Method Development and Application of Mass Spectrometry-based Omics Analyses

By

Pingli Wei

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

University of Wisconsin-Madison

2020

Date of Final Oral Examination: May 20th, 2020

This dissertation is approved by the following members of the Final Oral Exam Committee:

Lingjun Li, Professor, Pharmacy and Chemistry

Bill Chiu, Professor, Neuroscience

Vaishali P Bakshi, Professor, Psychiatry

Jun Dai, Professor, Pharmacy

Acknowledgments

First of all, I would like to thank my Ph.D. advisor, Dr. Lingjun Li, for her continued support and guidance in these five years. It is my honor to join Dr. Li's lab with innovative research and cutting-edge instrumentation. More importantly, Dr. Li is such a wonderful mentor who really cares about her students' growth and always tries to provide the best resources to her students. She gives me so many opportunities for diverse projects and has always been incredibly supportive of my research direction. None of the work in this dissertation could be possible without her. I will be forever grateful for her and proud to be a Li lab member.

I also would like to thank all my committee members, Dr. Bill Chiu, Dr. Vaishali P Bakshi, and Dr. Jun Dai, for their guidance and suggestions on the thesis background oral (TBO), original research proposal (RP), 4th-year committee meeting, and, finally, this dissertation. Their support is vital in my development as a qualified scientist.

In addition to my committee members, I wish to thank other collaborators who I have been fortunate to work with on various exciting bioanalytical projects (names not listed in order). I wish to thank Prof. William J Burlingham, Dr. Ewa Jankowska-Gan, Diego Lema, and Prof. Matthew Brown for their great support on the post-transplant survival rate improvement project. Thank you to Prof. Cameron R. Currie and Jennifer R. Bratburd for the human microbiome function study. I wish to thank Prof. Wade Bushman and Prof. William Ricke for the LUTS biomarker investigation projects. Thank you to Prof. Vaishali P Bakshi and Prof. Brian Baldo for the PTSD and cocaine withdrawal projects. Thank you to Prof. Woo-Ping Ge from the University of Texas Southwestern

Medical Center for the SCO neuro-function exploration project. I would like to thank all of them whole heartily for providing exciting research projects and wonderful opportunities to collaborate.

It is my honor to work alongside incredibly talented mentors and colleagues in our lab during my graduate study. At the very beginning of my graduate study, Dr. Fengfei Ma taught me everything about mammalian neuropeptide and proteomics study. Dr. Ling Hao led me into the metabolomics method development, and Dr. Qing Yu mentored me on the instrumentation. I sincerely appreciate all of their mentoring. Without their omics instruction, there is no chance I could conduct my further multi-omics integration study. Dr. Caitlin Keller is an expert on the microbiome, and it was a great pleasure to work with her for three microbiome-related projects. Dr. Yu Feng, Kellen Delaney, and I joint the Li lab together, and we had many good discussions during these five years. It is so lucky to be friends with them from the beginning to the end. I also want to thank Dr. Erin Gemperline and Dr. Samuel Thomas's help for editing my manuscripts. Dr. Bingming Chen and Dr. Jill Johnson gave me many great suggestions about preparing good quality presentations. I am happy to get a chance to work with two visiting scholars in our lab, Prof. Chunlan Tang, from the Medical School of Ningbo University for the multi-omics study for Alzheimer's disease (AD) research and Prof. Rui Liu from the Nanjing University of Chinese Medicine for the gut-brain axis study. I would like to give my special thanks to the instrument maintenance team, Zihui (Jerry) Li, Dr. Yatao Shi, Chris Sauer, Dylan Tabang and Bin Wang, and DiLeu synthesis team, Dr. Amanda Buchberger, Yusi Cui and Dr. Dustin Frost for their great support. Thank you to Graham Delafield for his great photos. Moreover, I have benefited from day-to-day interaction and discussions with other colleagues in the Li lab: Dr. Chenxi Jia, Dr. Xuefei Zhong, Dr. Xiaofang Zhong, Dr. Min

Ma, Dr. Tony Chen, Dr. Yang Liu, Dr. Qinjingwen Cao, Qinying Yu, Meng Xu, Mike Li, Nhu Vu, Jiabao Guo, Jericha T Mill, and Hannah Miles.

Beyond the lab, Dr. Cameron Scarlett and Molly Pellitteri-Hahn from the Analytical Instrumentation Center in the School of Pharmacy, who were beyond helpful in with their instrumentation expertise. I would also like to thank Prof. Melgardt De Villiers and lab coordinator Bonnie J Fingerhut for all the caring and support when I worked with them as a teaching assistant during my early stage of the graduate career. I would like to extend my gratitude to the UW-Madison's Graduate School and Chemistry Department Graduate Student-Faculty Liaison Committee (GSFLC) for financial support of research conference travel.

Finally, I would like to thank all of my family and friends who have supported me through thick and thin during the last five years of graduate school. In particular, I want to thank my parents, my sibling, my uncles, and my aunts' families, their unconditional love and support help me overcome the difficulties in my life and make me become a better person. It should also be noted that I would have never come to graduate school without the support of my former supervisor, Qi Gu, who introduced me to the world of research and encouraged me to pursue my Ph.D. study. Most importantly, I want to thank my husband, Zhe (Jason) Liu, who continuously support me through graduate school stress. Each tough moment, he is by my side. I am forever grateful that God brought him into my life! At last, I want to thank my angelic baby, Grace Wei Liu, without her support, this thesis should be finished at least one month earlier (LOL).

To all those I have mentioned, I dedicate this thesis.

Table of Contents

Acknowledgments		i
Table of Contents		iv
Abstract		v
Chapter 1	Introduction and Research Summary	1
Chapter 2	Urinary Amine Metabolomics Characterization of a Hormone-Induced Urinary Obstruction Mouse Model with Custom 12-plex Isobaric DiLeu labeling	8
Chapter 3	Urinary Metabolomics and Proteomics Analyses in a Mouse Model of Prostatic Inflammation	35
Chapter 4	Development of a Novel Mass Spectrometry Approach for Comprehensive Neuropeptidome Characterization and its Application to the Analysis of Human Pituitary Tumor	67
Chapter 5	Neuropeptidomic Study of the Mammalian Subcommissural Organ (SCO) by High-Resolution Mass Spectrometry	98
Chapter 6	Neuropeptides in Microbiome-Gut-Brain Axis and Their Influence on Host Immunity and Stress	120
Chapter 7	Integrated, Multi-Omics Strategy to Study the Host Immune Response Induced by Human Microbiome during *Salmonella enterica* Typhimurium and *Candida albicans* Infection	156
Chapter 8	Urine Biomarker Investigation of Lower Urinary Tract Symptoms with Mass Spectrometry (WISL Chapter)	192
Chapter 9	Conclusions and Future Directions	212
Appendix I	List of Publications and Presentations	220
Appendix II	Metandem: A novel online software platform for mass spectrometry-based isobaric labeling metabolomics	224
Appendix III	Absolute Quantification of Amine Metabolites in Human Cerebrospinal Fluid via MS1-centric Isotopic DiLeu (iDiLeu) Labeling	245
Appendix IV	Neuropeptidomics of the Rat Habenular Nuclei	271
Appendix V	Integrated Label-Free and Ten-Plex DiLeu Tag Quantitative Methods for Profiling Changes in the Mouse Hypothalamic Neuropeptidome and Proteome under Different Gut Microbiota Environments	298
Appendix VI	Extraction Optimization for Combined Metabolomics, Peptidomics, and Proteomics Analysis of Cecum Gut Microbiota Samples	335

Method Development and Application of Mass Spectrometry-based Omics Analyses

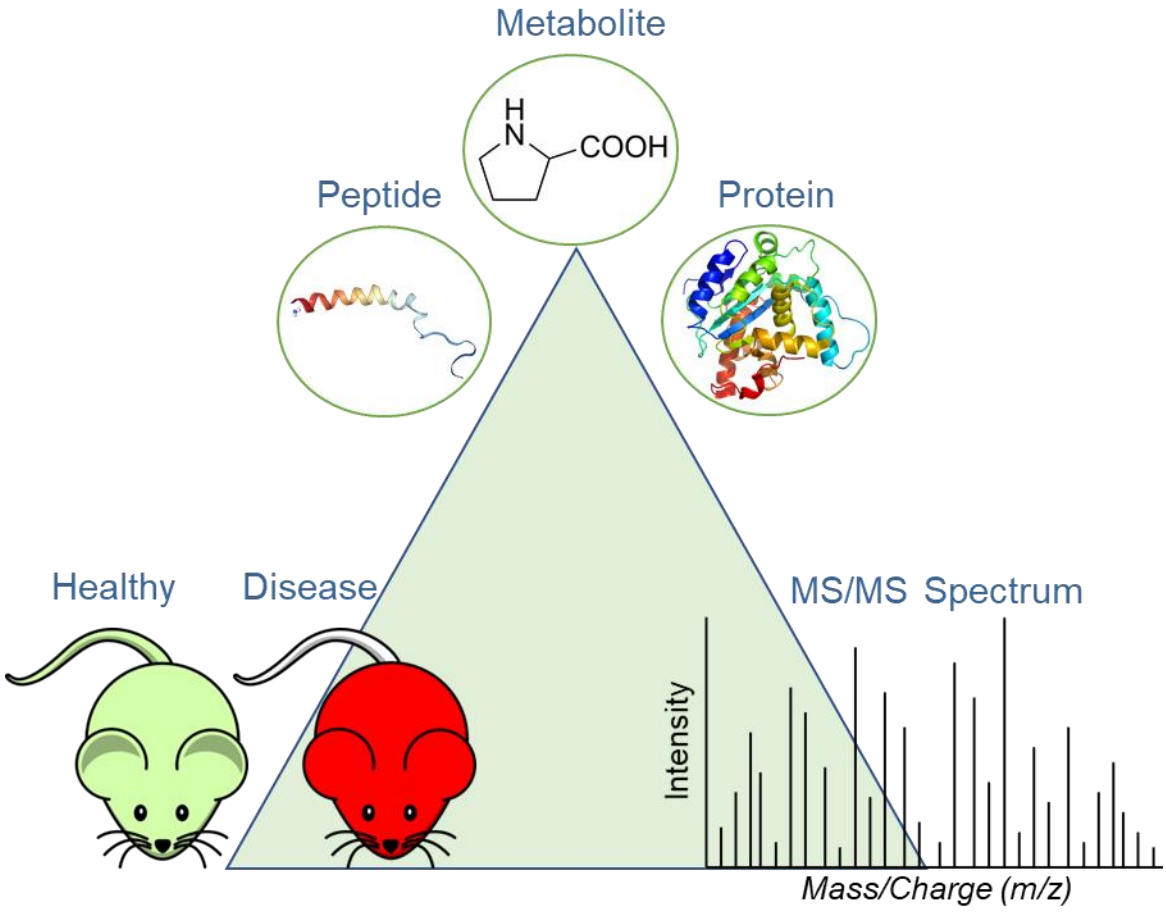
Pingli Wei
Under the Supervision of Professor Lingjun Li
University of Wisconsin-Madison

Abstract

Endogenous biomolecules, such as metabolites, neuropeptides, and proteins, play important roles in biological and physiological processes. Mass spectrometry (MS) has become a central technology for the study of biomolecules and their biological functions involved in disease mechanisms. However, there are still challenges for biomolecular analyses due to their poor ionization efficiency, low concentration, and various modifications. This dissertation is devoted to the development of novel and improved MS-based omics methodologies and the application to human disease studies. Specifically, this work established a multiplex isobaric dimethylated leucine (DiLeu) strategy combined with advanced bioinformatics tool for metabolomics studies and extended its application to biomarker discovery in mouse models for lower urinary tract symptoms (LUTS). This work also developed a novel MS approach for comparative neuropeptide characterization with a decision-tree driven MS method and applied it to peptidomic analyses of human pituitary tumors and subcommissural organ. Moreover, metabolomics, peptidomics, and proteomics were integrated to investigate how the host immune system is shaped by human microbiome during infection. In summary, this work not only improves current analytical methodologies for characterizing and quantifying biomolecules but also demonstrates its potential applications in the pharmaceutical industry and clinical settings.

Chapter 1

Introduction and Research Summary



Introduction

While mass spectrometry (MS) has been proven to be a powerful analytical technique for biomolecule identification and quantification, analytical challenges are still present due to the biomolecules' poor ionization efficiency, low concentration, and various modifications. To address some of these challenges, a multiplex isobaric dimethylated leucine (DiLeu) strategy combined with an advanced bioinformatics tool was established in this dissertation for metabolomics studies; a decision-tree driven MS approach was also developed for comprehensive neuropeptide characterization; moreover, multi-omics integration method was investigated to study in vivo molecular network. These novel and improved MS-based methodologies have been successfully applied to lower urinary tract symptoms (LUTS) biomarker discovery, neural function exploration of human pituitary tumor and subcommissural organ (SCO), and human microbiome role investigation.

Research Summary

Chapter 1 provides a general introduction with summaries of the major findings of each study. **Chapters 2** and **3** focus on the MS-based LUTS biomarkers discovery (1, 2). LUTS are common among aging men. The etiology of LUTS is considered to be multifactorial (3), and age-related hormone change and prostatic inflammation are two plausible etiologies. While patients may suffer a variety of etiologic mechanisms, the current subjective symptom score diagnosis method usually could not distinguish the pathogenic factors from different individuals. As a result, patients are treated empirically with either medication or surgery. Objective biomarkers associated with specific etiologic mechanisms are greatly needed to refine the diagnostic approach and provide personalized treatment. The purpose of these two chapters is to characterize the contribution of

hormonal change and prostatic inflammation to the urine metabolome and proteome in mouse models of hormone-induced urinary obstruction (**Chapter 2**) and bacterial-induced prostatic inflammation (**Chapter 3**). In Chapter 2, accurate and high throughput quantification of amine-containing metabolites was achieved by twelve-plex DiLeu isobaric labeling (4). In Chapter 3, the investigation of DiLeu labeling on metabolites allowed metabolomics and proteomics analysis on the same LC-MS platform. Metandem (5) provides a useful tool for large-scale stable isotope labeling-based metabolomics data analysis. Moreover, significantly changed metabolites and proteins were found to be candidate biomarkers for prostatic inflammation induced-LUTS and hormone-induced urinary obstruction, which could be further used as the LUTS patient stratification.

Neuropeptides are known to modulate neuronal communication and play a key role in regulating the activities of the brain. Recently, MS has evolved as a powerful tool to characterize neuropeptides, due to its high speed, great sensitivity, and unbiased detection of a wide range of peptides present in a complex mixture. However, the comprehensive characterization of neuropeptides is challenging due to their low abundance, biochemical complexities, and limited neuropeptide database (6). Therefore, a sensitive MS method is in high demand, and every step in the workflow, from sampling to data analysis, needs to be optimized. Although various sample preparation methods have already been reported, a systematic comparative investigation is still lacking. Also, EThcD is a novel ion fragmentation scheme, and its performance on neuropeptide characterization has not been evaluated yet. In **Chapter 4**, a novel workflow was developed, integrating a comparison of three common neuropeptide extraction methods, EThcD ion preference investigation, followed by application to neuropeptide characterization in the human

pituitary tumor. Furthermore, novel neuropeptides were explored with a custom-made neuropeptide preprohormone database (7).

Chapters 5 also focuses on the neuropeptide study. The mammalian subcommissural organ (SCO) is a circumventricular organ located in the dorsocaudal region of the third ventricle. The functionality of the SCO is not fully understood, although it has been implicated in guiding the development of the brain and spinal cord (8). Since neuropeptides are neuronal signaling molecules that coordinate the activities of the brain and whole body, elucidating the functions of the SCO at the molecular level requires knowledge of its neuropeptidome, along with their critical post-translational modifications (PTMs). In this chapter, we performed the first neuropeptidomic analysis of the SCO, revealing unique information on the neuropeptide content and PTM modulation of the SCO, which will help to better understand the roles of the SCO in different biological processes (9).

In recent decades, neuropeptides also have been found to play a significant role in communication along the gut-brain axis. Various neuropeptides are expressed in the central and peripheral nervous systems, where they facilitate the crosstalk between the nervous systems and other major body systems. In addition to being critical to communication from the brain in the nervous systems, neuropeptides actively regulate immune functions in the gut in both direct and indirect ways, allowing for communication between the immune and nervous systems. **Chapter 6** is a mini-review. The role of several neuropeptides, including calcitonin gene-related peptide (CGRP), pituitary adenylate cyclase-activating polypeptide (PACAP), corticotropin-releasing hormone

(CRH) and phoenixin (PNX), in the gut-brain axis and their functions in immunity and stress, were discussed and summarized (10).

Chapters 7 integrates metabolomics, peptidomics, and proteomics analyses to investigate how the host immune system is shaped by the human microbiome during infections (11). The identified bioactive peptides (hormones) allow us to understand better how the human microbiome enhances host immune response through the gut-brain axis. Furthermore, this multi-omics analysis enabled by advanced bioinformatics tools helps to generate a more comprehensive metabolic pathway map involved in the immune response.

In collaboration with the Wisconsin Institute for Scientific Literacy, **Chapter 8** describes the LUTS biomarker discovery projects for a more general audience. Finally, **Chapter 9** concludes the thesis and discusses future research directions. Other interesting projects, including Metandem software development (5), metabolite absolute quantification method development, neuropeptidomics characterization of rat habenular nuclei (12), extraction optimization for gut microbiota multi-omics study, and multi-omics study of different gut microbiota environments are presented in **Appendix II, III, IV, V and VI, respectively.** All my publications and presentations resulting from this work and beyond, are summarized in **Appendix I.**

References

1. Wei P, Hao L, Ma F, Yu Q, Buchberger AR, Lee S, Bushman W, Li LL.Urinary Metabolomic and Proteomic Analyses in a Mouse Model of Prostatic Inflammation. *Journal of Urine, accepted.*

2. Wei P, Hao L, Thomas S, Buchberger AR, Hyman S, Strinkr L, Ricke WA, Li LL. Urinary Amine Metabolomics Characterization of a Hormone-Induced Urinary Obstruction Mouse Model with Custom 12-plex Isobaric DiLeu Labeling. *Journal of the American Society for Mass Spectrometry, under reviewed.*

3. Harley SJD, Wittert G, Brook NR, Secombe P, Campbell J, Lockwood C. 2017. Identifying predictors of change in the severity of untreated lower urinary tract symptoms in men. JBI Database Syst Rev Implement Reports 15:1585–1592.

4. Frost DC, Greer T, Li L. 2015. High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. Anal Chem 87:1646–1654.

5. Hao L, Zhu Y, Wei P, Johnson J, Buchberger A, Frost D, Kao WJ, Li L. 2019. Metandem: An online software tool for mass spectrometry-based isobaric labeling metabolomics. Anal Chim Acta 1088:99–106.

6. Li L, Sweedler J V. 2008. Peptides in the Brain: Mass Spectrometry–Based Measurement Approaches and Challenges. Annu Rev Anal Chem 1:451–483.

7. Wei P, Yu Q, Sun H, Ma F, Bakshi VP, Sun W, Zheng Z, Zeng Chun, Li LL. Development of a Novel Mass Spectrometry Approach for Comprehensive Neuropeptidome Characterization and its Application to the Analysis of Human Pituitary Tumor. *Journal of the American Chemical Society, to be submitted.*

8. Gobron S, Creveaux I, Meiniel R, Didier R, Dastugue B, Meiniel A. 1999. SCO-spondin is evolutionarily conserved in the central nervous system of the chordate phylum. Neuroscience 88:655–664.

9. Wei P, Ma F, Ge W, Li L. Neuropeptidomic Study of the Mammalian Subcommissural Organ (SCO) by High-Resolution Mass Spectrometry. *To be submitted.*

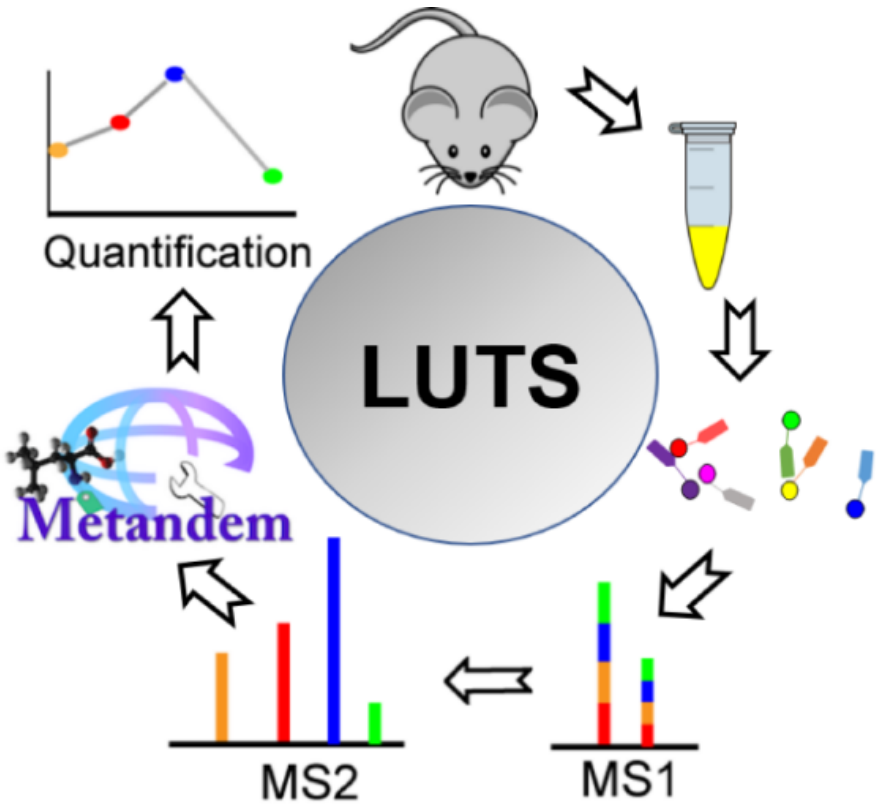
10. Wei P, Keller C, Li L. 2020. Neuropeptides in gut-brain axis and their influence on host immunity and stress. Comput Struct Biotechnol J 18:843–851.

11. Wei P, Keller C, Bratburd JR, Liu R, Vivas E, Gemperline E, Rey FE, Currie CR, Li LL. Integrated , Multi-Omics Strategy to Study the Host Immune Response Induced by Human Microbiome during Salmonella enterica Typhimurium and Candida albicans Infection. *To be submitted.*

12. Yang N, Anapindi KDB, Rubakhin SS, Wei P, Yu Q, Li L, Kenny PJ, Sweedler J V. 2018. Neuropeptidomics of the Rat Habenular Nuclei. J Proteome Res 17:1463–1473.

Chapter 2

Urinary Amine Metabolomics Characterization of a Hormone-Induced Urinary Obstruction Mouse Model with Custom 12-plex Isobaric DiLeu Labeling



Adapted from **Wei P**, Hao L, Thomas S, Buchberger AR, Hyman S, Strinkr L, Ricke WA, Li LL. "Urinary Amine Metabolomics Characterization of a Hormone-Induced Urinary Obstruction Mouse Model with Custom 12-plex Isobaric DiLeu Labeling. " *Journal of the American Society for Mass Spectrometry, under reviewed.* **Wei P**, designed and conducted the study under the supervision of LL; Hao L provided expertise in metabolomics, Buchberger AR systhesized the DiLeu tags, Strinkr L provide the mouse urine sample, Thomas S, Hyman S, and Ricke WA provided insights in LUTS.

