	HINT1	HINT3	HJV	HLA-DPA1
HMGB1	HMGB2	HMGN1	HMGN2	HMGN3	IFT88	IGKV1-6	IGKV2D-24	IGKV3D-11
IGSF8	IL18BP	KIF20B	KIF4B	KLK11	KRAS	LAMP1	LAMP2	LHFPL2
LIN7A	LOX	LRRC26	LY6D	LY6E	LYN	LYNX1	LYPD1	MAN1A1
MANSC1	MARCKSL1	MAST2	MBP	MCFD2	MDK	MFAP2	MINDY1	MRPL43
MUC16	MUC5B	MVB12A	MYH7	MYO6	NCSTN	NDRG1	NDUFV3	NME1
NPDC1	NPTN	NTRK2	OBI1	OGFRL1	OR1D5	PACSIN3	PATJ	PCOLCE
PDZK1IP1	PIK3IP1	PODXL2	PPP1R7	PRKAR1A	PRNP	PRR4	PVR	RAP1B
RHOG	RNASE6	RPL29	RPS10	S100P	S100PBP	SAA2	SDC1	SDC4
SDCBP2	SEPTIN9	SERINC5	SERPINA5	SH3BGRL	SIPA1L3	SIRPA	SLC15A2	SLC22A5
SLC2A9	SLC38A10	SLC39A8	SLC41A3	SLC9A3R2	SMIM22	SPINT2	STX3	STXBP1
STYK1	SVIP	SYNE2	SYNJ2BP-COX16	SZT2	TEKT4	TMC5	TNFRSF14	TOLLIP
TOM1L1	TRAJ56	TRAM1	TRIM4	TSG101	TSPAN14	TSPAN4	TSPAN9	VCAM1
VPS4B	VTCN1	WASL	XYLT1	YBX3				