ABSTRACT

Lower urinary tract symptoms (LUTS) are common in aging males. Disease etiology is largely unknown, but likely includes inflammation and age-related changes in steroid hormones. Diagnosis is currently based on subjective symptom scores, and mainstay treatments can be ineffective and bothersome. Biomarker discovery efforts could facilitate objective diagnostic criteria for personalized medicine and new potential druggable pathways. To identify urine metabolite markers specific to hormone-induced bladder outlet obstruction, we applied our custom synthesized multiplex isobaric tags to monitor the development of bladder outlet obstruction across time in an experimental mouse model of LUTS. Mouse urine samples were collected before treatment and after 2, 4, and 8 weeks of steroid hormone treatment and subsequently analyzed by nanoflow ultrahigh-performance liquid chromatography coupled to tandem mass spectrometry. Accurate and high throughput quantification of amine-containing metabolites was achieved by twelve-plex DiLeu isobaric labeling. Metandem, a novel online software tool for large-scale isobaric labeling-based metabolomics, was used for identification and relative quantification of labeled metabolites. A total of 59 amine-containing metabolites were identified and quantified, 9 of which were changed significantly by the hormone treatment. Metabolic pathway analysis showed three metabolic pathways were potentially disrupted. Among them, the arginine and proline metabolism pathway was significantly dysregulated both in this model and in a prior analysis of LUTS patient samples. Proline and citrulline were significantly changed in both samples and serve as attractive candidate biomarkers. 12-plex DiLeu isobaric labeling with Metandem data processing presents an accessible and efficient workflow for amine-containing metabolome study in biological specimens.

KEYWORDS: urine, metabolomics, lower urinary tract symptoms, mass spectrometry, isobaric labeling, Metandem

INTRODUCTION

Lower urinary tract symptoms (LUTS) is a major public health problem in the aging population. Symptoms include frequency, urgency, nocturia, weak urinary stream, straining to void, and a sense of incomplete emptying (1–3). Recent studies have shown that the etiology of LUTS is multi-factorial (4, 5). In men, prostate enlargement, prostatic inflammation, age-related changes in detrusor function and steroid hormones, and heightened sensitivity to bladder filling may all contribute (4–8). The complex and variegated composition of the LUTS patient population makes it challenging to tease out and validate contributions of different mechanisms. Currently no objective biomarkers exist to inform treatment strategies, so patients are generally treated with drugs to decrease prostate size (5α-reductase inhibitors) or relax smooth muscle (alpha blockers) (9, 10). These treatments are not completely effective, durable in response, nor are they without complication (11). Patients with LUTS that persists after these mainline treatments may undergo invasive transurethral resection of the prostate (12). Overall, better biomarkers and treatment options need to be found. Here, we attempt to characterize the contribution of hormone changes to the urine metabolome under controlled, experimental conditions in a mouse model of bladder outlet obstruction (BOO). Understanding the urine signature of hormone-induced BOO will inform both future efforts to study druggable pathways and future efforts to validate LUTS biomarkers for patient stratification and personalized treatments.

Our hormone-induced BOO mouse model relies on slow-release, subcutaneous implants of testosterone and 17β-estradiol in adult male C57/BL6 mice to generate marked increases in urinary frequency, bladder volume, bladder mass, and prostate mass as well as decreased urinary volume/void before the 8 week timepoint (13). Additionally, this model, if treated longer (4

months), develops prostatic intraepithelial neoplasia lesions, making this model useful for the study of carcinogenesis in the prostate (14). To characterize the urine metabolome of hormone-induced BOO, we collected samples across the development of this phenotype: before treatment and after 2, 4, and 8 weeks of treatment.

Tandem mass spectrometry (MS/MS)-based relative quantification by isobaric labeling is a useful technique for comparative quantitative metabolomics in biological systems (15–17). *N*,*N*-dimethyl leucine (DiLeu) labels are isobaric. This means that analyte precursor *m/z* between channels is indistinguishable at low MS resolution, but distinct reporter ions are apparent in the low *m/z* region upon MS^2 fragmentation. The intensities of these reporter ions in MS^2 spectra reflect the labeled metabolites' abundance in each sample and, thus, can be compared for relative quantification. Compared to label-free metabolomics, multiplexed isobaric labeling of metabolites greatly reduces run-to-run variation, enhances ionization efficiency, improves chromatographic separation of polar metabolites, and decreases instrument time demand (by 12-fold when using 12-plex DiLeu labeling) (18). DiLeu utilizes a triazine ester to label primary and secondary amines, making this derivatization scheme applicable to many metabolites. Additionally, DiLeu labeling enables polar metabolites to be separated and detected on nanoUPLC systems, improving chromatographic resolution and detection sensitivity over typical standard flow separation techniques (19). Twelve-plex DiLeu isobaric labels, which are designed and synthesized in our lab (20, 21), were employed here for the relative quantification of amine-containing metabolites in urine samples via nanoflow ultrahigh-performance liquid chromatography coupled to tandem mass spectrometry (nanoUPLC-MS/MS).

Isobaric labeling has been widely adopted for quantitative proteomics and peptidomics, but it has only recently been applied to metabolomics analyses (22–24). Thus, data processing workflows for isobaric labeling in proteomics and peptidomics have matured more rapidly than for metabolomics. To date, few metabolomics software tools can process stable isotope label-based metabolomics data, particularly when using reporter ions produced by MS/MS for quantification. Therefore, our lab developed a novel online software tool for isobaric labeling-based metabolomics, called Metandem, which integrates metabolite quantification, identification, and statistical analysis in the same software package and is freely available at http://metandem.com/web/ (25). Metandem is also the first omics data analysis software to provide straightforward, online parameter optimization functionality for customization to a particular datasets (25). Here, we employed the Metandem software tool to analyze the 12-plex DiLeu-labeled urinary metabolites at multiple time points after hormone treatment. Metabolite identification, quantification, and statistical analysis were achieved in less than 15 min using our Metandem software tool. It is expected that these powerful tools may identify clinically useful biomarkers of hormone-induced bladder outlet obstruction and new targets for drug treatment.

MATERIALS AND METHODS

Hormone-induced urinary obstruction mouse model

All animal procedures were approved by the University of Wisconsin-Madison Animal Care and Use Committee. Adult male C57BL/6 mice were used for this study (n = 3) (Charles River, Wilmington, MA). Urine was collected from each animal via metabolic cage (26) for a two-hour period. After the pre-treatment urine collection, mice were treated with subcutaneous, slow-release implants containing compressed testosterone (T, 25 mg) and 17β-estradiol (E2, 2.5 mg + 22.5 mg

cholesterol binder), as described previously (13). Urine samples from the same mice were then collected at 2, 4, and 8 weeks post-treatment, as above. All samples were stored at -80 °C until further processing.

Mouse urine sample preparation

Urine samples were thawed on ice and mixed well by vortexing. Raw urine (220 μL) was centrifuged at 10,000 *xg* for 10 min to remove particulates and cellular debris. Centrifugal filters (3 kDa, Millipore Amicon Ultra, Burlington, MA) were pre-rinsed 3 times with 500 μL of Milli-Q water at 14,000 *xg* for 20 min. Clarified supernatant (200 μL) was added to the filter unit and centrifuged at 14,000 *xg* for 30 min, followed by 2 rinses with Milli-Q water (200 μL; same centrifugation speed and time) to obtain the metabolite fraction (~600 μL total). Osmolality was determined via a freezing-point depression osmometer (Osmometer Model 3250, Advanced Instruments, Norwood, MA) (27) and all samples were normalized to 50 mOsmoles/kg H_2O, aliquoted, lyophilized, and stored at -80 °C until labeling.

DiLeu synthesis and labeling procedure

Twelve-plex isobaric DiLeu reagents were synthesized and used for labeling reaction as previously described (28). Briefly, the DiLeu 12-plex reagents were synthesized, aliquoted in inactivated form at 4 ℃ in a dry box and activated prior to labeling. Lyophilized urine metabolite samples were redissolved in 0.5 M triethylammonium bicarbonate solution prior to derivatization of primary and secondary amines by excess activated DiLeu reagent (**Fig. 1D**). The organic: aqueous ratio was maintained at ~70% via anhydrous dimethylformamide. Reactions were allowed to proceed for 2 hrs at room temperature with vigorous vortexing. Labeling reactions were quenched with 0.25% hydroxylamine (v/v), and labeled samples were combined in equal ratios (v/v) to form pooled 12-

plex samples. Excess DiLeu reagents were removed from pooled 12-plex samples via SCX Ziptips (OMIX-SCX, Agilent, Santa Clara, CA) as previously described (19) before lyophilization and storage at -80 °C until analysis.

LC-ESI-MS analysis

Twelve-plex pooled samples were reconstituted in 3% acetonitrile, 0.1% formic acid (v/v) in water before injection. UPLC-MS/MS analysis was conducted using a Thermo Dionex UltiMate^{TM} 3000 nanoLC system coupled to a Thermo Q Exactive^{TM} HF Orbitrap MS. The analytical column was self-made with an integrated emitter tip and dimensions of 75 μm inner diameter × 15 cm length, packed with 1.7 μm, 150 Å, BEH C_{18} material (Waters, Milford, MA). Mobile phase A was 0.1% formic acid in water, and mobile phase B was 0.1% formic acid in acetonitrile (Optima Solvents, Thermo, Waltham, MA). The flow rate was 0.3 μL/min, and the 70-min gradient was as follows: 0-16 min, 3% solvent B; 16-20 min, 3-25% B; 20-30 min, 25-45% B; 30-50 min, 45-70% B; 50-56 min, 70-95% B; 56-60 min 95% B; 60-60.5 min, 95-3% B; 60.5-70 min, 3% B. Full MS scans were acquired from *m/z* 180 to 1000 at a resolution of 60 K, automatic gain control (AGC) at 1×10^6, and maximum injection time of 50 ms. The top 20 precursors were selected for higher-energy C-trap dissociation tandem mass spectrometry (HCD MS^2) analysis with an isolation window of 1 *m/z*, normalized collision energy (NCE) of 30, resolving power of 60k, AGC target of 1×10^5, maximum injection time of 30 ms, and a lower mass limit of 110 *m/z*.

Data analysis

Raw data files were acquired using Thermo Scientific Xcalibur^{TM} software and converted into .txt format via the COMPASS software suite (29). Metandem was then used to batch-process three technical replicates of each 12-plex sample for metabolite quantification. Accurate mass of

reporter ions was obtained by averaging the mass across several MS/MS spectra. Then, the mass tolerance for metabolite identification was 20 ppm. DiLeu label purity was predetermined, and correction for each channel was performed as previously described (19). Average precursor mass shift due to labeling was 145.1280 Da. Data analysis parameters, such as reporter ion mass tolerance, batch mass tolerance, and retention time tolerance were optimized using the parameter optimization graphs in the Metandem software. Output files with reporter ion information were merged and median-normalized. Repeated-measures analysis of variance (RM-ANOVA) (30, 31), which helps to eliminate some of the interindividual variability, was used to compare metabolite abundance between timepoints. Metabolites were considered significantly modulated when *p*-values were < 0.05 (RM-ANOVA). MetaboAnalyte 4.0 software (32) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database were used for metabolic pathway analysis. Pathways were considered dysregulated when more than two metabolites were identified in the pathway and at least one was significantly changed due to the treatment. Candidate biomarkers and dysregulated pathways were compared with prior patient analyses and other previous reports.

RESULTS AND DISCUSSION

Efficacy of twelve-plex DiLeu for metabolomics

Twelve-plex DiLeu isobaric labeling allowed multiplexed metabolite identification and quantification in mouse urine samples while also reducing instrumentation time demand, decreasing run-to-run variation, and also improving quantification accuracy. The same labeled metabolite from 12 urine samples showed a single peak in the MS^1 spectrum with a mass shift of 145.1280 Da (**Fig. 1A**). For this peak, twelve distinct reporter ion peaks are present in the MS^2 low *m/z* region (**Fig. 1B**). The intensity of each reporter ion in MS^2 spectra reflects the labeled metabolites' abundance in each sample and, thus, can be compared for relative quantification (**Fig.**

1C). The absence of unlabeled metabolite ions in the MS1 (mass difference of 145.1280 compared to unlabeled counterparts) suggested highly efficient labeling of the metabolites by the DiLeu reagent.

Metandem parameter optimization

Metandem is a newly developed custom software platform for large-scale stable isotope labeling-based metabolite identification and quantification. It is also the first omics data analysis software that contains functionality to perform online parameter optimization for customization to a dataset. Results of automated parameter optimization were as follows: optimal reporter ion mass tolerance, 0.5 mDa (**Fig. 2A**); optimal batch processing mass tolerance, 5 ppm **(Fig. 2B**); and optimal batch processing retention time tolerance, 0.5 min (**Fig. 2C**). Metandem also provides the histogram for retention time (**Fig. S1A**) and detected precursor (**Fig. S1B**) distribution results.

Mouse urine metabolite identification and quantification

Three technical replicates of each 12-plex injection were merged in Metandem. A total of 312 features were identified as putative metabolites after accurate mass matching against the Human Metabolome Database (HMDB). Of these, 59 were primary or secondary amine-containing metabolites (**Table S1**). Thirty-seven of these 59 were documented in the urine metabolome database (33), and the twenty-two additional amine-containing metabolites were, to our knowledge, first reported here. After comparing metabolite abundance between timepoints, 9 metabolites were identified as statistically significant biomarker candidates (RM-ANOVA p-value < 0.05) (**Table 1**). Among them, eight metabolites were generally increased at all timepoints, while only one metabolite was generally decreased at all timepoints (**Fig. 3**). For the increased expression patterns, leucine and 5-aminopentanamide have the highest concentration at 2 weeks; *N*-acetylputrescine at

4 weeks; and proline, citrulline, *D*-alanyl-*D*-alanine, O-phosphohomoserine, and 2-Aminobenzoic acid peaked at 8 weeks.

Metabolic pathway analysis

Metabolic pathway analysis is based on the association between identified metabolites and their related biological processes (34). Herein, all identified metabolites were input into the MetaboAnalyte 4.0 software for metabolic pathway analysis. Three potentially perturbed metabolic pathways were identified: (1) the arginine and proline metabolism pathway; (2) the aminoacyl-tRNA biosynthesis pathway; and (3) the tryptophan metabolism pathway. (Fisher's exact test, p-value < 0.07) (**Table 2**).

Comparison with LUTS patient samples and other previous reports

Prior analyses have demonstrated some of the metabolomic and immunohistochemical features of LUTS in patient samples, which represent processes like fibrosis and inflammation (27). Both LUTS patients and this hormone-induced mouse model of BOO show perturbation of the arginine and proline metabolic pathway in the urine metabolome. In particular, the candidate biomarkers citrulline and proline were significantly changed in both LUTS and this mouse model (**Fig. 4**). For proline, it was increased in both LUTS patients and the hormone-treated mice, represents a strong candidate biomarker for hormone-induced BOO. Urinary proline was also increased in a mouse model of hepatic injury and fibrosis, indicating fibrosis is a potential relevant pathway in the present BOO model (35). Next, citrulline was decreased in LUTS patients but increased in this mouse model. Citrulline is poorly understood in the context of prostate diseases. Citrulline is a non-essential amino acid, a precursor to arginine, and displays antioxidant properties (36). Increased urine citrulline in these mice could be explained by inflammation-induced expression of

nitric oxide synthase 2, the enzyme responsible for conversion of arginine to citrulline (37), possibly in response to oxidative stress generated by catechol estrogen metabolites (38).

Other processes were significantly changed in the present study but not observed in LUTS patients. For example, the aminoacyl-tRNA biosynthesis pathway was disrupted in the urine metabolome of this mouse model but not in LUTS patients. This pathway is pivotal in determining how the genetic code is interpreted as amino acids (39). More specifically, leucine and proline were significantly changed in this metabolic pathway. The tryptophan metabolism pathway, represented by 2-aminobenzoic acid, was also disrupted in this mouse model. Overall, these pathway-level results were similar to a prior report of urine metabolomics in liver injury and fibrosis (40). Interestingly, leucine may play an important role in prostatic proliferation. Leucine is an essential branched-chain amino acid that signals through the mTOR pathway (41). This signaling is pro-proliferative in prostate cancer cells, and as such decreasing leucine transport into tumors is an attractive therapeutic target (42). It is possible the strong initial increases in leucine in this model contribute to the increased prostate mass (benign hyperplasia) or even the development of prostatic intraepithelial neoplasia observed later in this model (14).

CONCLUSIONS

Twelve-plex DiLeu isobaric labeling is an attractive high-throughput strategy for identification and quantification of amine-containing metabolites, and Metandem is a useful tool for large-scale stable isotope labeling-based metabolomics data analysis. Paired together, these tools offer a powerful and accessible method for relative quantification of amine-containing metabolites in disease biomarker research. In this study of urinary amine metabolomics of a hormone-induced

LUTS mouse model, we have identified and quantified 59 amine metabolites, and 22 of them were identified in urine for the first time. LUTS patients and this mouse model shared common pathways that are dysregulated compared to control groups, for instance, the arginine and proline metabolism pathway. Proline presents an especially attractive candidate biomarker for hormone-induced BOO, as it was significantly increased in both human LUTS and this mouse model. Future experiments will test the hypothesis that this hormone treatment results in fibrosis of the lower urinary tract, ultimately leading to the pronounced BOO phenotype.

ACKNOWLEDGEMENTS

The authors would like to thank DiLeu synthesis team in the Li Lab (Dr. Tyler Greer, Dr. Dustin Frost, and Dr. Amanda Rae Buchberger). The authors would also like to thank Seth Hyman for his assistance with urine sample collection and Zeeh Pharmaceutical Experiment Station in School of Pharmacy for instrument access. This work was supported in part by National Institutes of Health through Grants 1P20 DK097826, U54 DK104310, R01DK071801, RF1AG052324, P41GM108538, R01 DK093690, R01 DK099328, DK091193, and DK104310. The Q-Exactive Orbitrap instrument was purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531). L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

REFERENCES

1. Chapple CR, Roehrborn CG. 2006. A Shifted Paradigm for the Further Understanding, Evaluation, and Treatment of Lower Urinary Tract Symptoms in Men: Focus on the Bladder. Eur Urol 49:651–659.

2. Abrams P, Cardozo L, Fall M, Griffiths D, Rosier P, Ulmsten U, Van Kerrebroeck P, Victor

A, Wein A. 2003. The standardisation of terminology in lower urinary tract function: report from the standardisation sub-committee of the International Continence Society. Urology 61:37–49.

3. Speakman M, Kirby R, Doyle S, Ioannou C. 2015. Burden of male lower urinary tract symptoms (LUTS) suggestive of benign prostatic hyperplasia (BPH) - focus on the UK. BJU Int 115:508–519.

4. Harley SJD, Wittert G, Brook NR, Secombe P, Campbell J, Lockwood C. 2017. Identifying predictors of change in the severity of untreated lower urinary tract symptoms in men: a systematic review protocol. JBI database Syst Rev Implement reports 15:1585–1592.

5. Edgar AD, Levin R, Constantinou CE, Denis L. 2007. A critical review of the pharmacology of the plant extract of *Pygeum africanum* in the treatment of LUTS. Neurourol Urodyn 26:458–463.

6. Swerdloff RS, Wang C. 2004. Androgens and the ageing male. Best Pract Res Clin Endocrinol Metab 18:349–362.

7. Nickel JC, Roehrborn CG, Castro-Santamaria R, Freedland SJ, Moreira DM. 2016. Chronic Prostate Inflammation is Associated with Severity and Progression of Benign Prostatic Hyperplasia, Lower Urinary Tract Symptoms and Risk of Acute Urinary Retention. J Urol 196:1493–1498.

8. Noguchi N, Chan L, Cumming RG, Blyth FM, Handelsman DJ, Seibel MJ, Waite LM, Le Couteur DG, Naganathan V. 2016. Lower Urinary Tract Symptoms and Incident Falls in Community Dwelling Older Men: The Concord Health and Ageing in Men Project. J Urol 196:1694–1699.

9. Kim EH, Brockman JA, Andriole GL. 2018. The use of 5-alpha reductase inhibitors in the treatment of benign prostatic hyperplasia. Asian J Urol 5:28–32.

10. Lepor H. 2007. Alpha blockers for the treatment of benign prostatic hyperplasia. Rev Urol 9:181–90.

11. Ganzer CA, Jacobs AR, Iqbal F. 2015. Persistent Sexual, Emotional, and Cognitive Impairment Post-Finasteride. Am J Mens Health 9:222–228.

12. Cornu J-N, Ahyai S, Bachmann A, de la Rosette J, Gilling P, Gratzke C, McVary K, Novara G, Woo H, Madersbacher S. 2015. A Systematic Review and Meta-analysis of Functional Outcomes and Complications Following Transurethral Procedures for Lower Urinary Tract Symptoms Resulting from Benign Prostatic Obstruction: An Update. Eur Urol 67:1066–1096.

13. Nicholson TM, Ricke EA, Marker PC, Miano JM, Mayer RD, Timms BG, vom Saal FS, Wood RW, Ricke WA. 2012. Testosterone and 17β-Estradiol Induce Glandular Prostatic Growth, Bladder Outlet Obstruction, and Voiding Dysfunction in Male Mice. Endocrinology 153:5556–5565.

14. Ricke WA, McPherson SJ, Bianco JJ, Cunha GR, Wang Y, Risbridger GP. 2008. Prostatic hormonal carcinogenesis is mediated by *in situ* estrogen production and estrogen receptor alpha signaling. FASEB J 22:1512–1520.

15. Xiang F, Ye H, Chen R, Fu Q, Li N. 2010. N,N-Dimethyl leucines as novel Isobaric tandem mass tags for quantitative proteomics and peptidomics. Anal Chem 82:2817–2825.

16. Gritsenko MA, Xu Z, Liu T, Smith RD. 2016. Large-scale and deep quantitative proteome

profi ling using isobaric labeling coupled with two-dimensional lc- ms/ms. Methods Mol Biol 1410:237–247.

17. Rauniyar N, Yates JR. 2014. Isobaric labeling-based relative quantification in shotgun proteomics. J Proteome Res. American Chemical Society.

18. Ong S-E, Mann M, Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R, Li J, Steen H, Hansen KC, Schmitt-Ulms G, Chalkley RJ, Hirsch J, Baldwin MA, Burlingame AL. 2005. N,N-Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. Mol Cell Proteomics 1:2817–2825.

19. Hao L, Zhong X, Greer T, Ye H, Li L. 2015. Relative quantification of amine-containing metabolites using isobaric N,N-dimethyl leucine (DiLeu) reagents via LC-ESI-MS/MS and CE-ESI-MS/MS. Analyst 140:467–75.

20. Frost DC, Li L. 2016. High-throughput quantitative proteomics enabled by mass defect-based 12-plex dileu isobaric tags. Methods Mol Biol 1410:169–194.

21. Frost DC, Rust CJ, Robinson RAS, Li L. 2018. Increased N,N-Dimethyl Leucine Isobaric Tag Multiplexing by a Combined Precursor Isotopic Labeling and Isobaric Tagging Approach. Anal Chem 90:10664–10669.

22. Hao L, Johnson J, Lietz CB, Buchberger A, Frost D, Kao WJ, Li L. 2017. Mass defect-based n,n-dimethyl leucine labels for quantitative proteomics and amine metabolomics of pancreatic cancer cells. Anal Chem 89:1138–1146.

23. Moulder R, Bhosale SD, Goodlett DR, Lahesmaa R. 2018. Analysis of the plasma proteome using iTRAQ and TMT-based Isobaric labeling. Mass Spectrom Rev 37:583–606.

24. Zhang L, Elias JE. 2017. Relative protein quantification using tandem mass tag mass spectrometry, p. 185–198. *In* Methods in Molecular Biology. Humana Press Inc.