B

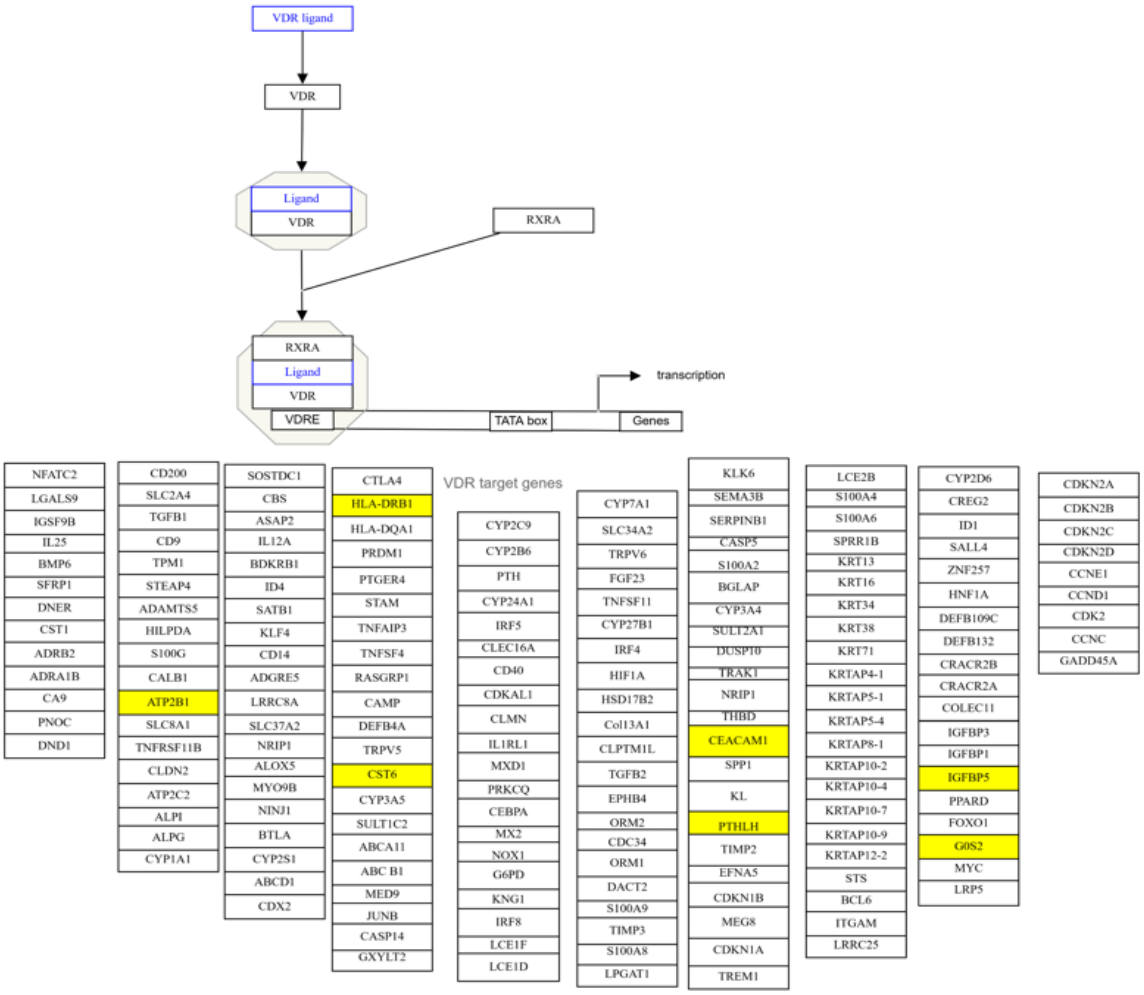


Figure S3-3. Vitamin D receptor (VDR) signaling pathway enrichment of PerCA-unique proteins. Network visualization showing the integration of PerCA-unique proteins within the curated VDR signaling pathway, with detected proteins highlighted and pathway connectivity illustrated through protein-protein interaction nodes.