25. Hao L, Zhu Y, Wei P, Johnson J, Buchberger A, Frost D, Kao WJ, Li L. 2019. Metandem: An online software tool for mass spectrometry-based isobaric labeling metabolomics. Anal Chim Acta 1088:99–106.

26. Li X, Knight J, Fargue S, Buchalski B, Guan Z, Inscho EW, Liebow A, Fitzgerald K, Querbes W, Todd Lowther W, Holmes RP. 2016. Metabolism of13C5-hydroxyproline in mouse models of Primary Hyperoxaluria and its inhibition by RNAi therapeutics targeting liver glycolate oxidase and hydroxyproline dehydrogenase. Biochim Biophys Acta - Mol Basis Dis 1862:233–239.

27. Hao L, Greer T, Page D, Shi Y, Vezina CM, Macoska JA, Marker PC, Bjorling DE, Bushman W, Ricke WA, Li L. 2016. In-Depth Characterization and Validation of Human Urine Metabolomes Reveal Novel Metabolic Signatures of Lower Urinary Tract Symptoms. Sci Rep 6:1–11.

28. Frost DC, Greer T, Li L. 2015. High-Resolution Enabled 12-Plex DiLeu Isobaric Tags for Quantitative Proteomics. Anal Chem 87:1646–1654.

29. Craig D. Wenger, Douglas H. Phanstiel, M. Violet Lee, Derek J. Bailey JJC. 2011. COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. Proteomics 92:1064–1074.

30. Lee KB, Lim SH, Kim KH, Kim KJ, Kim YR, Chang WN, Yeom JW, Kim YD, Hwang BY. 2015. Six-month functional recovery of stroke patients: a multi-time-point study. Int J Rehabil Res 38:173–80.

31. Mirowsky JE, Peltier RE, Lippmann M, Thurston G, Chen L-C, Neas L, Diaz-Sanchez D, Laumbach R, Carter JD, Gordon T. 2015. Repeated measures of inflammation, blood pressure, and heart rate variability associated with traffic exposures in healthy adults. Environ Heal 14:66.

32. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. 2018. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res 46:W486–W494.

33. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, Dame ZT, Poelzer J, Huynh J, Yallou FS, Psychogios N, Dong E, Bogumil R, Roehring C, Wishart DS. 2013. The Human Urine Metabolome. PLoS One 8:e73076.

34. Schilling CH, Schuster S, Palsson BO, Heinrich R. 1999. Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-genomic Era. Biotechnol Prog 15:296–303.

35. Gou X, Tao Q, Feng Q, Peng J, Zhao Y, Dai J, Wang W, Zhang Y, Hu Y, Liu P. 2013. Urine metabolic profile changes of CCl4-liver fibrosis in rats and intervention effects of Yi Guan Jian Decoction using metabonomic approach. BMC Complement Altern Med 13:123.

36. Allerton T, Proctor D, Stephens J, Dugas T, Spielmann G, Irving B, Allerton TD, Proctor DN, Stephens JM, Dugas TR, Spielmann G, Irving BA. 2018. l-Citrulline Supplementation: Impact on Cardiometabolic Health. Nutrients 10:921.

37. Wijnands K, Castermans T, Hommen M, Meesters D, Poeze M, Wijnands KAP, Castermans TMR, Hommen MPJ, Meesters DM, Poeze M. 2015. Arginine and Citrulline and the

Immune Response in Sepsis. Nutrients 7:1426–1463.

38. Bhat HK, Calaf G, Hei TK, Loya T, Vadgama J V. 2003. Critical role of oxidative stress in estrogen-induced carcinogenesis. Proc Natl Acad Sci U S A 100:3913–8.

39. Ibba M, Söll D. 2000. Aminoacyl-tRNA Synthesis. Annu Rev Biochem 69:617–650.

40. Chang H, Meng H, Liu S, Wang Y, Yang X, Lu F, Wang H. 2017. Identification of key metabolic changes during liver fibrosis progression in rats using a urine and serum metabolomics approach. Sci Rep 7:11433.

41. Zhang S, Zeng X, Ren M, Mao X, Qiao S. 2017. Novel metabolic and physiological functions of branched chain amino acids: a review. J Anim Sci Biotechnol 8:10.

42. Otsuki H, Kimura T, Yamaga T, Kosaka T, Suehiro J, Sakurai H. 2017. Prostate Cancer Cells in Different Androgen Receptor Status Employ Different Leucine Transporters. Prostate 77:222–233.

TABLES AND FIGURES

Table 1. Nine significantly changed urine metabolites after hormone treatment.

Compound	Molecular Weight	p-value	Expression Pattern	ppm	tr (min)	HMDB_ID
Proline	115.0623	0.006	Increase	9.2	13.6	HMDB00162
5-Aminopentanamide	116.0938	0.029	Increase	10.2	14.4	HMDB12176
N-Acetylputrescine	130.1108	0.042	Increase	1.2	15.2	HMDB02064
Citrulline	175.0936	0.005	Increase	11.7	15.5	HMDB00904
D-Alanyl-D-alanine	160.0827	0.005	Increase	13.1	16.5	HMDB03459
O-Phosphohomoserine	199.0278	0.001	Increase	16.1	16.7	HMDB03484
Purine	120.0450	0.001	Decrease	11.3	8.6	HMDB01366
2-Aminobenzoic acid	137.0464	0.023	Increase	9.7	19.0	HMDB01123
Leucine	131.0942	0.006	Increase	3.0	19.0	HMDB00687

Table 2. Potentially disrupted metabolic pathways via MetaboAnalyte 4.0 and KEGG pathway analysis, metabolites highlighted with red bold font are significantly changed metabolites (*p*-value is from Fisher's exact test).

Metabolic Pathway	KEGG ID	Matched Metabolites	*p*-value
Arginine and proline metabolism	Map00330	citrulline, N-acetylputrescine, proline, glutamic acid, creatine, GABA, 4-aminobutyraldehyde	4.90E-05
Aminoacyl-tRNA biosynthesis	Map00970	leucine, proline, cysteine, glycine, alanine, glutamic acid	4.88E-03
Tryptophan metabolism	Map00380	2-aminobenzoic acid, 3-hydroxyanthranilic acid, 5-hydroxyindoleacetic acid	6.81E-02

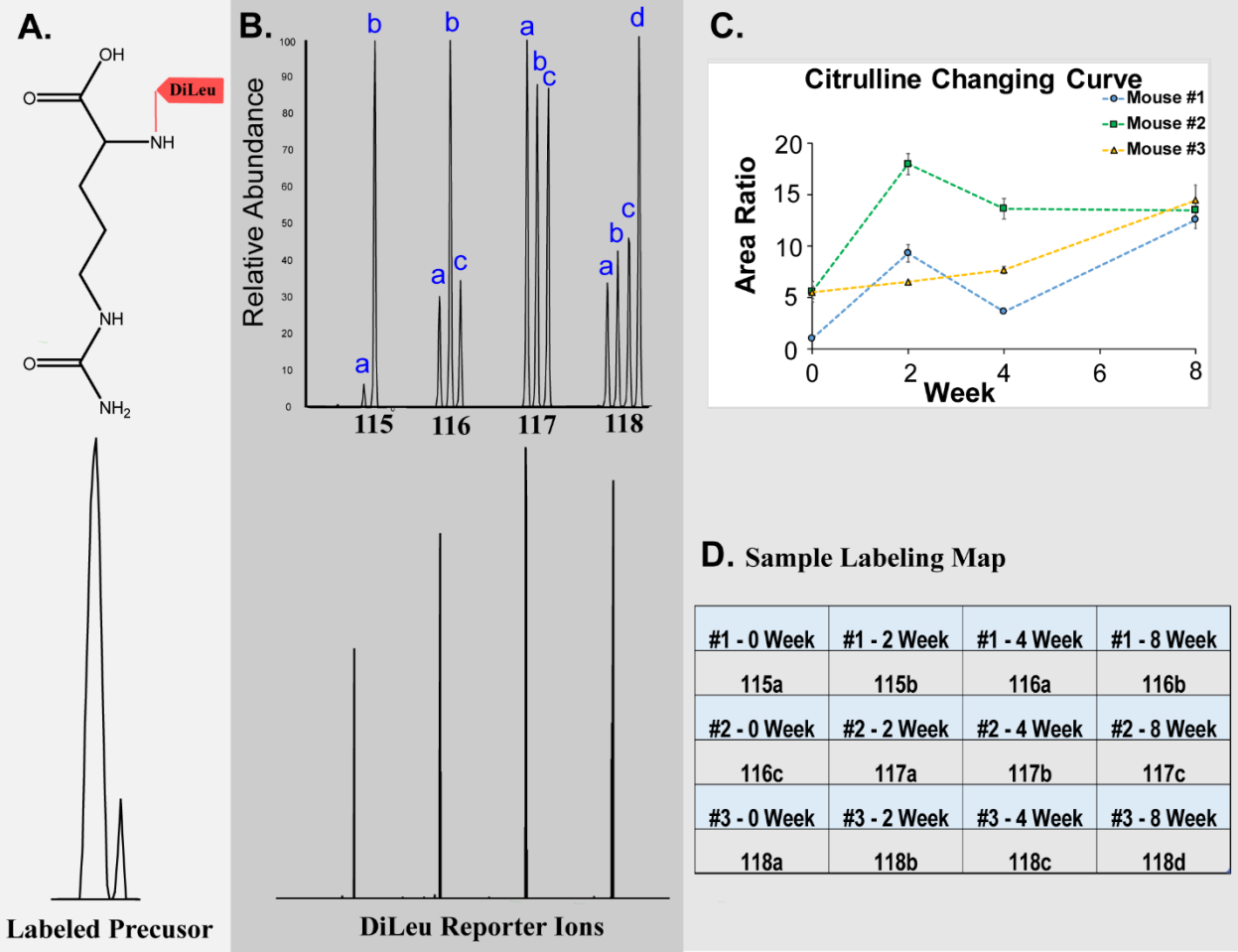
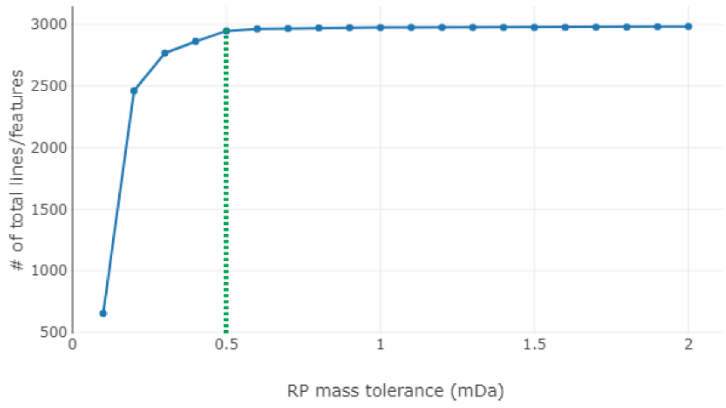
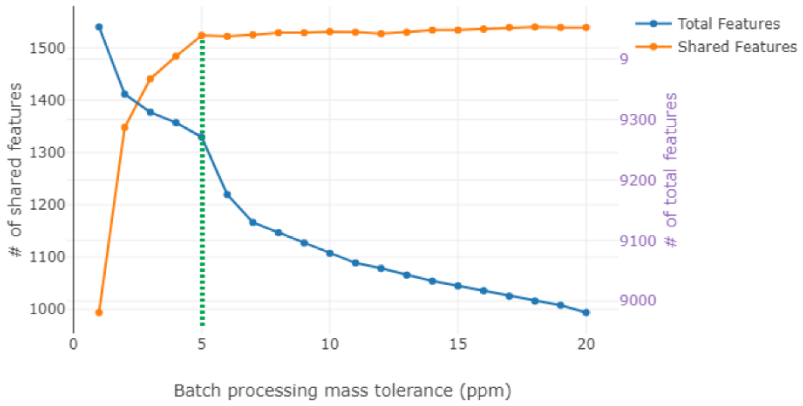


Figure 1. Details of 12-plex DiLeu labeling. A: Precursor ion of DiLeu-labeled citrulline; **B:** low *m/z* region showing DiLeu reporter channels at high resolution (top) and zoom in spectra (bottom); **C:** Citrulline changing trend from different time points of three biological replicates. **D:** Sample labeling map showing the 12-plex DiLeu tags and time point (randomized)

A. Optimizing reporter ion mass tolerance



B. Optimizing batch processing mass tolerance



C. Optimizing batch processing retention time tolerance

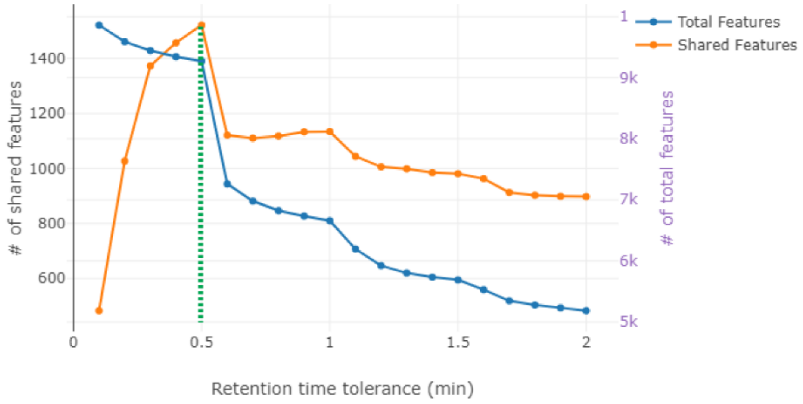


Figure 2. Metandem parameter optimization results: A: Optimizing reporter ion mass tolerance (0.5 mDa); **B:** Optimizing batch processing retention time tolerance (5 ppm); **C:** Optimizing batch processing retention time tolerance (0.5 min).

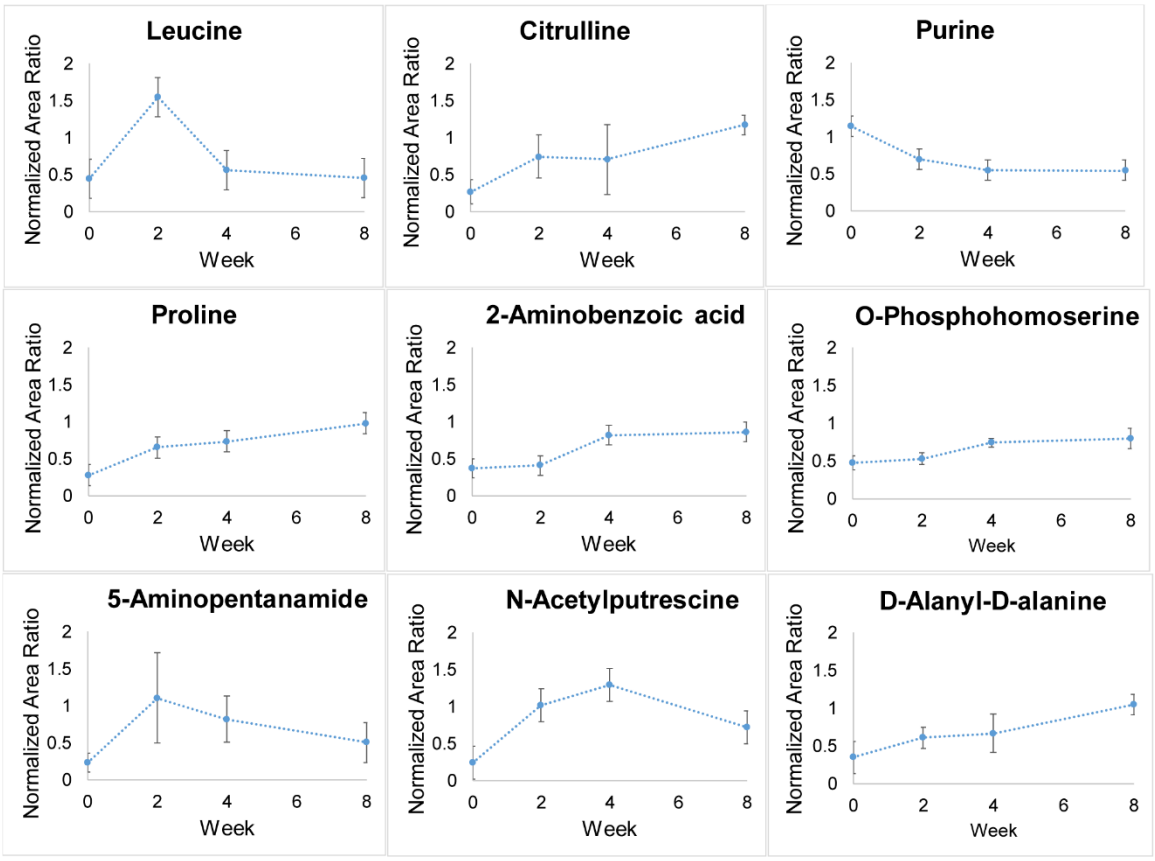


Figure 3. Quantification trends of 9 significantly changed urine metabolites (*n* = 3; RM-ANOVA, *p* < 0.05).

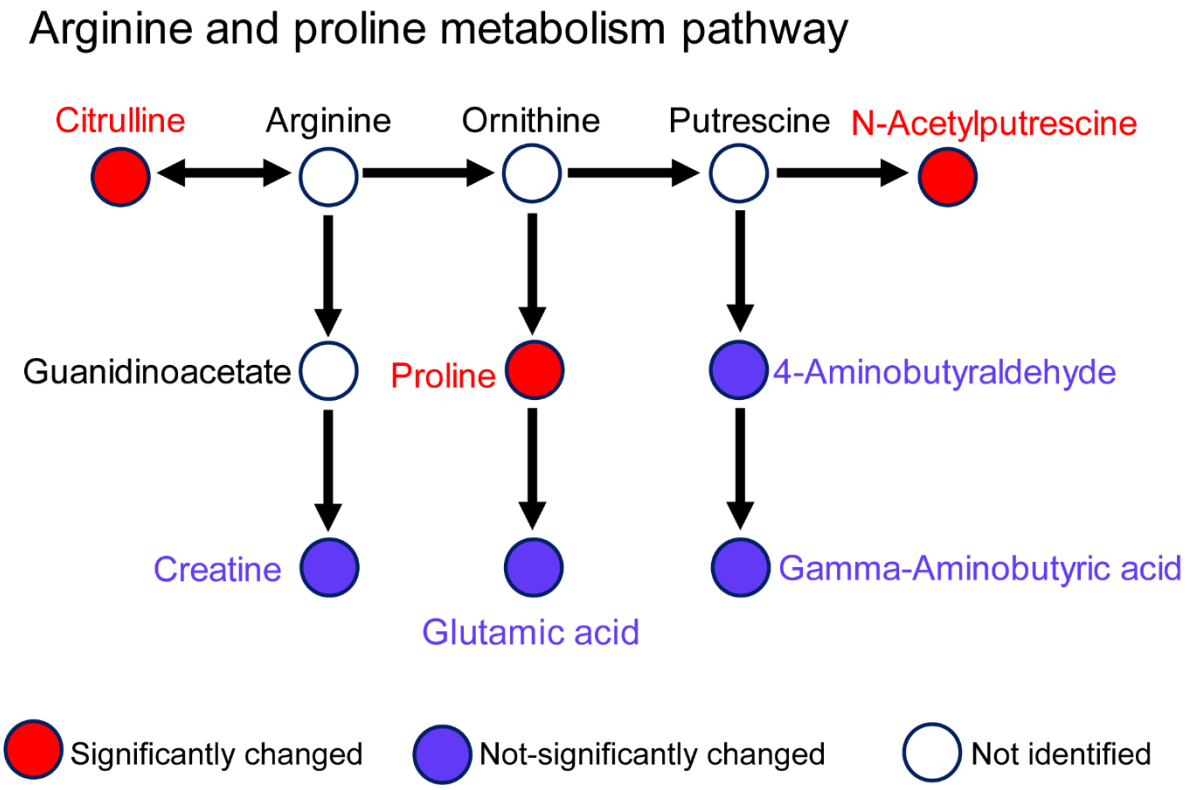


Figure 4. Arginine and proline metabolism pathway is potentially disrupted (MetaboAnalyte, KEGG; Fisher's exact test, $p < 0.07$).

SUPPLEMENTARY INFORMATION

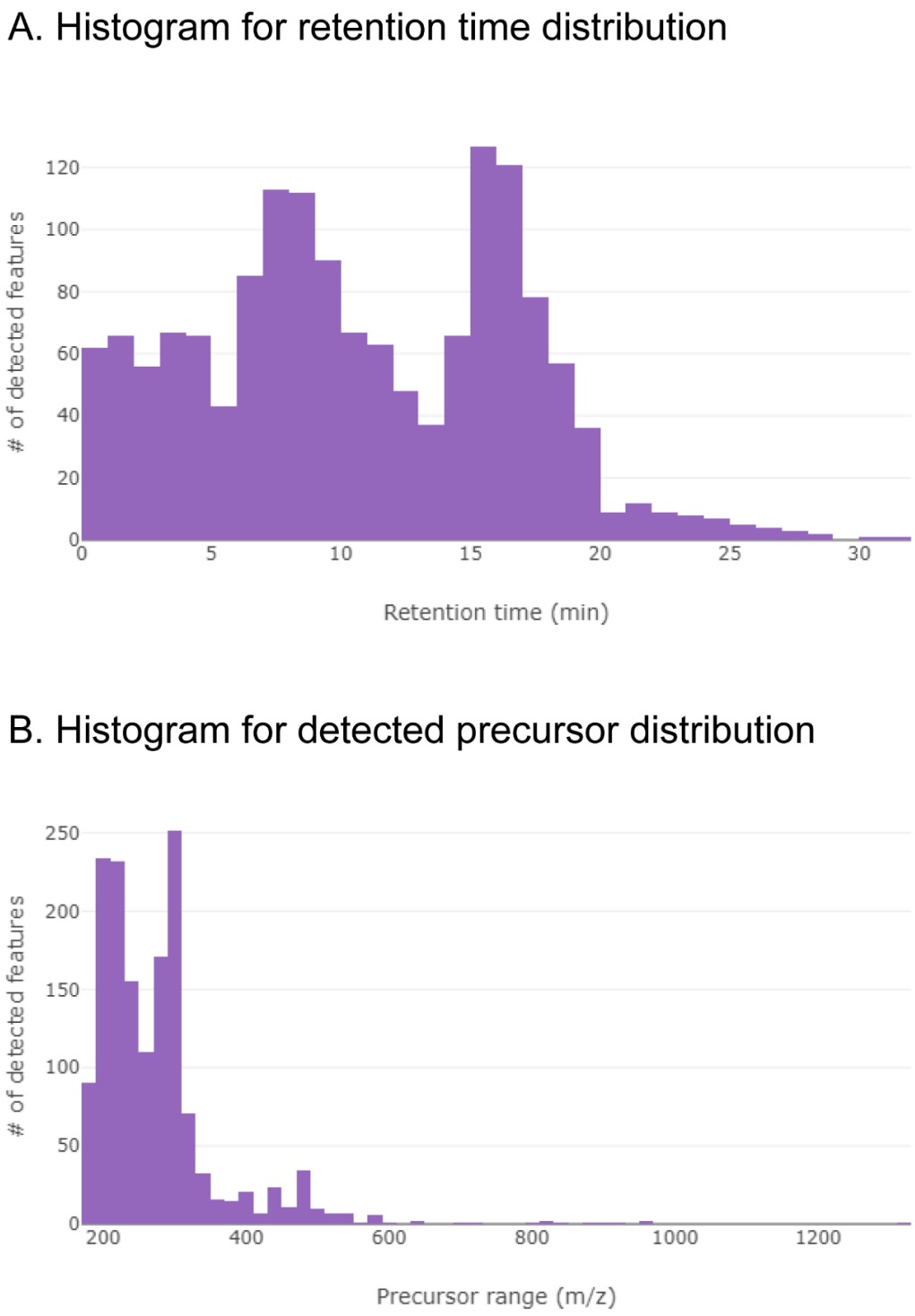


Figure S1. Additional statistical graphs from Metandem: A. Histogram for retention time distribution; **B.** Histogram for detected precursor distribution.