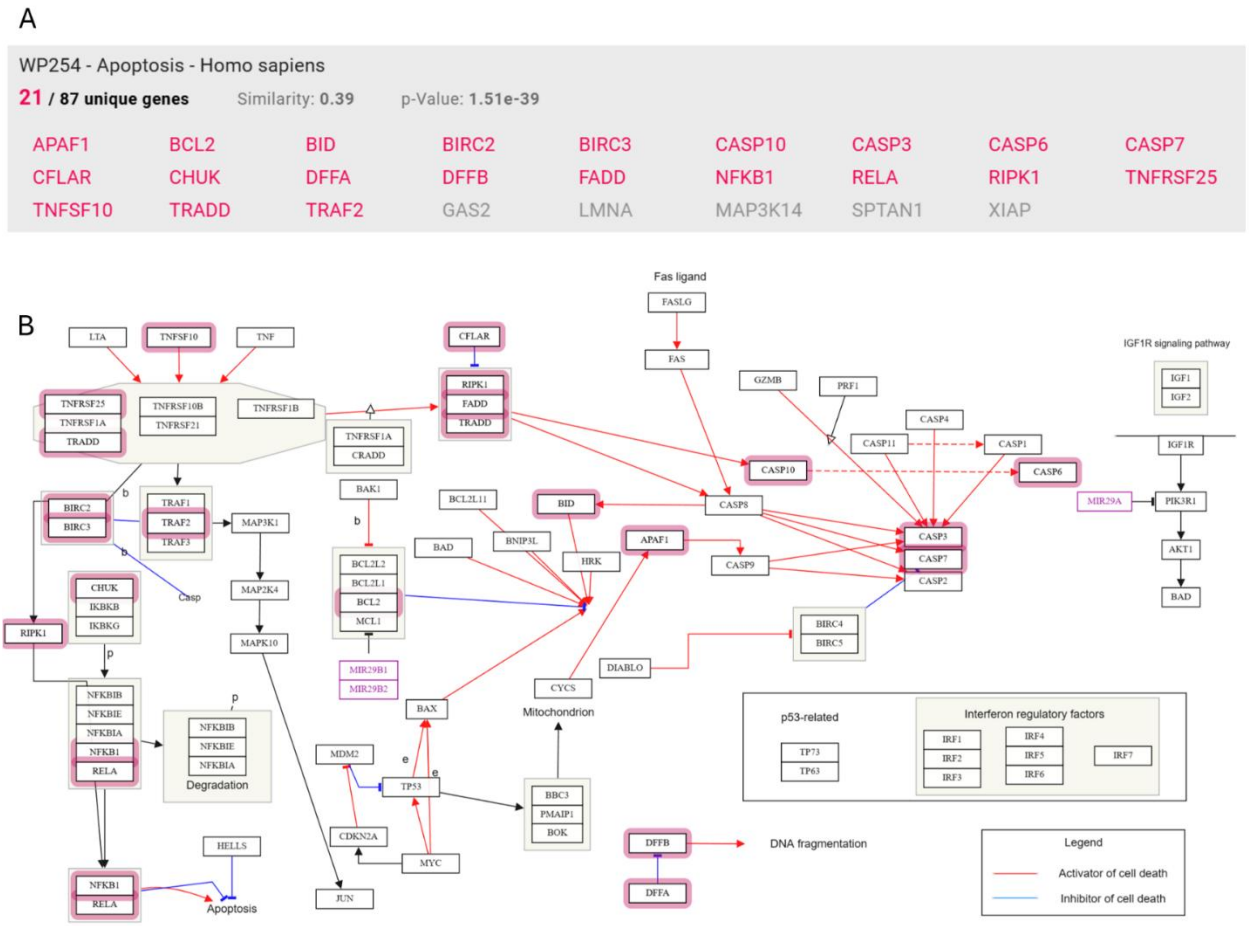


Figure S3-4. Apoptosis pathway enrichment of PerCA-unique proteins. Network visualization highlighting detected PerCA-unique proteins within the curated apoptosis signaling pathway, showing their positions within programmed cell death cascades including both intrinsic and extrinsic apoptotic routes.

A

WP2361 - Gastric cancer network 1 - Homo sapiens

2 / 28 unique genes Similarity: **0.04** p-Value: **3.72e-1**

KIF20B	S100P	ADGRD2	ADGRG2	AHNAK	ARHGAP18	ATP2B1	ATP2B2	BAIAP2
BPI	BROX	C18ORF32	CACNA2D1	CADM1	CAST	CD109	CD27	CD2AP
CD44	CD55	CDC14A	CEACAM1	CELSR1	CFAP53	CHMP2A	CHMP4B	CLIC4
COX5A	CRLF1	CSF1	CST6	CX3CL1	CXCL1	CXCL2	DAP	DENND4C
DGKK	EDF1	EHD1	EIF2A	EIF2S3	EIF5AL1	ELANE	ENG	ENSA
EPCAM	EVA1B	FABP5	FAM107B	FOLH1	FSCN1	FTH1	G0S2	GGCT
GNAI2	GNG5	H1-0	H1-5	HINT1	HINT3	HJV	HLA-DPA1	HLA-DRB1
HMGB1	HMGB2	HMGN1	HMGN2	HMGN3	IFT88	IGFBP5	IGKV1-6	IGKV2D-24
IGKV3D-11	IGSF8	IL18BP	KIF4B	KLK11	KRAS	LAMP1	LAMP2	LHFPL2
LIN7A	LOX	LRRC26	LY6D	LY6E	LYN	LYNX1	LYPD1	MAN1A1
MANSC1	MARCKSL1	MAST2	MBP	MCFD2	MDK	MFAP2	MINDY1	MRPL43
MUC16	MUC5B	MVB12A	MYH7	MYO6	NCSTN	NDRG1	NDUFV3	NME1
NPDC1	NPTN	NTRK2	OBI1	OGFRL1	OR1D5	PACSIN3	PATJ	PCOLCE
PDZK1IP1	PIK3IP1	PODXL2	PPP1R7	PRKAR1A	PRNP	PRR4	PTHLH	PVR
RAP1B	RHOG	RNASE6	RPL29	RPS10	S100PBP	SAA2	SDC1	SDC4
SDCBP2	SEPTIN9	SERINC5	SERPINA5	SH3BGRL	SIPA1L3	SIRPA	SLC15A2	SLC22A5
SLC2A9	SLC38A10	SLC39A8	SLC41A3	SLC9A3R2	SMIM22	SPINT2	STX3	STXBP1
STYK1	SVIP	SYNE2	SYNJ2BP-COX16	SZT2	TEKT4	TMC5	TNFRSF14	TOLLIP
TOM1L1	TRAJ56	TRAM1	TRIM4	TSG101	TSPAN14	TSPAN4	TSPAN9	VCAM1
VPS4B	VTCN1	WASL	XYLT1	YBX3				

B

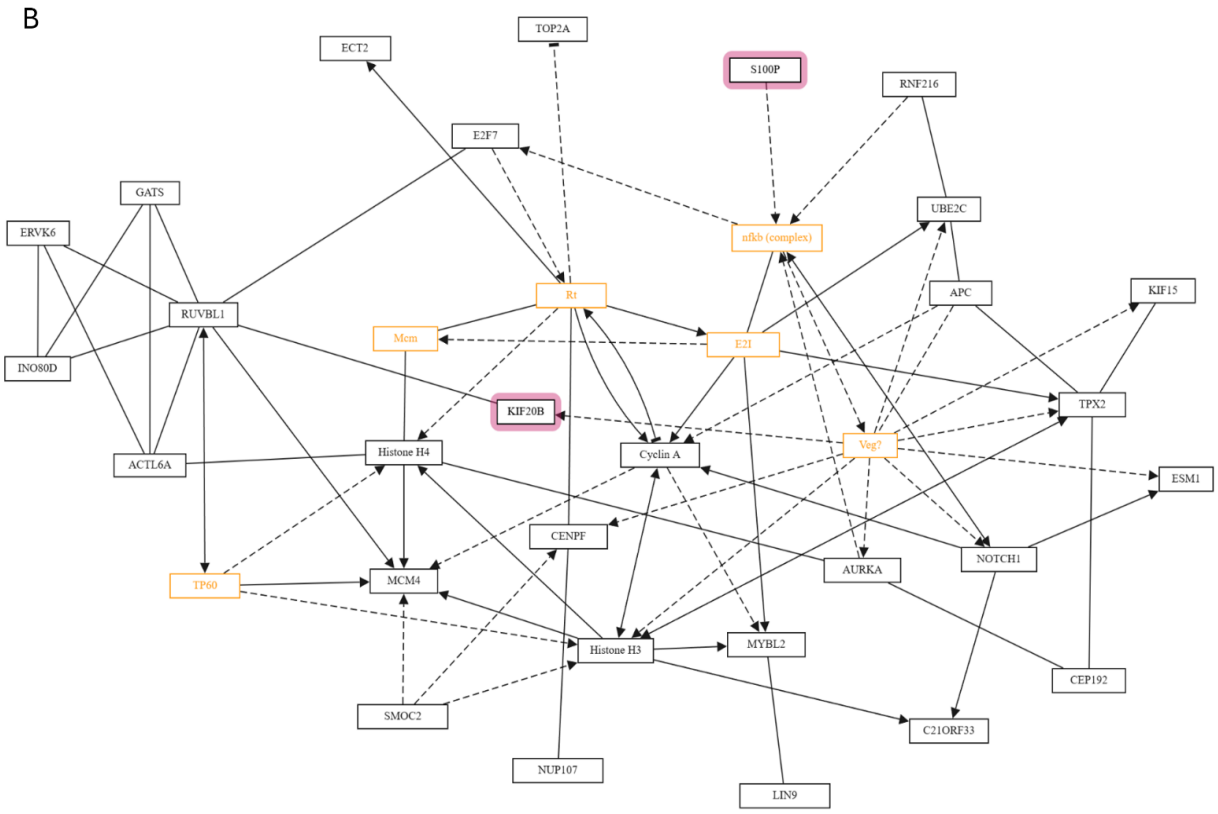


Figure S3-5. Gastric cancer pathway network of PerCA-unique proteins. Network diagram showing the integration of PerCA-unique proteins within the curated gastric

cancer pathway, illustrating their connectivity to oncogenic and tumor-suppressive nodes within this disease-associated network.



Figure S3-6. Breast cancer protein-protein interaction network of PerCA-unique proteins. Network visualization highlighting PerCA-unique proteins within curated breast cancer interaction modules, showing functional clustering and connectivity among proteins with documented roles in breast cancer pathways.