Table S1. List of 59 identified metabolites of interest.

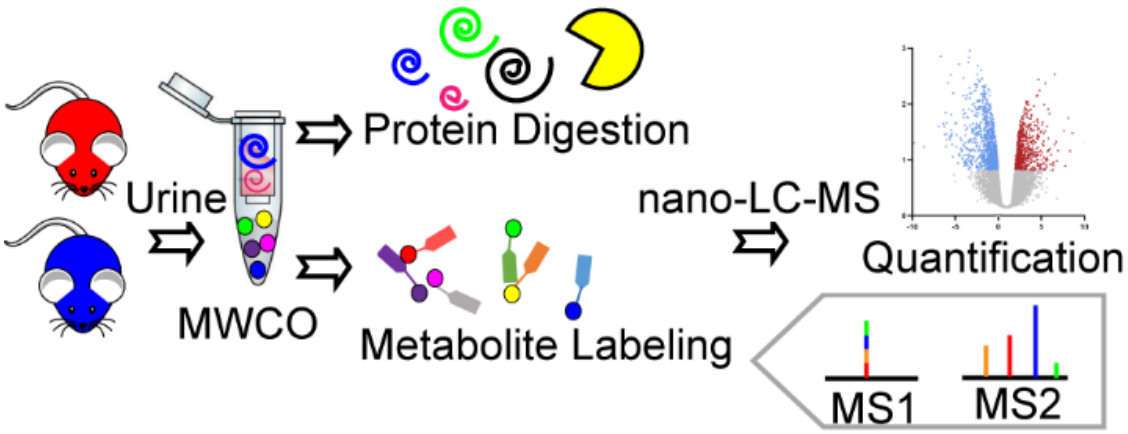
Compound	Molecular Weight (Da)	Mass of labeled metabolite	Δppm	Time (min)	HMDB_ID
2-Aminobenzoic acid	137.0464	283.1822	9.7	19.0	HMDB01123
Leucine	131.0942	277.2301	3.0	19.0	HMDB00687
Purine[a]	120.0450	266.1808	11.3	8.6	HMDB01366
Citrulline	175.0936	321.2295	11.7	15.5	HMDB00904
Proline	115.0623	261.1981	9.2	13.6	HMDB00162
5-Aminopentanamide[a]	116.0938	262.2296	10.2	14.4	HMDB12176
N-Acetylputrescine	130.1108	276.2466	1.2	15.2	HMDB02064
D-Alanyl-D-alanine[a]	160.0827	306.2185	13.1	16.5	HMDB03459
O-Phosphohomoserine	199.0278	345.1636	16.1	16.7	HMDB03484
Glutathione	307.0788	453.2147	16.2	18.1	HMDB00125
GABA	103.0618	249.1977	14.5	15.9	HMDB00112
Creatine	131.0698	277.2057	2.6	19.7	HMDB00064
Alanine	89.0467	235.1826	10.8	7.9	HMDB00161
Taurine	125.0142	271.1501	3.4	8.8	HMDB00251
Formamide[a]	45.0216	191.1574	2.1	8.7	HMDB01536
Glycine	75.0307	221.1665	17.8	9.4	HMDB00123
Cysteinylglycine	178.0421	324.1779	4.8	10.5	HMDB00078
Glutamic acid	147.0550	293.1908	12.6	16.0	HMDB00148
Cysteine	121.0175	267.1533	19.0	15.3	HMDB00574
4-Aminobutyraldehyde[a]	87.0675	233.2034	10.1	13.7	HMDB01080
1,2,3,4-Tetrahydroisoquinoline[a]	133.0889	279.2248	1.5	14.9	HMDB12489
3-Amino-2-piperidone	114.0784	260.2143	7.9	14.9	HMDB00323
6-Dimethylaminopurine[a]	163.0876	309.2234	11.0	15.4	HMDB00473
3-Methylindole[a]	131.0752	277.2111	13.2	15.4	HMDB00466
Dihyroxy-1H-indole glucuronide I	325.0784	471.2143	4.1	15.5	HMDB59997
Aminoadipic acid	161.0676	307.2034	7.8	15.6	HMDB00510
5-L-Glutamyl-taurine[a]	254.0583	400.1942	4.3	15.7	HMDB04195
Homocysteic acid[a]	183.0211	329.1569	5.1	15.7	HMDB02205
Diaminopimelic acid	190.0932	336.2290	11.4	16.0	HMDB01370
2,4-Diaminobutyric acid	118.0727	264.2085	13.1	15.9	HMDB02362
2,3-Diaminopropionic acid	104.0600	250.1958	13.7	15.9	HMDB02006
N-Carboxyethyl-g-aminobutyric acid[a]	175.0822	321.2180	13.2	16.1	HMDB02201
N-Acetylcadaverine	144.1251	290.2610	7.9	16.5	HMDB02284
4-Acetylimidazo[4,5-c]pyridine[a]	161.0597	307.1956	5.2	16.6	HMDB34888
S-(2-carboxypropyl)-Cysteamine	163.0637	309.1995	18.5	16.6	HMDB02169
7-Aminoclonazepam[a]	285.0642	431.2000	9.4	16.8	HMDB41817

Compound	Molecular Weight (Da)	Mass of labeled metabolite	Δppm	Time (min)	HMDB_ID
Glutamylalanine[a]	217.0867	363.2225	19.6	16.7	HMDB03764
2-Aminomuconic acid[a]	157.0361	303.1719	9.1	17.0	HMDB01241
2-Indolecarboxylic acid	161.0490	307.1848	8.2	17.0	HMDB02285
3-Indole carboxylic acid glucuronide	337.0792	483.2150	1.7	17.4	HMDB13189
S-Nitrosoglutathione[a]	336.0755	482.2114	4.7	17.4	HMDB04645
Alanyltryptophan	275.1218	421.2577	18.7	17.8	HMDB13209
N2,N2-Dimethylguanosine	311.1264	457.2622	11.0	17.9	HMDB04824
5-Hydroxy-6-methoxyindole glucuronide	339.0946	485.2305	2.3	17.9	HMDB10363
S-(Hydroxymethyl)glutathione[a]	337.0926	483.2284	5.3	17.9	HMDB04662
Glutamyl-Tyrosine[a]	309.1034	455.2392	17.0	17.9	HMDB28831
Inodxyl glucuronide	309.0813	455.2172	11.3	18.0	HMDB10319
Tyramine-O-sulfate	217.0398	363.1756	5.1	18.1	HMDB06409
Glutaminyl-Arginine[a]	302.1691	448.3050	3.7	18.7	HMDB28791
L-alpha-glutamyl-L-hydroxyproline	260.0977	406.2336	11.9	19.3	HMDB11161
S-3-oxodecanoyl cysteamine	245.1482	391.2840	13.1	20.9	HMDB59773
4-Thialysine[a]	164.0616	310.1974	2.3	16.6	HMDB29178
3-Hydroxyanthranilic acid	153.0428	299.1786	1.3	16.8	HMDB01476
5-Hydroxyindoleacetic acid	191.0579	337.1937	1.9	16.8	HMDB00763
2'-Deoxysepiapterin	221.0929	367.2288	7.5	17.2	HMDB00389
Cystathionine	222.0673	368.2031	0.6	17.2	HMDB00099
S-Formylglutathione[a]	335.0761	481.2120	7.8	17.4	HMDB01550
Homolanthionine[a]	268.0679	414.2037	18.7	17.8	HMDB02034
Tyramine glucuronide	313.1120	459.2479	13.1	19.2	HMDB10328

[a] First identified in urine samples

Chapter 3

Urinary Metabolomic and Proteomic Analyses in a Mouse Model of Prostatic Inflammation



Adapted from **Wei P**, Hao L, Ma F, Yu Q, Buchberger AR, Lee S, Bushman W, Li LL. "Urinary Metabolomic and Proteomic Analyses in a Mouse Model of Prostatic Inflammation. " *Journal of Urine, accepted.* **Wei P**, designed and conducted the study under the supervision of LL, Hao L provided expertise in metabolomics, Ma F, provided expertise in proteomics, Yu Q provided expertise instrumentation, Buchberger AR systhesized the DiLeu tags, Lee S provided the mouse urine sample, Bushman W provided insights in LUTS.

ABSTRACT

Lower urinary tract symptoms (LUTS) are common among aging men. Since prostatic inflammation is one of its etiologies, it is plausible that urinary metabolite and protein biomarkers could be identified and used to diagnose inflammation-induced LUTS. We characterized the urine metabolome and proteome in a mouse model of bacterial-induced prostatic inflammation. Mass Spectrometry (MS)-based multi-omics analysis was employed to discover urinary protein and metabolite-based biomarkers. The investigation of isobaric dimethylated leucine (DiLeu) labeling on metabolites allowed metabolomics and proteomics analysis on the same liquid chromatography (LC)-MS platform. In total, 143 amine-containing metabolites and 1058 urinary proteins were identified and quantified (data are available via ProteomeXchange with identifier PXD018023); among them, 14 metabolites and 168 proteins were significantly changed by prostatic inflammation. Five metabolic pathways and four inflammation-related biological processes were potentially disrupted. By comparing our findings with urinary biomarkers identified in a mouse model of genetic-induced prostate inflammation and with those previously found to be associated with LUTS in older men, we identified creatine, haptoglobin, immunoglobulin kappa constant and polymeric Ig receptor as conserved biomarkers for prostatic inflammation associated with LUTS. These data suggest that these putative biomarkers could be used to identify men in which prostate inflammation is present and contributing to LUTS.

KEYWORDS: urine, metabolomics, proteomics, lower urinary tract symptoms, prostatic inflammation, mass spectrometry.

ABBREVIATIONS:

LUST, Lower urinary tract symptoms; DiLeu, *N, N*-dimethyl leucine; NCE, normalized collision energy; DTT, dithiothreitol; IAA, iodoacetamide; TFA, trifluoroacetic acid; KEGG, Kyoto Encyclopedia of Genes and Genomes; HCD, higher-energy C-trap dissociation; GO, gene ontology; HMDB, Human Metabolome Database; IL-1, interleukin-1.

1. INTRODUCTION

Lower urinary tract symptoms (LUTS) are highly prevalent, affecting nearly 50% to 90% of men aged 50 years or older (1). Urinary symptoms include irritative symptoms such as urinary frequency, urgency and nocturia; and obstructive symptoms such as a weak urinary stream, straining to void, and a sense of incomplete emptying (2–4). The etiology of LUTS in men is considered to be multifactorial. Prostatic enlargement and age-related changes of the bladder detrusor are considered major factors, but many patients have symptoms without these conditions. Recently, prostate inflammation has been implicated as a possible cause of irritative voiding symptoms. We and others have shown that prostate inflammation in a mouse model is associated with detrusor overactivity and increased voiding frequency, an effect attributable to neural cross-talk between the prostate and bladder at the level of the dorsal root ganglia (5–8). Although patients are suffering a variety of etiologic mechanisms, the current subject symptom score diagnosis method usually could not distinguish the pathogenic factors from different individuals. As a result, patients are treated empirically with either medication or surgery. Objective biomarkers associated with specific etiologic mechanisms are greatly needed to refine the diagnostic approach

and provide personalized treatment (9–11). The purpose of this study is to characterize the contribution of prostatic inflammation to the urine metabolome and proteome under controlled, experimental conditions in a mouse model of bacterial-induced prostatic inflammation. The long-term goal of this work is to validate LUTS biomarkers for patient stratification and personalized treatments.

Acute and chronic inflammation is common in the prostate of adult men. The etiology is thought to be multifactorial and may include bacterial colonization, viral infection, changes in serum testosterone, and reflux of noxious chemicals in the urine (12). In this model, prostatic inflammation is induced by a single transurethral inoculation with uropathogenic *E.coli* 1677 in adult male C57BL/6J mice. Nitrofurantoin is administered beginning one day prior to bacterial inoculation to help eliminate inflammation of the bladder or urethra (8, 13). To characterize the urine metabolome and proteome of the bacterial-induced prostatic inflammation model, we collected urine samples from both the control group and the bacterial-induced prostatic inflammation group on day seven post-instillation.

Twelve-plex *N, N*-dimethyl leucine (DiLeu) isobaric labeling has been proven to be an accurate and high throughput quantification method for metabolomics studies in biological systems (14). It could increase the retention of polar metabolites and enable them to be separated and detected on nanoUPLC systems, meaning improved chromatographic resolution and detection sensitivity over typical standard flow separation techniques (15). Additionally, Metandem, a novel online software tool developed in our lab and freely available at http://metandem.com/web/, was created for isobaric labeling-based

metabolomics quantification, identification, and statistical analysis (16). Here, we paired the DiLeu labeling method and Metandem software tool to analyze the dataset of DiLeu labeled urinary metabolites in controlled, experimental samples from mice with bacterial-induced prostatic inflammation.

In this study, metabolomics and proteomics analyses have been integrated in analysis of urine from a mouse model of bacterial-induced inflammation. Urine normally contains low protein concentration (17) (18), but our previous study revealed that proteins involved in inflammation and prostate fibrosis were upregulated in men with LUTS (19). This combined MS analysis casts a wide net to identify both proteins and metabolites entering the urine from the bladder, urethra, and/or prostate in mice with prostatic inflammation (20). It is expected that this strategy may identify clinically useful biomarkers of prostatic inflammation. In addition, an integrated analysis of both metabolic and proteomics urinary biomarkers may provide insight into the mechanism for LUTS associated with prostatic inflammation and identify new targets for drug treatment.

2. MATERIALS AND METHODS

2.1 Mouse model of bacterial-induced prostatic inflammation

All animal procedures were approved by the University of Wisconsin-Madison Animal Care and Use Committee. 12-week old male C57BL/6J mice were used for this study (Jackson Lab, Bar Harbor, MA). *E. coli* 1677 ($1x10^8$ bacteria /ml, 20 µl) was instilled into the bladders of anesthetized mice (n=4) via a transurethral catheter, as previously described (13, 21, 22). Control mice (n=4) were inoculated with an equal volume of PBS. 27.2µg/g

nitrofurantoin (Sigma-Aldrich, St. Louis, MO) was injected into both groups via subcutaneous twice daily. The administration of nitrofurantoin (3.4 mg/kg, sc, twice daily) was begun one day prior to transurethral instillation and continued until the end of the experiment. Seven days after instillation, urine was collected from each animal housed in a metabolic cage (23) for a two-hour period. All samples were stored at -80 °C until further processing.

2.2 Mice urine sample preparation

Urine samples (~110 µL) were thawed on ice and mixed well before centrifugation (10,000 *xg,* 10 min, 4 °C). Clarified supernatant (~100 µL) was aliquoted from each sample and loaded onto pre-rinsed (14) centrifugal filters (3 kDa, Millipore Amicon Ultra, Burlington, MA). The fluid that went through centrifugal filters during centrifugation (14,000 *xg,* 30 min, 4 °C) was collected as the metabolite fraction (<3 kDa). The centrifugal filters were then rinsed twice with Milli-Q water (200 µL; same centrifugation speed and time), and the flow-through was combined with the previously collected metabolite fraction (total volume ~ 500 µL). Next, 200 µL Milli-Q water was used to dissolve the contents left on the membrane of centrifugal filters. The filters were then flipped over and centrifuged at 10,000 *xg* for 2 min to collect the contents as the protein fraction (>3 kDa).

2.3 Metabolite fraction sample preparation

Osmolality was used to normalize all the metabolite fraction to 50 mOsmoles/kg H_2O (24). All the samples were then aliquoted, lyophilized, and stored at -80 °C until labeling. Eight

channels were randomly chosen from the twelve-plex isobaric labeling reagents to match the eight samples. Four channels (117a, 117 c, 118a, and 118b) were used to label samples from bacterial-induced prostatic inflammation mouse, and four channels (115b, 116a, 116b, 116c) to label control samples. Isobaric labeling was performed as previously described (25). Then equal amounts of eight labeled samples were combined, and excess DiLeu reagents were removed from the pooled sample via SCX Ziptips (OMIX-SCX, Agilent, Santa Clara, CA) (15). After lyophilization, the pooled metabolite sample was stored at -80 °C until analysis.

2.4 Protein fraction sample preparation

The protein fraction was lyophilized and resuspended in 8 M urea lysis buffer. The total protein concentration of each urine sample was determined by the BCA assay (Pierce™ BCA Protein Assay Kit, Thermo, Waltham, MA) (26). An equal amount of protein (200 μg) was aliquoted from each sample. The disulfide bonds were reduced with 5 mM dithiothreitol (DTT) for 1 h at room temperature followed by alkylation with 15 mM iodoacetamide (IAA) in the dark for 30 min. The alkylation was quenched by 5 mM DTT. Next, the protein sample was diluted with 50 mM Tris solution (pH = 8) to a concentration of urea less than 1 M. Trypsin was used for protein digestion (protein/enzyme ratio 50:1) at 37 °C for 17 h. The digestion was quenched by 10% trifluoroacetic acid (TFA) to a pH lower than 2. Then, the digested proteins were desalted with Sep-Pak C18 (Sep Pak C18 Cartridges, Waters, Milford, MA), lyophilized, and stored at -80 °C until LC-MS/MS analysis.

2.5 Metabolomics LC-MS Acquisition and Data Analysis

For the metabolite fraction, the pooled sample was reconstituted in 3% acetonitrile, 0.1% formic acid (v/v) in water and inject into Thermo Dionex UltiMate^{TM} 3000 nanoLC system coupled to a Thermo Q Exactive^{TM} HF Orbitrap MS. The LC-MS acquisition method was as previously described (14). For the metabolomics data analysis, Metandem (16) was used to batch-process three technical replicates of pooled sample for metabolite identification and quantification. Two-tailed Student's t-test (unequal variance) was used to compare metabolite abundance between groups (*i.e., E. coli* infected vs control mice). Metabolites were considered significantly modulated when fold changes were >1.5 and p-values were <0.05 (t-test). MetaboAnalyte 4.0 software (26) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (27) were used for metabolic pathway analysis. Pathway was considered dysregulated when p-value was <0.05 (Fisher's exact test).

2.6 Proteomics LC-MS Acquisition and Data Analysis

The same LC-MS platform was applied to protein fraction with minor modification. An optimized 142 min gradient used was as follows: 0-16 min, 3% solvent B; 16-107 min, 3-30% B; 107-107.5 min, 30-75% B; 107.5-117 min, 75% B; 117-117.5 min, 75-95% B; 117.5-127 min 95% B; 127-127.5 min, 95-3% B; 127.5-142 min, 3% B. Full MS scans were acquired from *m/z* 300 to 1500 at a resolution of 60 K, AGC at 1×10^6, and maximum injection time (IT) of 100 ms. The top 15 precursors were then selected for higher-energy C-trap dissociation tandem mass spectrometry (HCD MS2) analysis with an isolation

window of 1.4 *m/z*, a normalized collision energy (NCE) of 30, a resolving power of 15 K, an AGC target of 1×10^5, a maximum injection time of 100 ms, and a lower mass limit of 120 *m/z*.

For the proteomics data analysis, the raw data files from the Orbitrap MS analysis were searched against a mouse proteome database downloaded from Uniprot (April 2017, entry number 24977) (28) utilizing PEAKS 8.5 software (29). A precursor mass error tolerance of 10 ppm and a fragment mass error tolerance of 0.01 Da were allowed. Trypsin was chosen as the enzyme, and max missed cleavages were two. Carbamidomethylation was set as fixed modification, and oxidation (M) was set as variable modifications and three maximum PTMs were allowed per peptide. Parameters for confident peptide identification were Ascore (PTM site confidence) higher than 20, FDR lower than 1%, and the presence of at least one unique peptide. The threshold for the quantification of significant proteins was the same as metabolites ($p<0.05$) using a t-test. DAVID (26) was used for gene ontology (GO) annotation of identified proteins. Candidate biomarkers, dysregulated pathways, and biological processes were compared with prior human patient analyses and other previous reports. Both the mass spectrometry metabolomics and proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (30) partner repository with the dataset identifier PXD018023.

3. RESULTS AND DISCUSSION

3.1 Mouse urine metabolite identification and quantification

DiLeu isobaric labeling allowed both multiplexed metabolite identification and quantification in a single injection (**Table S1,** sample labeling map), which reduces demand for instrument time and improves relative quantification accuracy by decreasing run-to-run variation. Metandem, a newly developed custom software platform, was successfully used to process this large dataset for global stable isotope labeling-based metabolite identification and quantification. Results of automated parameter optimization (**Figure S1)** were as follows: reporter ion mass tolerance, 0.4 mDa; batch processing mass tolerance, 10 ppm; and batch processing retention time tolerance, 0.5 min.

Three technical replicates of each pooled sample injection were merged via Metandem. Only metabolites which displayed reporter ions in every channel were used for further analysis. A total of 143 primary or secondary amine-containing metabolites (**Table S2**) were identified after matching for accurate mass against the Human Metabolome Database (HMDB) (31). Of these, 102 have been previously documented in the urine metabolome database (32); the remaining 41 amine-containing metabolites are, to our knowledge, identified in urine for the first time. Of the metabolites with omnipresent reporter ions, 14 were identified as statistically significant biomarker candidates (Student's t-test, fold change>1.5, p-value<0.05) (**Figure 1**). Among them, nine were upregulated, while five were downregulated.

3.2 Metabolic pathway analysis

Metabolic pathway analysis is an effective way to understand the cellular metabolic network and metabolic objectives (33). All identified metabolites were entered into the MetaboAnalyte 4.0 software, and metabolic pathways with at least one significantly changed metabolite (p-value<0.05; Fisher's exact test) were considered to be perturbed. In total, five perturbed metabolic pathways were identified (**Table 1)**: tryptophan metabolism pathway; arginine and proline metabolism pathway; histidine metabolism pathway; glycine, serine and threonine metabolism pathway, and alanine, aspartate and glutamate metabolism pathway.

3.3 Mouse urine protein identification and quantification

Both labeling and label-free methods of protein identification were performed and compared. While the labeling method is more high-throughput, we found that the label-free method identified more peptides and proteins in the urine samples. Proteomics analysis of the mouse urine samples with the label-free method enabled us to identify 6309 unique peptide sequences belonging to 1058 unique proteins (**Table S3**). Comparison of protein abundance in the prostatic inflammation and the control groups identified 168 proteins statistically significant candidate biomarkers (Student's t-test, fold change>1.5) (**Figure 2)**. Among these, 164 were upregulated, while 4 were downregulated. The 30 proteins showing the greatest magnitude of change are listed in **Table 2**.

3.4 Proteomics Gene Ontology analysis

Although urinary proteomic analysis does not directly implicate any metabolic or signaling pathway, it is nonetheless instructive to perform GO classification of the significantly changed proteins to obtain insight into their possible roles in prostate inflammation and LUTS (34). Cell component analysis and biology process analysis of the 168 significantly change are shown in **Figure 3**. Cell component analysis identified most proteins as belonging to the extracellular component. Biological process analysis identified proteins involved in acute inflammation, oxidative stress, fibroblast proliferation and wound healing.