A

Lymphocyte cell-cell adhesion

2 / 13 unique genes Similarity: **0.09** p-Value: **3.63e-1**

CD27	CD44	ADGRD2	ADGRG2	AHNAK	ARHGAP18	ATP2B1	ATP2B2	BAIAP2	BPI	BROX
C18ORF32	CACNA2D1	CADM1	CAST	CD109	CD2AP	CD55	CDC14A	CEACAM1	CELSR1	CFAP53
CHMP2A	CHMP4B	CLIC4	COX5A	CRLF1	CSF1	CST6	CX3CL1	CXCL1	CXCL2	DAP
DENND4C	DGKK	EDF1	EHD1	EIF2A	EIF2S3	EIF5AL1	ELANE	ENG	ENSA	EPCAM
EVA1B	FABP5	FAM107B	FOLH1	FSCN1	FTH1	G0S2	GGCT	GNAI2	GNG5	H1-0
H1-5	HINT1	HINT3	HJV	HLA-DPA1	HLA-DRB1	HMGB1	HMGB2	HMGN1	HMGN2	HMGN3
IFT88	IGFBP5	IGKV1-6	IGKV2D-24	IGKV3D-11	IGSF8	IL18BP	KIF20B	KIF4B	KLK11	KRAS
LAMP1	LAMP2	LHFPL2	LIN7A	LOX	LRRC26	LY6D	LY6E	LYN	LYNX1	LYPD1
MAN1A1	MANSC1	MARCKSL1	MAST2	MBP	MCFD2	MDK	MFAP2	MINDY1	MRPL43	MUC16
MUC5B	MVB12A	MYH7	MYO6	NCSTN	NDRG1	NDUFV3	NME1	NPDC1	NPTN	NTRK2
OBI1	OGFRL1	OR1D5	PACSIN3	PATJ	PCOLCE	PDZK1IP1	PIK3IP1	PODXL2	PPP1R7	PRKAR1A
PRNP	PRR4	PTHLH	PVR	RAP1B	RHOG	RNASE6	RPL29	RPS10	S100P	S100PBP
SAA2	SDC1	SDC4	SDCBP2	SEPTIN9	SERINC5	SERPINA5	SH3BGRL	SIPA1L3	SIRPA	SLC15A2
SLC22A5	SLC2A9	SLC38A10	SLC39A8	SLC41A3	SLC9A3R2	SMIM22	SPINT2	STX3	STXBP1	STYK1
SVIP	SYNE2	SYNJ2BP-COX16	SZT2	TEKT4	TMC5	TNFRSF14	TOLLIP	TOM1L1	TRAJ56	TRAM1
TRIM4	TSG101	TSPAN14	TSPAN4	TSPAN9	VCAM1	VPS4B	VTCN1	WASL	XYLT1	YBX3

B

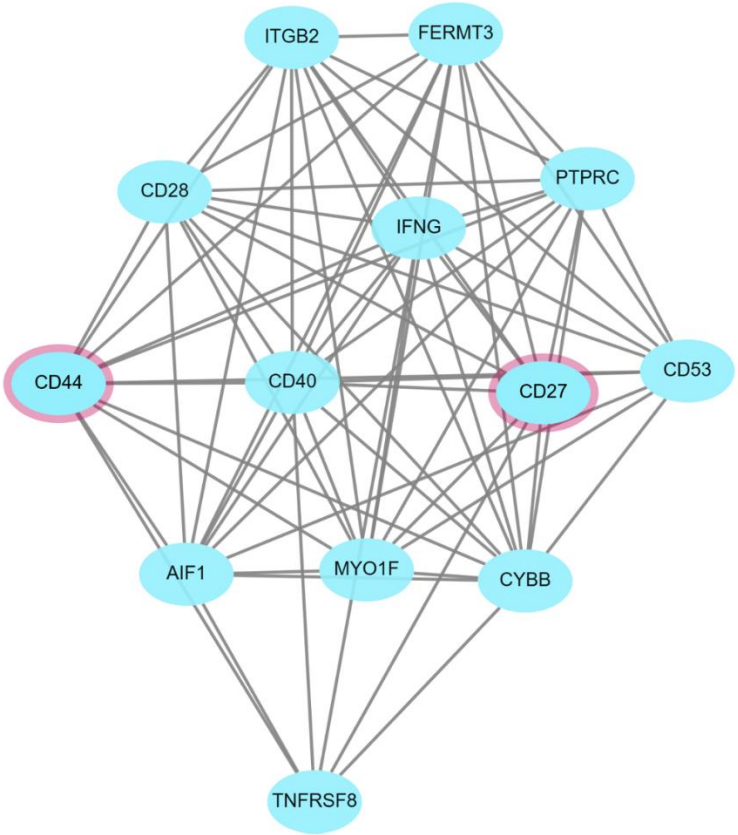


Figure S3-7. Lymphocyte cell-cell adhesion pathway network of PerCA-unique proteins. Network diagram illustrating immune-related PerCA-unique proteins within the lymphocyte cell-cell adhesion pathway, highlighting their roles in immune cell interactions and adhesion-mediated signaling relevant to mucosal immunity in the neonatal gut.