3.5 Correlation between metabolic and proteomic analyses identifies biomarker profiles of inflammation and fibrosis

Metabolomic analysis revealed perturbation of the tryptophan metabolism pathway (**Table 1**). Tryptophan metabolism and, most especially, its derivative serotonin has been implicated in age-related chronic inflammation (35–39). Proteomic analysis revealed, upregulation of a group of proteins involved in the inflammatory response to injury, including acute-phase response (four proteins), wound healing (six proteins) and oxidative stress response (six proteins; **Figure 3**). Metabolomic analysis also revealed perturbation of arginine and proline metabolism pathways (**Table 1**). These pathways are known to be related to the synthesis of collagen (40) and it is therefore notable that the group of differentially expressed proteins includes several collagens (collagen alpha-1(I) chain, collagen alpha-1(XV) chain, collagen alpha-1(XXIV) chain) **(Table S3).** In addition, the proteomics GO analysis identified nine of the differentially expressed proteins are positive

regulators of fibroblast proliferation. Taken together, these data comprise what appears to be a profile of biomarkers reflecting a fibrotic response to prostatic inflammation.

It has been speculated that collagen accumulation (fibrosis) of the prostate, possibly occurring as a consequence of inflammation, can contribute to LUTS by partially causing bladder outlet obstruction (41–43). The inflammatory response to a prostate infection in our mouse model of bacterial-induced prostatic inflammation has been well-characterized and has been shown to induced collagen synthesis and accumulation. Accordingly, these potential biomarkers are likely relevant to LUTS occurring in association with prostatic inflammation.

In order to identify metabolic pathway changes and differential protein expression that are conserved features of prostate inflammation, we compared our findings here with previous studies in a mouse model of prostate inflammation-induced by transgenic overexpression of interleukin-1 (IL-1) in the prostate epithelium (44, 45). This comparison analysis revealed that creatine, a product of the arginine and proline metabolism pathway, was up-regulated in both inflammation mouse models (45). The significant changes in creatine were 1.5- and 1.6-fold increases in the transgenic mouse model and our bacterial-induced mouse model, respectively. We also identified three proteins whose expression was significantly increased in both models: haptoglobin, immunoglobulin kappa constant, and polymeric Ig receptor.

In order to determine if the changes we observed here are specific for prostatic inflammation or represent a generalized response to perturbations of the lower urinary tract, we compared them to changes observed in a mouse model of hormonally induced prostate hyperplasia and voiding dysfunction (16) (46). Metabolomic analysis of the urine revealed that this model also showed perturbation of the arginine and proline metabolic pathway, but different metabolites were found to be disrupted in these two different models. Creatine, putrescine, and argininosuccinic acid were significantly changed in urine samples from prostatic inflammation mouse model as compared to its control samples, whereas, citrulline, N-acetylputrescine, and proline were significantly changed in urine samples from hormone-induced mouse model as compared to its control samples. These observations support the conclusion that the different profiles of urinary biomarkers identified by metabolomic and proteomic analysis in the different models likely reflect the unique etiologic mechanisms.

3.6 Comparison with findings in men with LUTS

We previously performed metabolomic and proteomic analysis of urine from men with LUTS (19, 24). To determine whether our findings in the mouse model are recapitulated in patients with LUTS, we compared the changes observed in our mouse model of bacterial-induced prostatic inflammation with those found in men with LUTS (**Figure 4**). Metabolomic analysis revealed perturbation of the arginine and proline metabolic pathway in the urine metabolome, similar to that observed in the mouse models described above, but the specific metabolites exhibiting changes appeared to be different (**Figure 4**). Proteomic analysis revealed significant changes in processes related to fibrosis and

inflammation (**Figure 4**), similar to the mouse models of prostate inflammation but, again, the specific proteins that exhibited significantly changed expression were not the same. This is not unexpected given that men with LUTS present a complex mix of patients suffering a variety of etiologic mechanisms and exhibiting a wide spectrum of symptoms. Changes within a small component of that population that mimic those found in a specific mouse model are likely to be obscured. In order to utilize the biomarkers identified here to stratify patients with LUTS, it will first be necessary to stratify patients by mechanism or by clinical features. For example, urine could be obtained from patients prior to prostate surgery, and the proteomic and metabolomic analysis performed on those patients in whom substantial inflammation is present in the removed prostate tissue. Alternatively, an analysis could be performed on patients who exhibit much greater irritative than obstructive symptoms. Once a correspondence is established with a subsegment of the LUTS patients, then the biomarkers might be used as a tool for stratification and individualized treatment.

4. CONCLUSIONS

Our aim was to perform urine metabolomics and proteomics analyses to determine LUTS biomarkers representing multi-factorial etiologies of LUTS and specifically targeting prostatic inflammation-induced LUTS. This aim was pursued by discovering novel prognostic metabolites and protein biomarkers in urine and linking them to the pathogenesis of LUTS through disrupted metabolic pathways and biological processes. The benefit of our study showing urine-based metabolomics and proteomics designed for LUTS is expected to be four-fold. First of all, the investigation of DiLeu labeling on metabolites allowed metabolomics and proteomics analysis on the same LC-MS platform. Second, the

combined metabolomics and proteomics analyses provided an informative method for the study of disease molecular mechanisms, which contributed to biomarker discovery and patient stratification according to etiologies. Third, metabolite creatine and proteins haptoglobin, immunoglobulin kappa constant, and polymeric Ig receptor were found to be candidate biomarkers for prostatic inflammation induced-LUTS and could be further used as the LUTS patient stratification. Fourth, urine samples from voluntary voiding, meaning a non-invasive method for sample collection, have been used for metabolomics and proteomics. Future studies will focus on clinical validation of the metabolite and protein biomarkers found in this study with a separate large cohort of patient samples.

ACKNOWLEDGMENTS

The authors would like to thank the DiLeu synthesis team in the Li Lab, including Dr. Tyler Greer, Dr. Dustin Frost, and Dr. Amanda Buchberger. The authors would also like to thank the Zeeh Pharmaceutical Experiment Station in the School of Pharmacy for instrument access. This work was supported in part by the National Institutes of Health through Grants 1P20 DK097826, U54 DK104310, R01DK071801, RF1AG052324, P41GM108538, R01 DK093690, R01 DK099328, DK091193, and DK104310. The Q-Exactive Orbitrap instrument was purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531). L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

REFERENCES

1. Parsons JK, Bergstrom J, Silberstein J, Barrett-Connor E. 2008. Prevalence and Characteristics of Lower Urinary Tract Symptoms in Men Aged ≥80 Years. Urology 72:318–321.

2. Speakman M, Kirby R, Doyle S, Ioannou C. 2015. Burden of male lower urinary tract symptoms (LUTS) suggestive of benign prostatic hyperplasia (BPH) - focus on the UK. BJU Int 115:508–519.

3. Abrams P, Cardozo L, Fall M, Griffiths D, Rosier P, Ulmsten U, Van Kerrebroeck P, Victor A, Wein A. 2003. The standardisation of terminology in lower urinary tract function: Report from the standardisation sub-committee of the International Continence Society. Urology. Elsevier Inc.

4. Chapple CR, Roehrborn CG. 2006. A shifted paradigm for the further understanding, evaluation, and treatment of lower urinary tract symptoms in men: Focus on the bladder. Eur Urol. Elsevier.

5. Park JS, Jin MH, Hong CH. 2018. Neurologic mechanisms underlying voiding dysfunction due to prostatitis in a rat model of nonbacterial prostatic inflammation. Int Neurourol J 22:90–98.

6. Ni J, Mizoguchi S, Bernardi K, Suzuki T, Kurobe M, Takaoka E, Wang Z, DeFranco DB, Tyagi P, Gu B, Yoshimura N. 2019. Long-lasting bladder overactivity and bladder afferent hyperexcitability in rats with chemically-induced prostatic inflammation. Prostate 79:872–879.

7. Funahashi Y, Takahashi R, Mizoguchi S, Suzuki T, Takaoka E, Ni J, Wang Z, DeFranco DB, Groat WC, Tyagi P, Yoshimura N. 2019. Bladder overactivity and

afferent hyperexcitability induced by prostate-to-bladder cross-sensitization in rats with prostatic inflammation. J Physiol 597:2063–2078.

8. Lee S, Yang G, Bushman W. 2015. Prostatic inflammation induces urinary frequency in adult mice. PLoS One 10:1–12.

9. Shi T, Quek SI, Gao Y, Nicora CD, Nie S, Fillmore TL, Liu T, Rodland KD, Smith RD, Leach RJ, Thompson IM, Vitello EA, Ellis WJ, Liu AY, Qian WJ. 2017. Multiplexed targeted mass spectrometry assays for prostate cancer-associated urinary proteins. Oncotarget 8:101887–101898.

10. Luan H, Liu LF, Tang Z, Zhang M, Chua KK, Song JX, Mok VCT, Li M, Cai Z. 2015. Comprehensive urinary metabolomic profiling and identification of potential noninvasive marker for idiopathic Parkinson s disease. Sci Rep 5:1–11.

11. Hawkridge AM, Muddiman DC. 2009. Mass Spectrometry–Based Biomarker Discovery: Toward a Global Proteome Index of Individuality. Annu Rev Anal Chem 2:265–277.

12. Bushman WA, Jerde TJ. 2016. The role of prostate inflammation and fibrosis in lower urinary tract symptoms. Am J Physiol Physiol 311:F817–F821.

13. Johny E. Elkahwaji WZ, Walter J. Hopkins WB. 2007. Chronic Bacterial Infection and Inflammation Incite ReactiveHyperplasia in aMouseModelofChronic Prostatitis. Prostate 104:14–21.

14. Wei P, Hao L, Thomas S, Buchberger AR, Hyman S, Strinkr L, Ricke WA, Li LL. Urinary Amine Metabolomics Characterization of a Hormone-Induced Urinary Obstruction Mouse Model with Custom 12-plex Isobaric DiLeu Labeling.

15. Hao L, Zhong X, Greer T, Ye H, Li L. 2015. Relative quantification of amine-

containing metabolites using isobaric N,N-dimethyl leucine (DiLeu) reagents via LC-ESI-MS/MS and CE-ESI-MS/MS. Analyst 140:467–475.

16. Hao L, Zhu Y, Wei P, Johnson J, Buchberger A, Frost D, Kao WJ, Li L. 2019. Metandem: An online software tool for mass spectrometry-based isobaric labeling metabolomics. Anal Chim Acta 1088:99–106.

17. ZHAO M, PAN X, GAO Y, LIU Y. 2020. Why are there proteins in the urine of healthy people? Sci Sin Vitae 50:338–348.

18. Thomas S, Hao L, Ricke WA, Li L. 2016. Biomarker discovery in mass spectrometry-based urinary proteomics. Proteomics - Clin Appl. Wiley-VCH Verlag.

19. Greer T, Hao L, Nechyporenko A, Lee S, Vezina CM, Ricke WA, Marker PC, Bjorling DE, Bushman W, Li L. 2015. Custom 4-Plex DiLeu Isobaric Labels Enable Relative Quantification of Urinary Proteins in Men with Lower Urinary Tract Symptoms (LUTS). PLoS One 10:e0135415.

20. Wettersten HI, Hakimi AA, Morin D, Bianchi C, Johnstone ME, Donohoe DR, Trott JF, Abu Aboud O, Stirdivant S, Neri B, Wolfert R, Stewart B, Perego R, Hsieh JJ, Weiss RH. 2015. Grade-dependent metabolic reprogramming in kidney cancer revealed by combined proteomics and metabolomics analysis. Cancer Res 75:2541–2552.

21. Boehm BJ, Colopy SA, Jerde TJ, Loftus CJ, Bushman W. 2012. Acute bacterial inflammation of the mouse prostate. Prostate 72:307–317.

22. Jerde TJ, Bushman W. 2009. IL-1 induces IGF-dependent epithelial proliferation in prostate development and reactive hyperplasia. Sci Signal 2.

23. Cornu JN, Ahyai S, Bachmann A, De La Rosette J, Gilling P, Gratzke C, McVary

K, Novara G, Woo H, Madersbacher S. 2015. A systematic review and meta-analysis of functional outcomes and complications following transurethral procedures for lower urinary tract symptoms resulting from benign prostatic obstruction: An update. Eur Urol. Elsevier.

24. Hao L, Greer T, Page D, Shi Y, Vezina CM, Macoska JA, Marker PC, Bjorling DE, Bushman W, Ricke WA, Li L. 2016. In-Depth Characterization and Validation of Human Urine Metabolomes Reveal Novel Metabolic Signatures of Lower Urinary Tract Symptoms. Sci Rep 6:30869.

25. Frost DC, Greer T, Li L. 2015. High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. Anal Chem 87:1646–1654.

26. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC. 1985. Measurement of protein using bicinchoninic acid. Anal Biochem 150:76–85.

27. Kanehisa M. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30.

28. Bateman A. 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515.

29. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B. 2012. PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics 11.

30. Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dianes JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolomé S, Apweiler

R, Omenn GS, Martens L, Jones AR, Hermjakob H. 2014. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol. Nature Publishing Group.

31. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, MacInnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L. 2007. HMDB: The human metabolome database. Nucleic Acids Res 35.

32. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, Dame ZT, Poelzer J, Huynh J, Yallou FS, Psychogios N, Dong E, Bogumil R, Roehring C, Wishart DS. 2013. The Human Urine Metabolome. PLoS One 8:e73076.

33. Schilling CH, Schuster S, Palsson BO, Heinrich R. 1999. Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-genomic Era. Biotechnol Prog 15:296–303.

34. Yu LR, Stewart NA, Veenstra TD. 2010. Proteomics. The Deciphering of the Functional Genome., p. 89–96. *In* Essentials of Genomic and Personalized Medicine. Elsevier Inc.

35. Rose DP. 1972. Aspects of tryptophan metabolism in health and disease: a reviewJ. clin. Path.

36. Xie J, Zhang A hua, Qiu S, Zhang T lei, Li X na, Yan G li, Sun H, Liu L, Wang X

jun. 2019. Identification of the perturbed metabolic pathways associating with prostate cancer cells and anticancer affects of obacunone. J Proteomics 206.

37. Wolf H, Madsen P 0, Price JM. 1968. Studies on the Metabolism of Tryptophan in Patients with Benign Prostatic Hypertrophy or Cancer of the ProstateTHE JOURNAL OF UROLOGY.

38. Sorgdrager FJH, Naudé PJW, Kema IP, Nollen EA, De Deyn PP. 2019. Tryptophan metabolism in inflammaging: From biomarker to therapeutic target. Front Immunol. Frontiers Media S.A.

39. He Q, Wang Z, Liu G, Daneshgari F, MacLennan GT, Gupta S. 2016. Metabolic syndrome, inflammation and lower urinary tract symptoms: Possible translational links. Prostate Cancer Prostatic Dis. Nature Publishing Group.

40. Barbul A. 2008. Proline Precursors to Sustain Mammalian Collagen Synthesis. J Nutr 138:2021S-2024S.

41. Boehm BJ, Colopy SA, Jerde TJ, Loftus CJ, Bushman W. 2012. Acute bacterial inflammation of the mouse prostate. Prostate 72:307–317.

42. Wong L, Hutson PR, Bushman W. 2014. Prostatic inflammation induces fibrosis in a mouse model of chronic bacterial infection. PLoS One 9.

43. Ma J, Gharaee-Kermani M, Kunju L, Hollingsworth JM, Adler J, Arruda EM, MacOska JA. 2012. Prostatic fibrosis is associated with lower urinary tract symptoms. J Urol 188:1375–1381.

44. Hao L, Thomas S, Greer T, Vezina CM, Bajpai S, Ashok A, De Marzo AM, Bieberich CJ, Li L, Ricke WA. 2019. Quantitative proteomic analysis of a genetically induced prostate inflammation mouse model via custom 4-plex DiLeu

isobaric labeling. Am J Physiol - Ren Physiol 316:F1236–F1243.

45. Hao L, Shi Y, Thomas S, Vezina CM, Bajpai S, Ashok A, Bieberich CJ, Ricke WA, Li L. 2018. Comprehensive urinary metabolomic characterization of a genetically induced mouse model of prostatic inflammation. Int J Mass Spectrom 434:185–192.

46. Thomas S, Hao L, Delaney K, McLean D, Steinke L, Marker PC, Vezina CM, Li L, Ricke WA. 2020. Spatiotemporal Proteomics Reveals the Molecular Consequences of Hormone Treatment in a Mouse Model of Lower Urinary Tract Dysfunction. J Proteome Res 19:1375–1382.

TABLES AND FIGURES

Table 1. Potentially disrupted metabolic pathways via MetaboAnalyte 4.0 and KEGG pathway analysis. Matched metabolites include significantly changed metabolites (red font) and other identified metabolites whose quantitative values were not significantly different between the two groups (p-value is from Fisher's exact test).

Metabolic Pathway	KEGG ID	Matched Metabolites	p-value
Tryptophan metabolism	Map00380	melatonin, serotonin, 2-aminomuconic acid semialdehyde, N-methylserotonin, L-kynurenine, indoleacetaldehyde, tryptamine, N-acetylserotonin, 4-(2-amino-3-hydroxyphenyl)-2,4-dioxobutanoic acid, 4-(2-Aminophenyl)-2,4-dioxobutanoic acid, 5-hydroxyindoleacetic acid, L-3-hydroxykynurenine	6.07E-06
Arginine and proline metabolism	Map00330	creatine, putrescine, argininosuccinic acid, citrulline, L-arginine, L-glutamic acid, N-acetylornithine, L-proline, hydroxyproline, gamma-aminobutyric acid, 4-aminobutyraldehyde, N-acetylputrescine	1.82E-05
Histidine metabolism	Map00340	histamine, 1-methylhistamine, Imidazoleacetic acid, 1-Methylhistidine, L-Glutamic acid	2.67E-04
Glycine, serine and threonine metabolism	Map00260	creatine, L-serine, L-threonine, aminoacetone, L-cysteine, L-2-amino-3-oxobutanoic acid	1.60E-02
Alanine, aspartate and glutamate metabolism	Map00250	argininosuccinic acid, L-alanine, L-glutamic acid, gamma-aminobutyric acid, 5-phosphoribosylamine	2.03E-02

Table 2. 30 proteins with the greatest magnitude of change. The p-values are from a paired Student's t-test.

Accession	Protein Name	Gene	Fold change	p-value (unequal)
Q06890	Clusterin	CLUS	2.38	1.1E-03
Q6UGQ3	Secretoglobin family 2B member 2	SG2B2	12.27	1.1E-03
Q924C5	Alpha-protein kinase 3	ALPK3	9.03	1.2E-03
E9PV24	Fibrinogen alpha chain	FIBA	26.68	1.8E-03
P21981	Protein-glutamine gamma-glutamyltransferase 2	TGM2	11.34	1.9E-03
Q8BZT5	Leucine-rich repeat-containing protein 19	LRC19	3.34	2.3E-03
Q61089	Frizzled-6	FZD6	13.42	2.5E-03
Q61592	Growth arrest-specific protein 6	GAS6	3.34	2.6E-03
Q91WR8	Glutathione peroxidase 6	GPX6	4.31	2.7E-03
Q01339	Beta-2-glycoprotein 1	APOH	3.76	2.7E-03
Q8CCH2	NHL repeat-containing protein 3	NHLC3	5.27	3.2E-03
Q9DAU7	WAP four-disulfide core domain protein 2	WFDC2	3.07	3.3E-03
P12961	Neuroendocrine protein 7B2	7B2	5.14	4.2E-03
Q9ESB3	Histidine-rich glycoprotein	HRG	2.62	4.7E-03
P63005	Platelet-activating factor acetylhydrolase IB subunit alpha	LIS1	8.02	4.8E-03
O70251	Elongation factor 1-beta	EF1B	13.10	4.8E-03
P16675	Lysosomal protective protein	PPGB	2.86	4.9E-03
P01132	Pro-epidermal growth factor	EGF	2.31	5.1E-03
Q6GU68	Immunoglobulin superfamily containing leucine-rich repeat protein	ISLR	4.81	5.4E-03
Q61029	Lamina-associated polypeptide 2 isoforms beta/delta/epsilon/gamma	LAP2B	12.40	5.6E-03
Q9R1Q9	V-type proton ATPase subunit S1	VAS1	7.63	5.7E-03
P81117	Nucleobindin-2	NUCB2	7.30	5.8E-03
Q9ESD1	Prostasin	PRSS8	3.85	6.5E-03
O70570	Polymeric immunoglobulin receptor	PIGR	6.88	6.7E-03
P11276	Fibronectin	FINC	4.60	7.0E-03
O88792	Junctional adhesion molecule A	JAM1	6.70	7.2E-03
Q61646	Haptoglobin	HPT	27.48	7.2E-03
Q04646	Sodium/potassium-transporting ATPase subunit gamma	ATNG	3.09	8.3E-03
Q8BFR4	N-acetylglucosamine-6-sulfatase	GNS	3.37	8.5E-03
P05533	Lymphocyte antigen 6A-2/6E-1	LY6A	4.97	9.2E-03

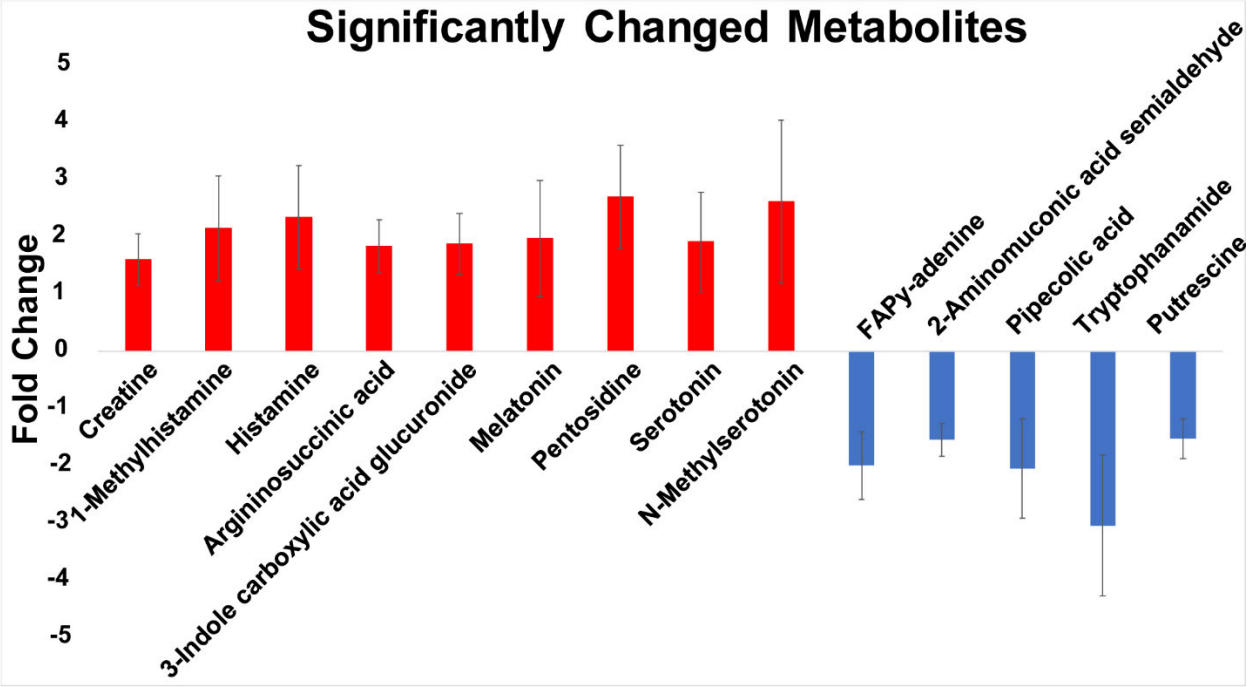


Figure 1. Significantly changed metabolites. 14 metabolites were identified as statistically significant biomarker candidates as shown here by a fold change (bacterial-induced prostatic inflammation mice with respect to control mice). 9 metabolites were up-regulated (red bars) and 5 metabolites were down-regulated (blue bars).

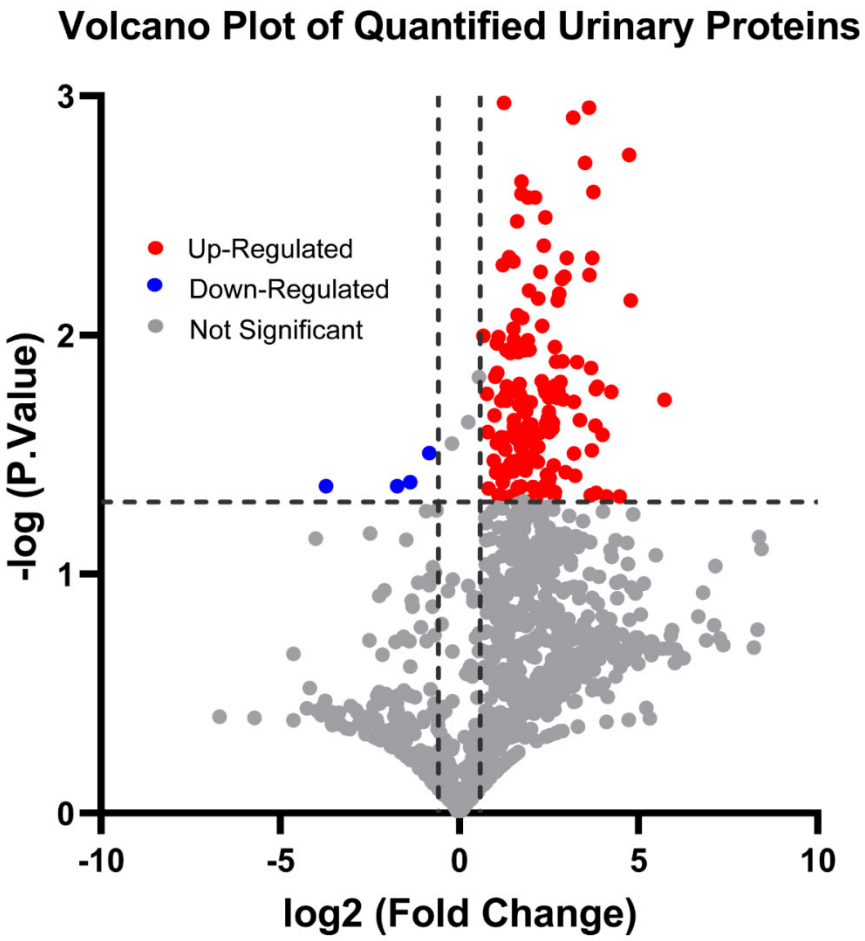


Figure 2. Volcano plot of quantified urinary proteins. The volcano plot of quantified urinary proteins reveals that 168 such proteins were significantly changed in bacterial-induced prostatic inflammation mice compared to control samples. Proteins with fold changes > 1.5 and p-values < 0.05 are shown (red spots: up-regulated; blue spots: down-regulated). Significant proteins were further characterized through literature searches and GO-term enrichment analysis.

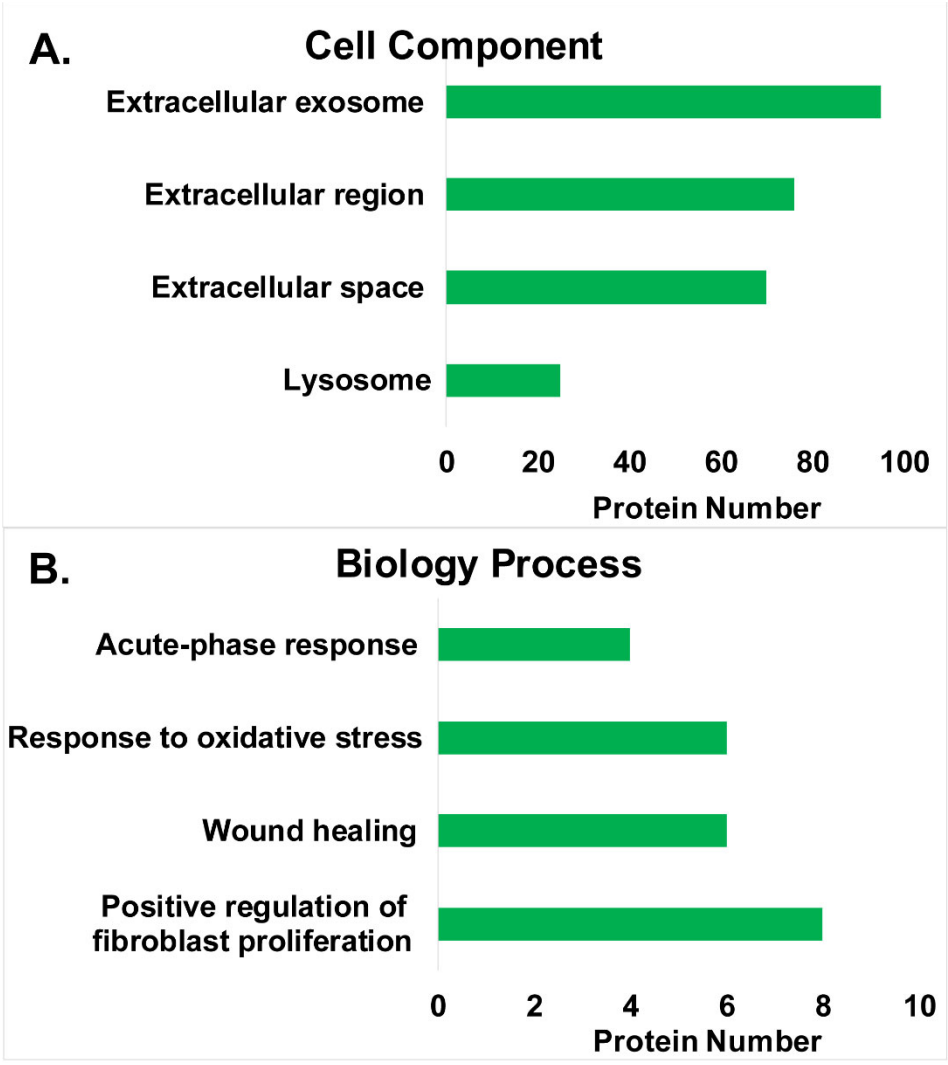


Figure 3. Gene ontology analysis of significantly changed proteins. A: Cell component (top four categories) of significantly changed proteins. **B:** Biology processes (inflammation-related) of significantly changed proteins.

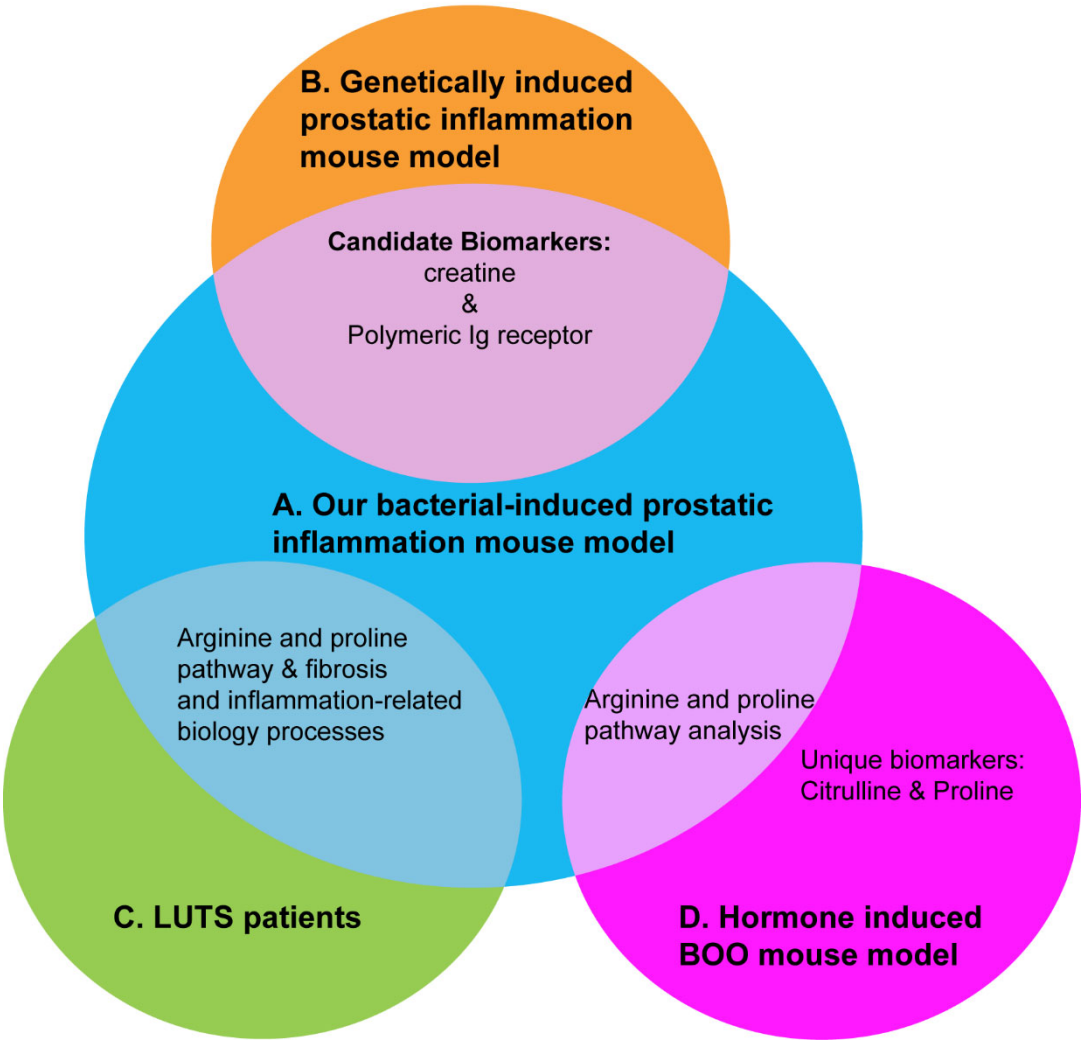
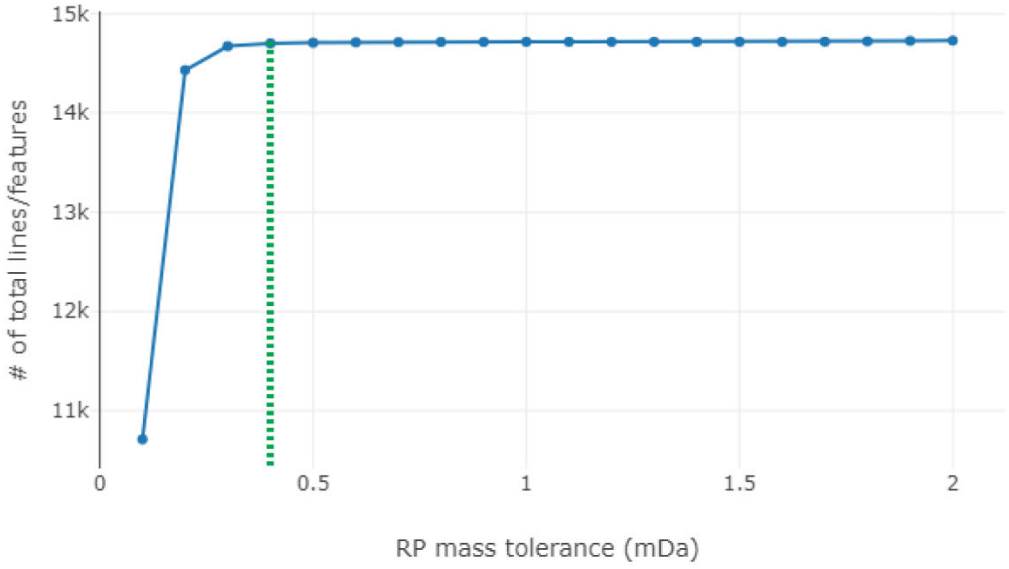


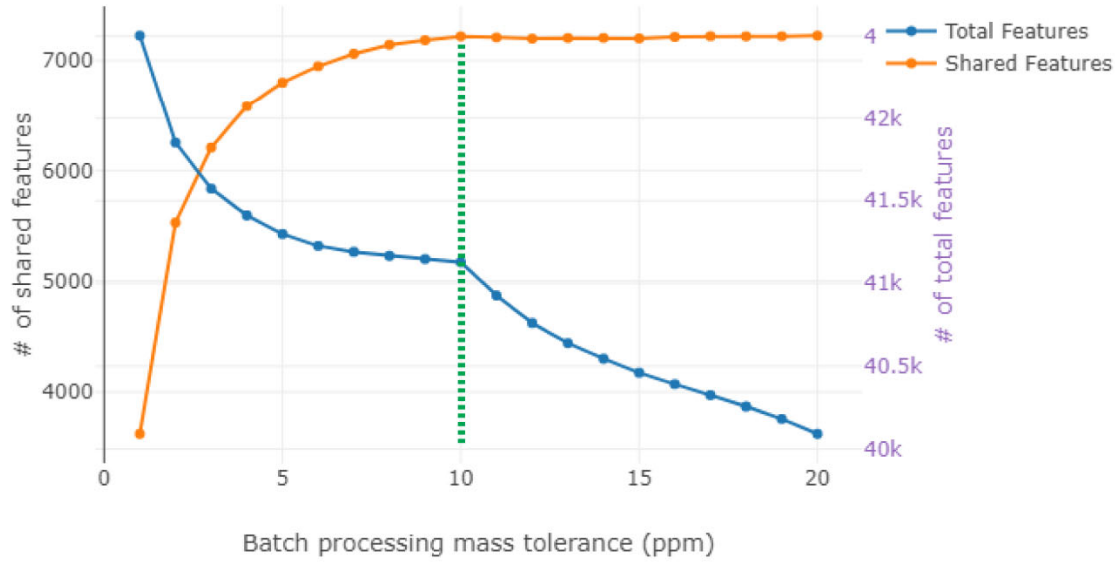
Figure 4. Result comparison with the previous study. A: Our bacterial-induced prostatic inflammation mouse model. **B:** Genetically induced prostatic inflammation mouse model. **C:** LUTS patients. **D:** Hormone induced BOO mouse model.

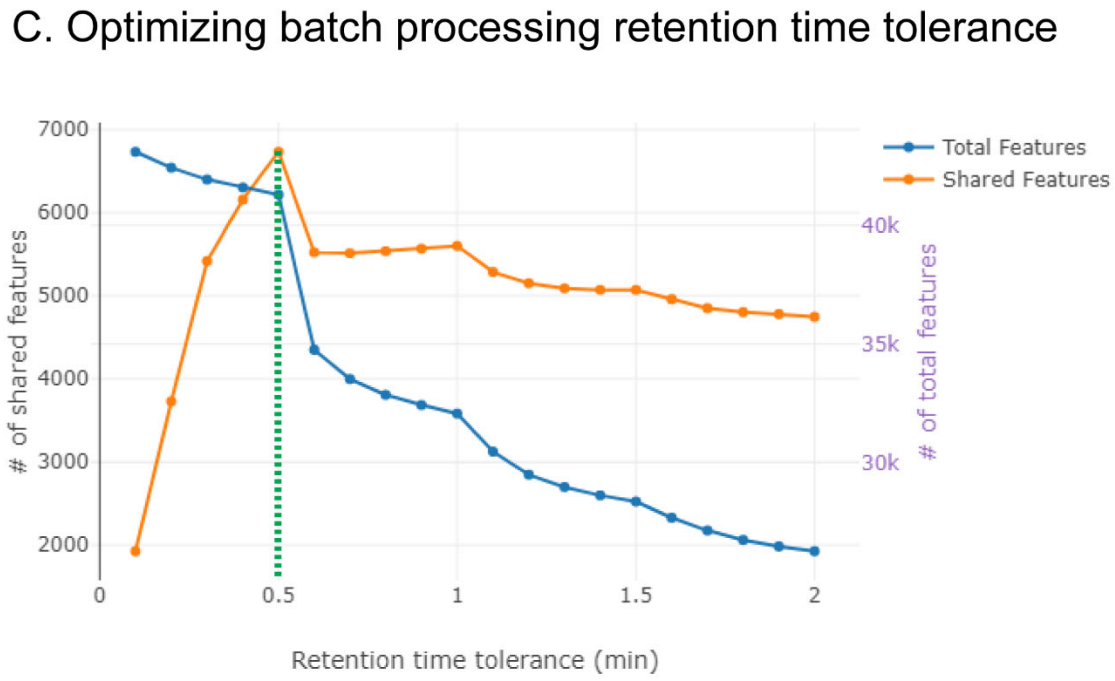
SUPPLEMENTARY INFORMATION

A. Optimizing reporter ion mass tolerance



B. Optimizing batch processing mass tolerance





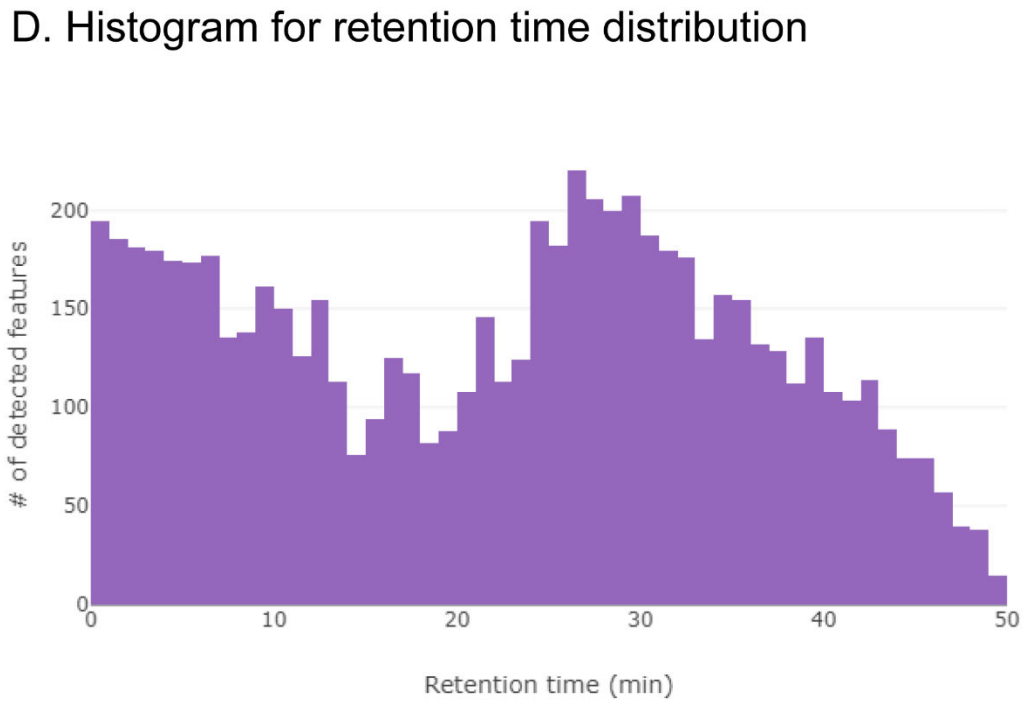


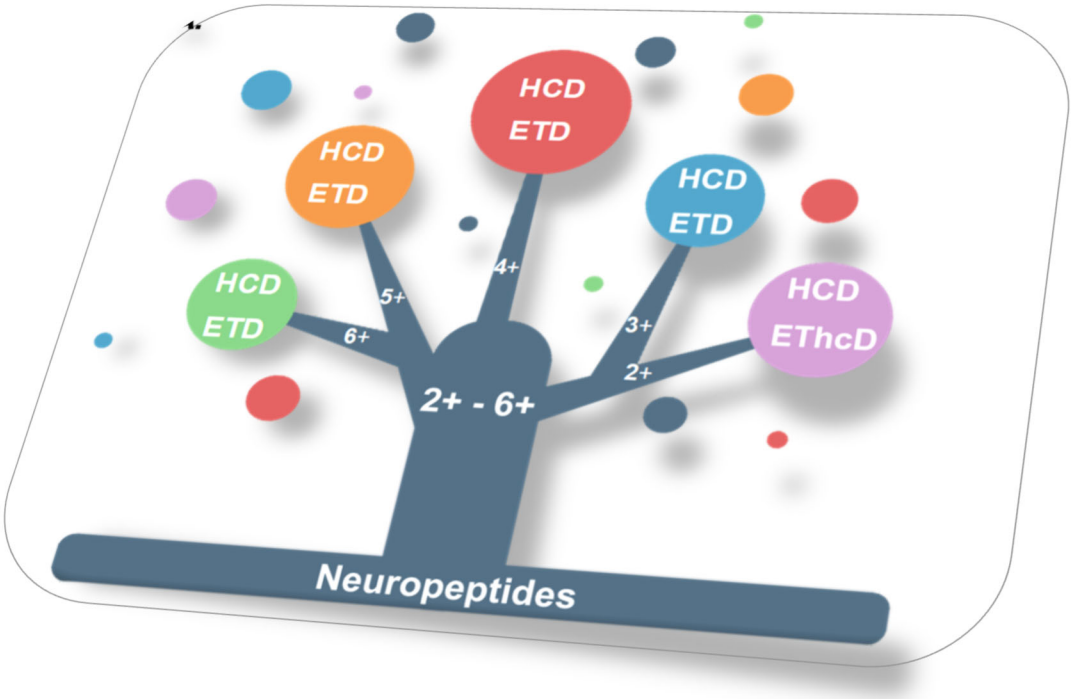
Figure S1. Metandem parameter optimization results: A: Optimizing reporter ion mass tolerance; 0.4 mDa was chosen for the final data analysis. **B:** Optimizing batch processing mass tolerance; 10 ppm was chosen. **C:** Optimizing batch processing retention time tolerance; 0.5 min was chosen. **D:** Histogram for retention time distribution.

Table S1. Sample labeling map (randomized) for metabolites samples

Sample Labeling Map			
Bacterial #1	Bacterial #2	Bacterial #3	Bacterial #4
118b	118a	117c	117a
Control #1	Control #2	Control #3	Control #4
116c	116b	116a	115b

Chapter 4

Development of a Novel Mass Spectrometry Approach for Comprehensive Neuropeptidome Characterization and its Application to the Analysis of Human Pituitary Tumor



Adapted from **Wei P***, Yu Q*, Sun H*, Ma F, Bakshi VP, Sun W, Zheng Z, Zeng Chun, Li LL. "Development of a Novel Mass Spectrometry Approach for Comprehensive Neuropeptidome Characterization and its Application to the Analysis of Human Pituitary Tumor. " *Journal of the American Chemical Society, to be submitted.* *cofirst authors. Wei P and Yu Q, designed and conducted the study under the supervision of LL; Sun H provided samples and helped with sample preparation, Ma F provied the expertise in neuropeptidomics, Bakshi VP Buchberger provided rat brain samples for the method development, Zeng Chun performed the surgury and collected the tumor samples. Sun W and Zheng Z provided insights in human pitutory tumor.