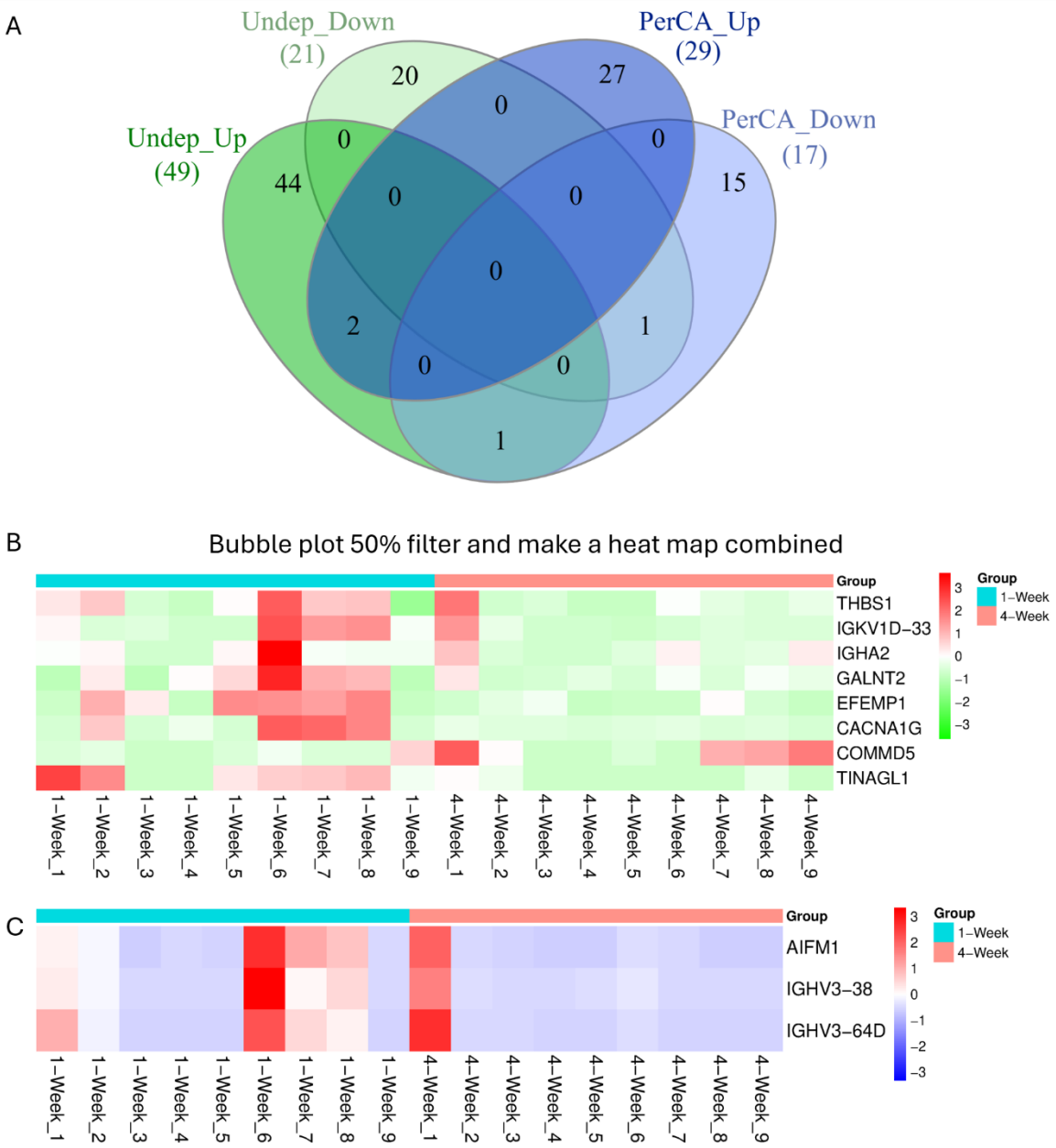
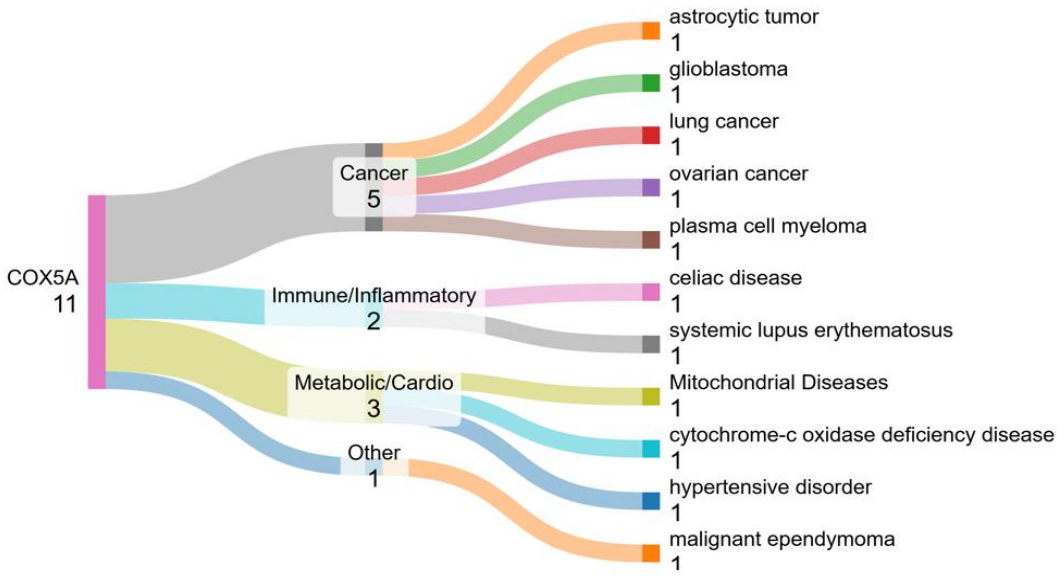