ABSTRACT

Comprehensive characterization of neuropeptides is challenging due to their low abundance, biochemical complexities, and limited neuropeptide database. Therefore, a sensitive mass spectrometry (MS) method is in great demand, and every step in the workflow, from sampling to instrumentation and data analysis, needs to be optimized. Although various sample preparation methods have already been reported, a systematic comparative investigation is still lacking. Also, EThcD is a relatively new ion fragmentation scheme, and its performance on neuropeptide characterization has not been evaluated yet. Herein, a novel MS approach was firstly developed by comparison of three common neuropeptide extraction methods, the one with the highest extraction efficiency was used in subsequent experiments. Deactivating protease activities was proven to decrease the sampling artifact of C-terminal methylation. Secondly, the performance of EThcD on neuropeptide characterization was evaluated in comparison with ETD and HCD, and a decision tree-driven MS scheme was developed to allow improved neuropeptide sequencing by alternating between these fragmentation techniques for each specific charge state. Furthermore, the raw data were searched against both mature neuropeptide database and custom-made neuropeptide preprohormone databases, then applied the detected preprohormones to NeuroPred to predict cleavage probability for novel neuropeptide exploration. With the assistance of this novel MS method, a total of 119 neuropeptides and 69 preprohormones were detected from four human pituitary tumor samples. Also, eleven novel sulfation modifications and four novel neuropeptides have been identified from human pituitary tumor samples. These results demonstrate the efficiency of this approach for both neuropeptide sequencing and post-translational modification (PTM) elucidation. The characterization of these candidate neuropeptides and new sulfation sites lay the foundation for future functional investigations.

KEYWORDS: neuropeptide, mass spectrometry, extraction, EThcD, decision tree, preprohormone, pituitary tumor, novel neuropeptide, sulfation

INTRODUCTION

Neuropeptides are known to modulate neuronal communication and play a key role in regulating the activities of the brain (1–5). Recently, mass spectrometry (MS) has evolved as a powerful tool to characterize neuropeptides, due to its high speed, great sensitivity and unbiased detection of a wide range of peptides present in a complex mixture (6–9). Nevertheless, a comprehensive characterization of neuropeptide remains to be a challenging task due to low abundances and biochemical complexities of neuropeptides, as well as the limited neuropeptide database available for study (9–12).

Increasing extraction efficiency, minimizing extraction artifact, and improving MS method sensitivity are critical to improving the characterization of low abundance neuropeptides in complex biochemical samples (13–16). C-terminal methylation, one of extraction artifacts, can occur during neuropeptide extraction and cause false identification. It could happen during the sample preparation process when both water and methanol exist, which may cause neuropeptide misidentifications. For example, NFDEIDRSGF*G-OMe* may be misidentified as NFDEIDRSGF*A*, since both sequences would yield nearly identical mass spectra (13). Therefore, the heat stabilization to deactivate protease activity method was evaluated to explore whether this step could inhibit this artifact when the extraction buffer is a water and methanol solution. Also, an extraction buffer, which allows sufficient neuropeptide recovery and preserves native

modification, is especially in need. Although various sample preparation methods have already been reported, a thorough comparative investigation is still lacking. In this work, three common extraction methods were compared and evaluated based on the number (No.) of neuropeptides identified. Since some of the brain tissue is tiny, the extraction method should be efficient for both large brain tissue and tiny brain tissue (<3mg). The better neuropeptide extraction method was further employed for the neuropeptidome analysis of the human pituitary tumor samples.

Fragmentation is essential for peptide sequence characterization since this technique allows the generation of the structure information of a peptide by ion dissociation (17). Smaller m/z fragments would be produced during this process, measuring the values of these pieces and then assembling them could generate the original peptide sequence (18). Therefore, choosing a suitable fragmentation technique is critical for comprehensive neuropeptide identification. Electron transfer dissociation (ETD) and beam-type collision-induced dissociation (HCD) are the two common fragmentation techniques employed in the peptidomics and proteomics studies. Electron Transfer Higher Energy Collision Dissociation (EThcD), a recently developed ion fragmentation scheme available on Thermo Scientific Orbitrap instruments, has emerged as a hybrid method combining ETD and HCD fragmentation techniques (23, 24). It is well-known that each fragmentation technique has its preferences regarding the charge state of a peptide/protein ion, the length, and the amino acid components (17). For example, ETD tends to perform better than HCD on higher charge states but yields the lowest number of total identifications due to its slower scan rate (26, 27). HCD facilitates isobaric labeling-based peptide quantification because of its ability for better detection of reporter ions in the low mass region (25). Since EThcD is a relatively new fragmentation technique, its ion preference and performance on

neuropeptide sequencing are to be explored. Herein we comparatively evaluated these three fragmentation techniques and developed a decision tree-driven MS scheme to allow improved neuropeptide sequencing by alternating between these fragmentation modes for each specific charge state.

Despite the large array of neuropeptides present in the brain (28) to precisely regulate animal behaviors (29), only 294 are included in the mature rat neuropeptide database and 283 in homo neuropeptide database (30). A lack of comprehensive database is a major bottleneck for neuropeptide identification. Based on the neuropeptide biosynthesis process, the N-terminal signal peptide sequence is firstly removed from preprohormone by signal peptidase. Then the obtained prohormone will be cleaved by endopeptidases at the N-terminus between two basic residues or the C-terminus of the paired basic residues. Finally, basic residue extensions are removed by carboxypeptidase E and Peptidyl-glycine alpha-amidating monooxygenase (PAM) to yield the bioactive neuropeptide (25) (31). Therefore, to be classified as a candidate neuropeptide, two features were required: 1) Its preprohormone needs a signal sequence; 2) At least one terminal of experimentally measured peptide flanked by basic cleavage residues. In this study, besides a mature neuropeptide database, a custom-made candidate preprohormone database, which includes all the signal sequence containing proteins, was also used for untargeted neuropeptide searching. Neuropred, a statistical neuropeptide prediction algorithm (32), was used to explore novel neuropeptides in human pituitary tumor samples.

MATERIALS AND METHODS

Tissue Isolation

All animal procedures were approved by the University of Wisconsin-Madison Animal Care and Use Committee (33). Male Sprague Dawley rats (Charles River, Wilmington, MA) were used for the method development study. After euthanized with isoflurane, rat brains were quickly removed after fast decapitation. The entire brain samples and hypothalamus samples were placed on dry ice immediately after isolation, then transferred to -80 ℃ for storage. Human pituitary tumor samples were collected during surgery by Beijing Tian Tan Hospital and stored at -80℃ until further processing.

Heat Denature Evaluation

Frozen rat brain tissues were used for both the denature group and the non-denature group to evaluate the effect of heat denaturation. Denature step was using STABILIZOR T1 with frozen tissue mode. Then both the denatured tissue and non-denature tissue were weighed, following with the same analysis step. Briefly, ice-cold extraction buffer (water/methanol/acetic acid (1:90:9, v/v/v)) (34–37) was added based on the ratio of buffer volume to tissue weight, 3 μL/1mg. After homogenizing by Sonic Dismembrator (8 secs on, 15 secs off, three circles, Fisher Scientific Model FB120), samples were centrifuged (20,000 *xg* at 4°C, 20min, Eppendorf Centrifuge 5424 R). Then the supernatant was loaded on the pre-rinsed Ultracel-30K 15ml Molecular Weight Cut Off (MWCO) filter pre-rinse with (i) 0.1M sodium hydroxide 1200ul, 4000G at 4°C, 5min (ii) Water/Methanol/Acetonitrile solution (50:30:20, v/v/v) 1200μL, 4000G at 4°C, 5min, repeat one time (iii) Extraction buffer, 4000 *xg* at 4°C, 5min. The flow-through part was collected and lyophilized. Sep-Pak C18 Cartridges were used for the sample desalting. Peptide Assay (Pierce™ Quantitative Colorimetric Peptide Assay) was used for peptide concentration determination and injection volume normalization before injecting into LC-MS. Ultimate 3000 UPLC system

coupling with Orbitrap Fusion™ Tribrid™ Mass Spectrometer. HCD fractionation method was adopted here for the heat denature evaluation.

Extraction Buffer Evaluation

Extraction buffers were evaluated with both large brain tissue (entire brain) and tiny brain tissue (hypothalamus). Three common neuropeptide extraction buffers, low acidified methanol (LAM, water/methanol/acetic acid (79.8:20:0.2, v/v/v) and water/methanol/acetic acid (49.8:50:0.2, v/v/v)) (38), high acidified methanol buffer (HAM, water/methanol/acetic acid (1:90:9, v/v/v)) and 8M Urea buffer (39), were applied to the brain tissues. Both HAM and 8M Urea buffers were used as extraction buffer and followed the workflow above. Since LAM method has two different buffers, to be consistent with the buffer volume-weight ratio with the other two methods, 1.5 μL/mg of water/methanol/acetic acid solution (79.8:20:0.2, v/v/v) and 1.5 μL/mg water/methanol/acetic acid solution (49.8:50:0.2, v/v/v) were added to tissue sample, respectively during homogenization step. The first step aqueous extractions and second step organic extractions were mixed before MWCO. The following steps were the same as the other two extraction buffer.

Also, to minimize sample loss for handling small tissues such as the hypothalamus, the buffer volume-weight ratio was increased to 20 μL/mg. MWCO filter was changed to 0.5 ml Microcon YM-30 (Millipore; 20 min, 15,000 *xg* at 4 °C). 0.5 ml MWCO filters were pre-rinsed with the same solutions as 15 ml filters but used 200 μL volume instead (15,000 *xg* at 4°C, 10min, Eppendorf Centrifuge 5424 R)

Evaluation of EThcD, ETD and HCD Fragmentation Patterns Performance

After BCA normalization, an equal amount of peptide sample was injected into LC-MS. For the MS fragmentation method, every charge stage from charge 2 to charge 6 was evaluated with EThcD, ETD, and HCD fragmentation patterns, respectively. LC-MS method has been optimized under each fractionation. Peaks 8.5 was employed to evaluate the fragmentation performance for each charge state. The following filters, false discovery rate (FDR) lower than 1%, PTM Ascore equal to or higher than 20, and at least one unique peptide was detected for each identified neuropeptide, were used to increase the confidence of identification. Mature neuropeptide database downloaded from NeuroPep was used for the database searching, and both the intact neuropeptides and truncated neuropeptides were used for the three fragmentation pattern evaluations.

Novel Neuropeptide Exploration

After the heat denature evaluation, extraction buffer evaluation, and three fragmentation performance evaluations, a novel method was developed and applied to the human pituitary tumor neuropeptide study. Besides the mature neuropeptide database, a custom-made candidate preprohormone database, which includes all the signal sequence containing proteins, was also used for novel neuropeptide exploration. A house-made candidate preprohormone database was begun with downloading the homo database from Uniprot (40). Then the signal sequence containing proteins were extracted with Signal P (41) to form the preprohormone database. After the preprohormone database searching, the identified preprohormones from human pituitary tumors were applied to Neuropred for cleavage probability predicted. Cleavage probability higher than 0.9 was used for the candidate neuropeptide exploration.

RESULTS AND DISCUSSION

Heat Denaturation Evaluation

Both denature and non-denature groups identified same 89 neuropeptides with a mature neuropeptide database. However, four C-terminal methylation modified peptides were identified from the non-denature group, while none was detected from the denature group. From **Table 1**, the four C-terminal methylated peptides identified in the non-denature group were also found in the denature group but without methylation modification, which demonstrates that those four C-methylation modifications did not belong to the original sequences but from the sample preparation artifact. Therefore, heating stabilization conducted rapid heating to preserve samples from human artifact during sampling, which was a recommended step to increase the accuracy and quality of analytical results.

Extraction Buffer Evaluation

HAM, LAM, and Urea buffers were comparatively evaluated for the neuropeptide recovery rate with both large brain tissue (the entire brain) and tiny tissue (hypothalamus). According to **Fig. 1 A** and **B**, Urea and HAM buffers showed better extraction efficiency in both the hypothalamus and the entire brain, when compared with the LAM buffer. Therefore, Urea and HAM was further considered. According to **Fig. 1 C** and **D,** HAM extraction method could enable the identification of more unique neuropeptides when compared with the Urea method in both the hypothalamus and the entire brain tissues. Furthermore, since brain tissue is lipid-rich tissue, and the urea buffer offers good solubility for both lipids and neuropeptides. When using urea buffer for neuropeptide extraction, the lipid in the extraction solution would stuck on the MWCO filters and stop the other neuropeptide from flowing through, which made the centrifuge time much longer than the HAM

buffer. Therefore, HAM exhibits the best performance for neuropeptide study in the brain, as it yields the highest recovery efficiency, and it is easy to work with.

Evaluation of EThcD, ETD and HCD Fragmentation Pattern Performance

After going through the ion source, neuropeptides were charged, and their charge states commonly ranged from 2 to 6, due to the relatively small size of most neuropeptides. Therefore, ions preference analysis of EThcD, ETD, and HCD was based on the charge states from 2 to 6. In **Fig 2**, firstly, the bar chart indicates HCD identified the highest number of peptides in every charge states, followed by EThcD and ETD. Secondly, although from the bar chart, the total peptide number identified by EThcD is higher than ETD, Venn diagram shows ETD, which has high charge ion preference as discussed in the background, identified more unique peptides at charge states 3, 4 and 5 than EThcD. Thirdly, from the Venn diagram, most identified peptides of EThcD are overlapped with HCD. Furthermore, EThcD and HCD exhibit similar trends, with identification rates decrease as the charge states increase, which indicates EThcD prefers lower charge states ions than higher charge state. Based on all the statistical comparison results from **Fig 2**, a data-dependent decision tree method was established and illustrated in **Fig 3A**. Briefly, HCD was used as the main fragmentation mode due to its superior performance at all charge states. EThcD was combined at charge state 2, and ETD at charge states 3-6 as a complementary fragmentation mode to maximize the peptide coverage for comprehensive neuropeptide characterization. This data-dependent decision tree approach was further evaluated. The same brain sample was analyzed by both the decision tree method and the MS methods with three fragmentation modes operated individually. The neuropeptide identification results in **Fig 3B** show that the single run of the decision tree method could identify similar numbers of neuropeptides

from the sum of three separate runs with ETD, EThcD, and HCD, which verified the high efficiency of neuropeptide characterization using this novel decision tree method.

Neuropeptidome Characterization in Human Pituitary Tumor

After the method evaluation and optimization, a novel approach was developed, which starting from a heat denaturation to remove the sampling artifact. Then HAM buffer was added based on the pituitary tumor weight. A high pH reversed-phase peptide fractionation kit was specifically used here to improve the neuropeptide discovery. Then injected peptide mixture into the LC-MS with the data dependence decision-tree method. Raw data were searched against both the mature neuropeptide database and the preprohormone database. Neuropeptides identification, neuropeptide PTM, and novel neuropeptide were further explored.

From the mature database searching, in total, 6480 peptides were detected from 119 mature neuropeptides, which belonging to 79 preprohormones. In **Fig 4A**, the bar graph shows the total identified neuropeptide number from each human pituitary tumor. The Venn diagram in **Fig 4B** shows the shared neuropeptides between four pituitary tumor samples. 39 neuropeptides **(Table S1)** were observed in all four pituitary tumors, including Neurosecretory protein VGF, which may regulate cell-cell interactions, Secretogranin-1, and Secretogranin-2 which could regulate the neuroendocrine secretion, and ProSAAS which control the neuroendocrine secretory pathway.

Furthermore, 27 sulfotyrosine peptides (**Table S2**) were detected that belong to 5 human neuropeptides. Notably, 11 novel sulfotyrosine sites were confidently identified. Seven of them

were from neuropeptide Secretogranin-1. When compared with the sulfation sites with the previous study (**Fig 5A**), three sulfation sites 321, 454, and 546 were reported in Uniprot by similarity. By similarity means, these sulfotyrosine sites have been found in other organisms but have not yet been detected in human neuropeptides. Along with the other four novel sulfotyrosine sites in 369, 506, 548, and 581, seven sulfotyrosine sites were identified in the human for the first time, which greatly expanded the knowledge of sulfotyrosine sites in neuropeptide Secretogranin-1. **Fig 5B** shows a mass spectrum of a sulfotyrosine peptide Y.Y(+79.96)DPLQWKSSHFE.R from neuropeptide Secretogranin-1. According to the neutral loss of 80 Da, sulfotyrosine could be distinguished from the phosphotyrosine with modification sites being localized. Commonly, PTM of neuropeptides could increase the complexity of neuropeptide structure and lead to more challenges for identification. Sulfation is especially tricky because of the similarity of sulfation and phosphorylation, exhibiting the same mass change after modification. However, this challenge was solved by this novel decision tree method, even without special sulfation enrichment. The alternation between different fragmentation modes could not only improve the efficiency of neuropeptide identification but also enable confident elucidation of the neuropeptide PTM and modification sites.

Novel Neuropeptides in Human Pituitary Tumor

After searching the raw data with the preprohormone database, 586 peptides that belong to 69 preprohormones were found in all four human pituitary samples. The results of the candidate preprohormone database were significantly expanded in comparison with the mature neuropeptide database searching. After submitting the FASTA formatted sequence of detected candidate preprohormones to NeuroPred under the mammalian model, the potential cleavage sites and

cleavage probability were exported (32). Basic cleavage site Lys-Arg (K-R) and possible pairs Arg-Arg, Lys-Lys, and Arg-Lys, which may also be cleaved in some prohormones to produce active neuropeptides, were included to evaluate the possible cleavage sites (42). **Table 2** shows the four cleavage sites were found with the cleavage probability of more than 90%, notably, three of the cleavage sites have cleavage probability of more than 95%. 90% was used here to avoid the false positive cleavage site.

Among those four, Platelet-derived growth factor D, Augurin, and Meteorin have not been reported as neuropeptide preprohormones, while two mature neuropeptides have been identified from Neuropeptide W. WYKHVASPRYHTVGRAAGLLMGL (NP03746) and WYKHVASPRYHTVGRAAGLLMGLRRSPYLW (NP03747). In particular, peptide SSQAGIPVRAP has also been detected an experiment designed for the discovery of bioactive peptides by experimentally identifying signal sequence cleavage sites and processing sites (43). Meteorin, a secreted protein that regulates glial cell differentiation and promotes axonal extension (44). It is relatively newly described as a neurotrophic factor and reported as over-expressed in the striatum following excitotoxic injury (45).

Augurin is a secreted peptide, a putative hormone that may induce senescence of oligodendrocyte and neural precursor cells. According to gene ontology annotation (GO), its molecular function is the neuropeptide hormone, and the biological process is the neuropeptide signaling pathway and negative regulation of cell proliferation. It is hypothesized that augurin is a novel peptide hormone likely to be involved in physiological homeostasis [29]. Also, it is reported as a novel hypothalamic

neuropeptide that stimulates the hypothalamus-pituitary-adrenal axis [30]. Notably, a hidden Markov model tool, a bioinformatics-driven discovery method has also been predicted that Augurin is a putative neuropeptide precursor from the human genome (48). Thus, its detection in our study further suggest that augurin is a candidate neuropeptide preprohormone.

Platelet-derived growth factor D (PDGF D) is a growth factor that plays an essential role in the regulation of embryonic development, cell proliferation, cell migration, survival, and chemotaxis. Although PDGF D activation and human cancer progression are currently under investigation, the elevated PDGF D expression which has been observed and reported in many ovarian, lung, renal, and brain cancer-derived cell lines, sera, and cancer tissues supports such a hypothesis that PDGF D is aberrantly localized in some specific tumor tissues and may play a role in tumorigenesis (49). Since PDGF D is a signaling molecule (growth factor) specifically involved in the tumor tissues, it is likely to act as a crucial tumor-related neuropeptide precursor that could be used for tumor therapy.

CONCLUSIONS

A novel approach was developed for more comprehensive neuropeptide characterization. Firstly, three extraction buffers were comparatively evaluated, and HAM has been selected as the best extraction method for neuropeptide sampling. HAM buffer has the highest extraction efficiency for both large and tiny brain tissues when compared with the LAM extraction buffer. It is also a friendly extraction buffer to use when compared with the 8M urea buffer. Secondly, from the comparative investigation of three fragmentation techniques, HCD has the superior performance

for neuropeptide characterization in all charge states. While EThcD could provide more complementary results at charge state 2 and ETD provide more identifications at charge states 3-6. Based on this result, a decision tree-driven MS scheme was developed to allow improved neuropeptide sequencing by alternating between these fragmentation techniques for each specific charge state. Thirdly, applying this novel MS approach to the human pituitary tumor study, a total of 119 neuropeptides were detected from four human pituitary tumor samples. Not only the neuropeptide sequencing but the site-specific PTMs characterization was also improved by this novel method. Eleven new sulfotyrosine sites were identified in neuropeptides from the human pituitary tumor samples without special enrichment steps. Furthermore, with the preprohormone database, four candidate neuropeptides have been identified in human pituitary tumor samples. The results demonstrated the efficiency of this novel approach for neuropeptide characterization. Future work will focus on novel neuropeptide evaluation and functional exploration.

ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health through Grant NIH R01 DK071801. The Orbitrap instruments were purchased through the support of an NIH shared instrument grant NIH-NCRR S10RR029531 (L.L.) and Office of the Vice-Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

REFERENCES

1. van den Pol AN. 2012. Neuropeptide transmission in brain circuits. Neuron 76:98–115.

2. Han B, Fang Y, Feng M, Hu H, Qi Y, Huo X, Meng L, Wu B, Li J. 2015. Quantitative Neuropeptidome Analysis Reveals Neuropeptides Are Correlated with Social Behavior Regulation of the Honeybee Workers.

3. Bakos J, Zatkova M, Bacova Z, Ostatnikova D. 2016. The Role of Hypothalamic Neuropeptides in Neurogenesis and Neuritogenesis. Neural Plast 2016:1–10.

4. Lovinger DM. 2008. Communication networks in the brain: neurons, receptors, neurotransmitters, and alcohol. Alcohol Res Health 31:196–214.

5. Thakker-Varia S, Alder J. 2009. Neuropeptides in depression: role of VGF. Behav Brain Res 197:262–78.

6. Schmerberg CM, Li L. 2013. Function-driven discovery of neuropeptides with mass spectrometry-based tools. Protein Pept Lett 20:681–94.

7. Secher A, Kelstrup CD, Conde-Frieboes KW, Pyke C, Raun K, Wulff BS, Olsen J V. 2016. Analytic framework for peptidomics applied to large-scale neuropeptide identification. Nat Commun 7:11436.

8. Li L, Kelley WP, Billimoria CP, Christie AE, Pulver SR, Sweedler J V., Marder E. 2003. Mass spectrometric investigation of the neuropeptide complement and release in the pericardial organs of the crab, Cancer borealis. J Neurochem 87:642–656.

9. Lee JE. 2016. Neuropeptidomics: Mass Spectrometry-Based Identification and Quantitation of Neuropeptides. Genomics Inform 14:12–9.

10. Zhang X, Petruzziello F, Zani F, Fouillen L, Andren PE, Solinas G, Rainer G. 2012. High

Identification Rates of Endogenous Neuropeptides from Mouse Brain. J Proteome Res 11:2819–2827.

11. Van Wanseele Y, De Prins A, De Bundel D, Smolders I, Van Eeckhaut A. 2016. Challenges for the *in vivo* quantification of brain neuropeptides using microdialysis sampling and LC–MS. Bioanalysis 8:1965–1985.

12. Van Eeckhaut A, Maes K, Aourz N, Smolders I, Michotte Y. 2011. The absolute quantification of endogenous levels of brain neuropeptides *in vivo* using LC–MS/MS. Bioanalysis 3:1271–1285.

13. Stemmler EA, Barton EE, Esonu OK, Polasky DA, Onderko LL, Bergeron AB, Christie AE, Dickinson PS. 2013. C-terminal methylation of truncated neuropeptides: An enzyme-assisted extraction artifact involving methanol. Peptides 46:108–125.

14. Che F-Y, Zhang X, Berezniuk I, Callaway M, Lim J, Fricker LD. 2007. Optimization of Neuropeptide Extraction from the Mouse Hypothalamus. J Proteome Res 6:4667–4676.