Figure S3-8. Overlap and filtered heatmap analysis of differentially regulated proteins. (A) Venn diagram showing the overlap of upregulated and downregulated proteins between undepleted and PerCA-depleted datasets at one week versus four weeks, illustrating the largely complementary nature of the two preparation workflows in capturing regulated proteins. (B) Heatmap of differentially regulated proteins in the undepleted dataset after applying a 50% detection filter. (C) Corresponding filtered heatmap for the PerCA-depleted dataset showing stage-specific abundance patterns after the same detection threshold was applied.

A


COX5A
11
Cancer
5
Immune/Inflammatory
2
Metabolic/Cardio
3
Other
1
astrocytic tumor
1
glioblastoma
1
lung cancer
1
ovarian cancer
1
plasma cell myeloma
1
celiac disease
1
systemic lupus erythematosus
1
Mitochondrial Diseases
1
cytochrome-c oxidase deficiency disease
1
hypertensive disorder
1
malignant ependymoma
1

B

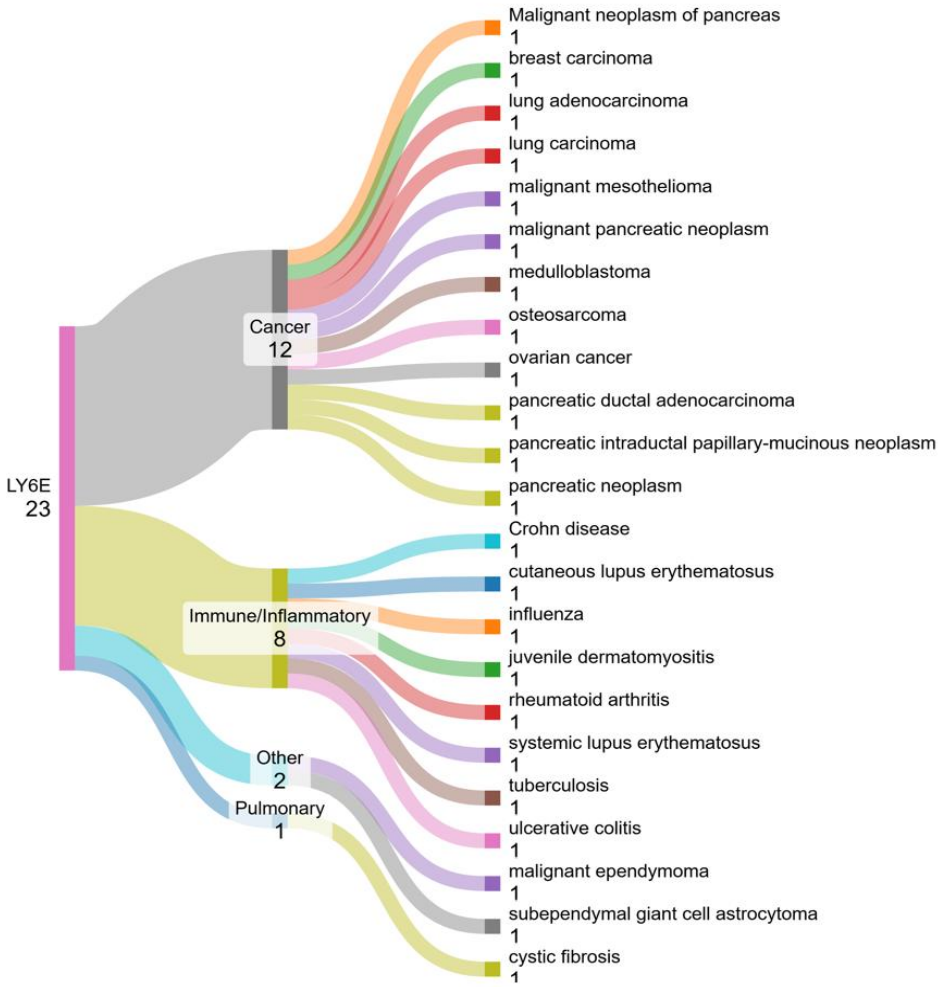

LY6E
23
Cancer
12
Immune/Inflammatory
8
Other
2
Pulmonary
1
Malignant neoplasm of pancreas
1
breast carcinoma
1
lung adenocarcinoma
1
lung carcinoma
1
malignant mesothelioma
1
malignant pancreatic neoplasm
1
medulloblastoma
1
osteosarcoma
1
ovarian cancer
1
pancreatic ductal adenocarcinoma
1
pancreatic intraductal papillary-mucinous neoplasm
1
pancreatic neoplasm
1
Crohn disease
1
cutaneous lupus erythematosus
1
influenza
1
juvenile dermatomyositis
1
rheumatoid arthritis
1
systemic lupus erythematosus
1
tuberculosis
1
ulcerative colitis
1
malignant ependymoma
1
subependymal giant cell astrocytoma
1
cystic fibrosis
1

Figure S3-9. Extended disease association Sankey diagrams for stage-specific proteins identified in the PerCA-depleted dataset. Sankey diagrams illustrating the disease category mapping of proteins uniquely upregulated at one week and four weeks postpartum, showing distribution across cancer, immune/inflammatory, pulmonary, neurological, and other disease categories derived from NIH Pharos and Target Central Resource Database classifications.

Supplementary tables: Chapter 3

Table S3-1. Proteins uniquely upregulated at each lactation stage in PerCA-depleted and undepleted datasets.

Unique Upregulated genes in 1-week PerCA	Unique Upregulated genes in 4-week PerCA
ARHGAP18	BPI
CLIC4	ELANE
COX5A	FAM107B
EIF5AL1	MINDY1
FSCN1	SLC41A3
GNAI2	TRAM1
LY6E	
MFAP2	
MUC5B	
PPP1R7	
PRKAR1A	
SLC22A5	
SYNJ2BP-COX16	

Table S3-2. Therapeutic compounds and drug targets associated with ELANE.

Drug name	Primary target/mechanism	Key description	Clinical/therapeutic use
Sivelestat	Neutrophil elastase inhibitor	Direct inhibitor of neutrophil elastase; structure reported in original source	Investigational/used for acute lung injury and inflammatory conditions
Quercetin	Broad antioxidant; indirect enzyme modulation	Flavonol widely distributed in plants; antioxidant; glycosylated forms include rutin	Nutraceutical; anti-inflammatory and antioxidant effects
Telaprevir	HCV NS3/4A serine protease inhibitor	Inhibits proteolytic processing of HCV polyprotein; IC_{50} ≈ 10 nM	Antiviral therapy for Hepatitis C
Boceprevir	HCV NS3/4A serine protease inhibitor	Covalent but reversible binding to NS3 active-site serine; Ki ≈ 14 nM	Antiviral therapy for Hepatitis C
Bortezomib	Proteasome inhibitor	Pyrazine–boronic acid derivative; reversible proteasome inhibition	Antineoplastic agent for multiple myeloma and mantle cell lymphoma

Supplementary Data 1 through 5. Comprehensive proteomic datasets for Chapter 3. All files available at MassIVE MSV000101302 (PerCA-depleted) and MSV000101301 (Undepleted) at https://massive.ucsd.edu.