15. El-Karim I, Lundy F., Linden G., Lamey P-J. 2003. Extraction and radioimmunoassay quantitation of neuropeptide Y (NPY) and vasoactive intestinal polypeptide (VIP) from human dental pulp tissue. Arch Oral Biol 48:249–254.

16. Sturm RM, Dowell JA, Li L. 2010. Rat Brain Neuropeptidomics: Tissue Collection, Protease Inhibition, Neuropeptide Extraction, and Mass Spectrometric Analysis, p. 217–226. *In* . Humana Press.

17. Lingdong Quan ML. 2013. CID,ETD and HCD Fragmentation to Study Protein Post-Translational Modifications. Mod Chem Appl 01:1–2.

18. Wysocki VH, Resing KA, Zhang Q, Cheng G. 2005. Mass spectrometry of peptides and proteins. Methods 35:211–222.

19. Frese CK, Altelaar AFM, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, Heck AJR, Mohammed S. 2011. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res 10:2377–2388.

20. Li W, Wysocki VH. 2012. ETD fragmentation features improve algorithm. Expert Rev Proteomics 9:241–3.

21. Shen Y, Toli N, Xie F, Zhao R, Purvine SO, Schepmoes a, Moore RJ, Anderson G a, Smith RD. 2011. peptidomic analysis : comparison of peptide identification methods Effectiveness of CID , HCD , and ETD with FT MS / MS for degradomic-peptidomic analysis : comparison of peptide identification methods 3929–3943.

22. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP. 2011. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. Mol Cell Proteomics 10:M111.009910.

23. Yu Q, Wang B, Chen Z, Urabe G, Glover MS, Shi X, Guo LW, Kent KC, Li L. 2017. Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. J Am Soc Mass Spectrom 28:1751–1764.

24. Frese CK, Zhou H, Taus T, Maarten Altelaar AF, Mechtler K, Heck AJR, Mohammed S. 2013. Unambiguous Phosphosite Localization using Electron-Transfer/ Higher-Energy Collision Dissociation (EThcD).

25. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP. 2011. Evaluation of

HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. Mol Cell Proteomics 10:M111.009910.

26. Lingdong Quan ML, Liu M. 2013. CID,ETD and HCD Fragmentation to Study Protein Post-Translational Modifications. Mod Chem Appl 01:1–2.

27. Guthals A, Bandeira N. 2012. Peptide identification by tandem mass spectrometry with alternate fragmentation modes. Mol Cell Proteomics 11:550–7.

28. Jarecki JL, Frey BL, Smith LM, Stretton AO. 2011. Discovery of Neuropeptides in the Nematode Ascaris suum by Database Mining and Tandem Mass Spectrometry. J Proteome Res 10:3098–3106.

29. Yu Q, OuYang C, Liang Z, Li L. 2014. Mass spectrometric characterization of the crustacean neuropeptidome. EuPA Open Proteomics 3:152–170.

30. Neuropeptide Database - Search.

31. Fleur L. Strand. Neuropeptides: Regulators of Physiological Processes - Fleur L. Strand - Page 47.

32. Southey BR, Amare A, Zimmerman TA, Rodriguez-Zas SL, Sweedler J V. 2006. NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. Nucleic Acids Res 34:W267-72.

33. Wang B, Zhang M, Takayama T, Shi X, Roenneburg DA, Craig Kent K, Guo LW. 2015. BET Bromodomain Blockade Mitigates Intimal Hyperplasia in Rat Carotid Arteries. EBioMedicine 2:1650–1661.

34. Xiang F, Ye H, Chen R, Fu Q, Li L. 2010. *N* , *N* -Dimethyl Leucines as Novel Isobaric

Tandem Mass Tags for Quantitative Proteomics and Peptidomics. Anal Chem 82:2817–2825.

35. Ma M, Bors EK, Dickinson ES, Kwiatkowski MA, Sousa GL, Henry RP, Smith CM, Towle DW, Christie AE, Li L. 2009. Characterization of the Carcinus maenas neuropeptidome by mass spectrometry and functional genomics. Gen Comp Endocrinol 161:320–334.

36. Ma M, Szabo TM, Jia C, Marder E, Li L. 2009. Mass spectrometric characterization and physiological actions of novel crustacean C-type allatostatins. Peptides 30:1660–1668.

37. Chen R, Ma M, Hui L, Zhang J, Li L. 2009. Measurement of neuropeptides in crustacean hemolymph via MALDI mass spectrometry. J Am Soc Mass Spectrom 20:708–18.

38. Zhang X, Petruzziello F, Zani F, Fouillen L, Andren PE, Solinas G, Rainer G. 2012. High identification rates of endogenous neuropeptides from mouse brain. J Proteome Res 11:2819–2827.

39. Secher A, Kelstrup CD, Conde-Frieboes KW, Pyke C, Raun K, Wulff BS, Olsen J V. 2016. Analytic framework for peptidomics applied to large-scale neuropeptide identification. Nat Commun 7:11436.

40. UniProt.

41. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–6.

42. Liu F, Baggerman G, Hertog WD, Verleyen P, Schoofs L, Wets G. 2006. In Silico Identification of New Secretory Peptide Genes in Drosophila melanogaster * □ 510–522.

43. Sasaki K, Satomi Y, Takao T, Minamino N. 2009. Snapshot peptidomics of the regulated

secretory pathway. Mol Cell Proteomics 8:1638–47.

44. Nishino J, Yamashita K, Hashiguchi H, Fujii H, Shimazaki T, Hamada H. 2004. Meteorin: a secreted protein that regulates glial cell differentiation and promotes axonal extension. EMBO J 23:1998–2008.

45. Wright JL, Ermine CM, Jørgensen JR, Parish CL, Thompson LH. 2016. Over-Expression of Meteorin Drives Gliogenesis Following Striatal Injury. Front Cell Neurosci 10:177.

46. Mirabeau O, Perlas E, Severini C, Audero E, Gascuel O, Possenti R, Birney E, Rosenthal N, Gross C. 2007. Identification of novel peptide hormones in the human proteome by hidden Markov model screening. Genome Res 17:320–7.

47. Tadross J, Patterson M, Suzuki K, Beale K, Boughton C, Smith K, Moore S, Ghatei M, Bloom S. 2010. Augurin stimulates the hypothalamo-pituitary-adrenal axis via the release of corticotrophin-releasing factor in rats. Br J Pharmacol 159:1663–1671.

48. Murphy D, Gainer H. 2016. Molecular neuroendocrinology : from genome to physiology.

49. Reardon DA, Michalkiewicz E, Boyett JM, Sublett JE, Entrekin RE, Ragsdale ST, Valentine MB, Behm FG, Li H, Heideman RL, Kun LE, Shapiro DN, Look AT, Wu F, Dhanabal M, Herrmann J, Lichenstein HS. 1997. Extensive genomic abnormalities in childhood medulloblastoma by comparative genomic hybridization. Cancer Res 57:4042–7.

TABLES AND FIGURES

Table 1. Heat denaturation evaluation results. C-terminal methylation was only found in the non-denature group when compared with the denature group.

Neuropeptide Accession	Non-Denature Group	Denature Group
NP01018|Chromogranin/secretogranin|Secretogranin-1	R.LLDEGHDPVHESPVD**(+14.02)**.T	R.LLDEGHDPVHESPVD.T
NP01013|Chromogranin/secretogranin|Chromogranin-A	W.SRMDQLAKELT**(+14.02)**.A	W.SRMDQLAKELT.A
NP01021|Chromogranin/secretogranin|Secretogranin-3	R.ELSAERPLNEQIAEAE**(+14.02)**.A	R.ELSAERPLNEQIAEAE.A
NP00883|CART|Cocaine_and_amphetamine-regulated	Q(-17.03)EDAELQPR**(+14.02)**.A	Q(-17.03)EDAELQPR.A

Table 2. Novel neuropeptide identified from human pituitary tumors. The cleavage sites with the cleavage probability higher than 90%, results from NeuroPred.

Protein Accession	Gene	Peptide	Unique	Start	End	Cleavage Probability
Q8N729|Neuropeptide W	NPW	R.SSQAGIPVRAP.R	Y	105	115	0.9783
Q9GZP0|Platelet-derived growth factor D	PDGFD	R.DESNHLTDLY.R	Y	41	50	0.9620
Q9H1Z8|Augurin	C2orf40	R.EAPVPTKTKVAVDEN.K	Y	42	56	0.9105
Q9UJH8|Meteorin	METRN	R.ALFLQATPHQDIS.R	Y	128	140	0.9962

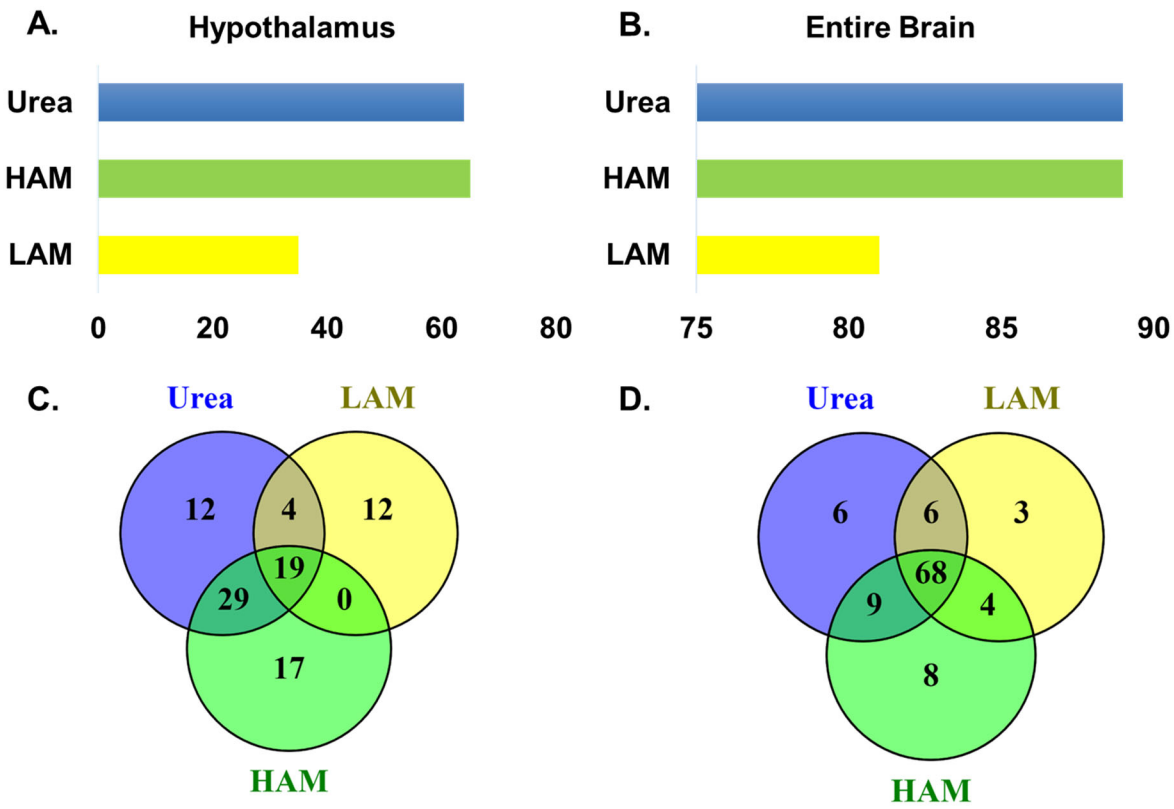


Figure 1. Extraction buffer evaluation result. Panels A and **C** show the neuropeptides identified from the hypothalamus (tiny tissue) with three extraction buffers and the overlap among each buffer extraction. **Panels B** and **D** show the neuropeptides identified from the entire brain (large tissue) with three extraction buffers and the overlap among each buffer extraction.

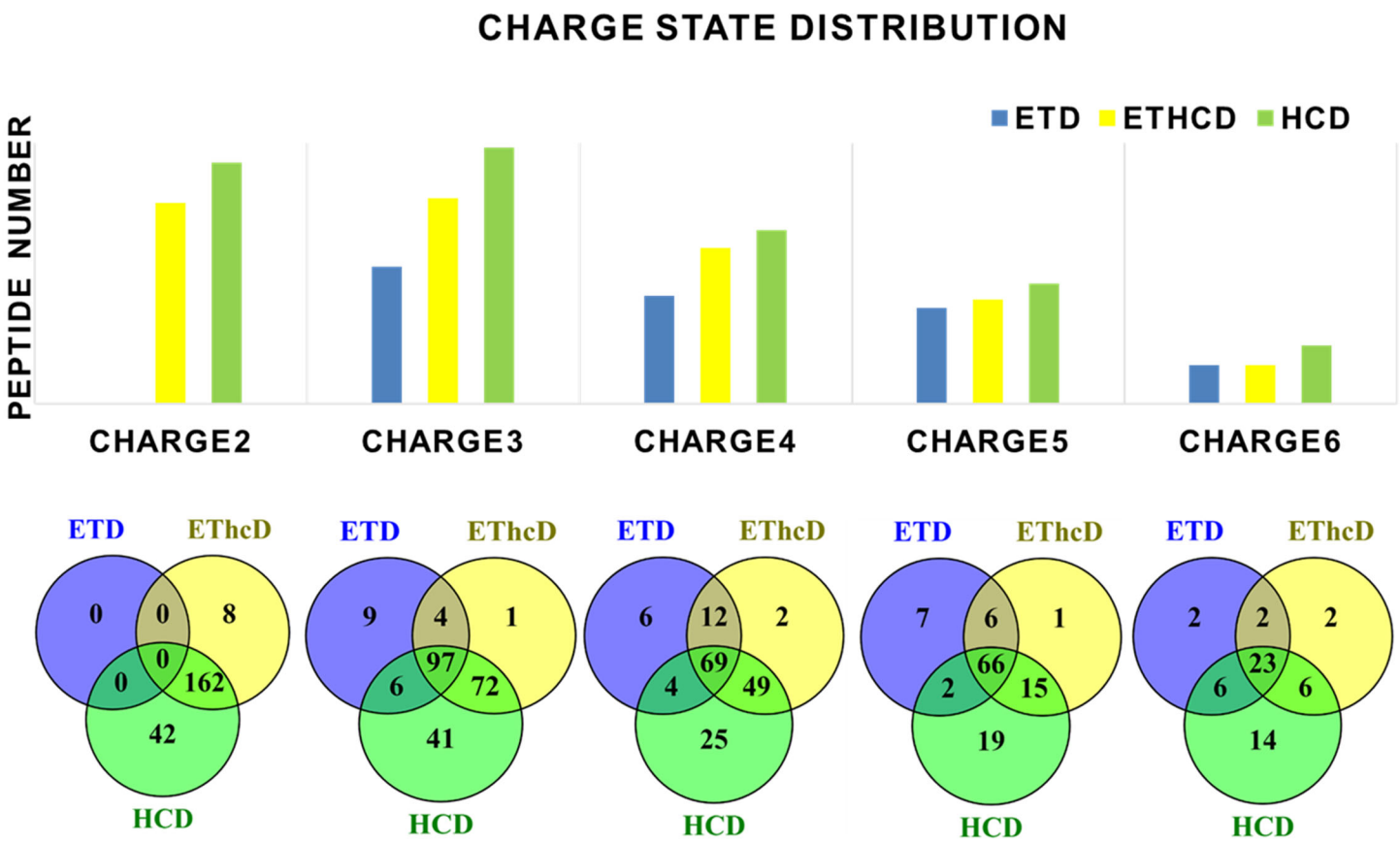


Figure 2. Charge stage distribution with ETD, EThcD, and HCD. The bar chart shows the total identified peptide No. of ETD, EThcD, and HCD at each charge state. The Venn diagram shows the unique and overlap of peptides identified at each charge state. Peptide here refers to truncated neuropeptide.

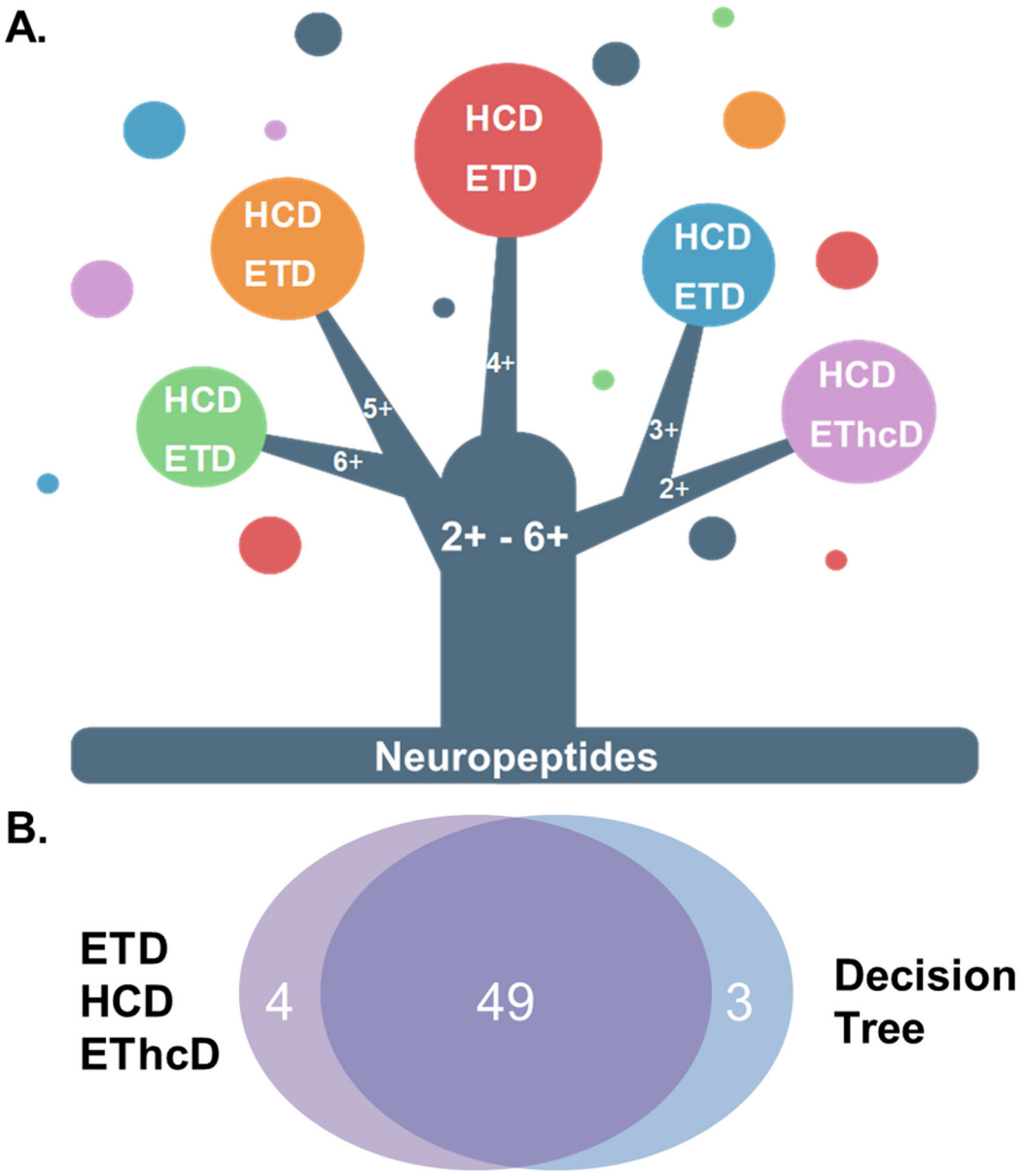


Figure 3. Data-dependent decision tree method. A: Alternation between ETD, EThcD, and HCD fragmentation techniques for each charge state; **B:** Evaluation of the decision tree method with three individual run.

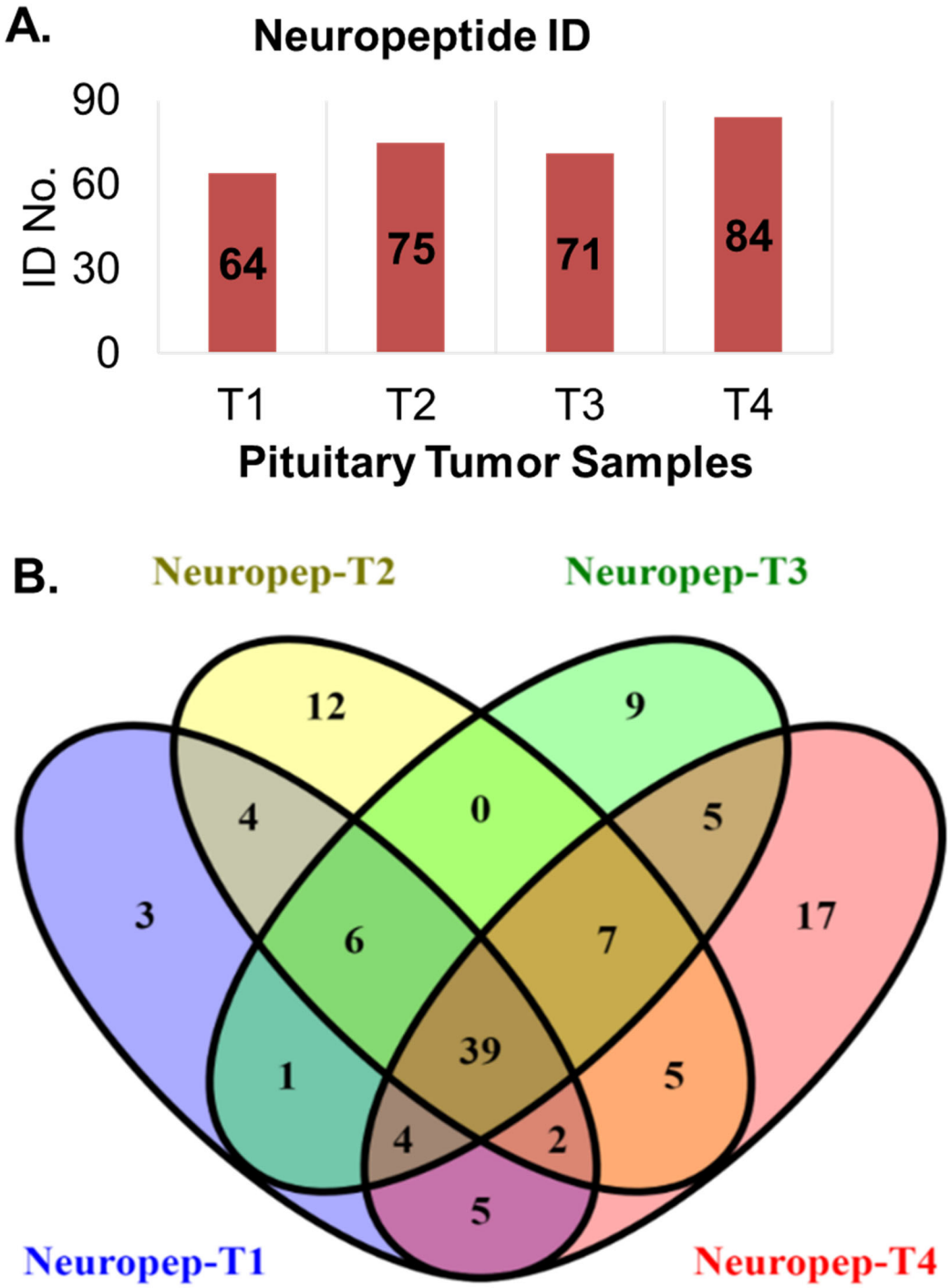


Figure 4. Neuropeptide identification in the human pituitary tumor. The bar chart in **A** shows the identified neuropeptide in four human pituitary tumors; The Venn diagram in **B** shows the overlap neuropeptides between four pituitary tumors.

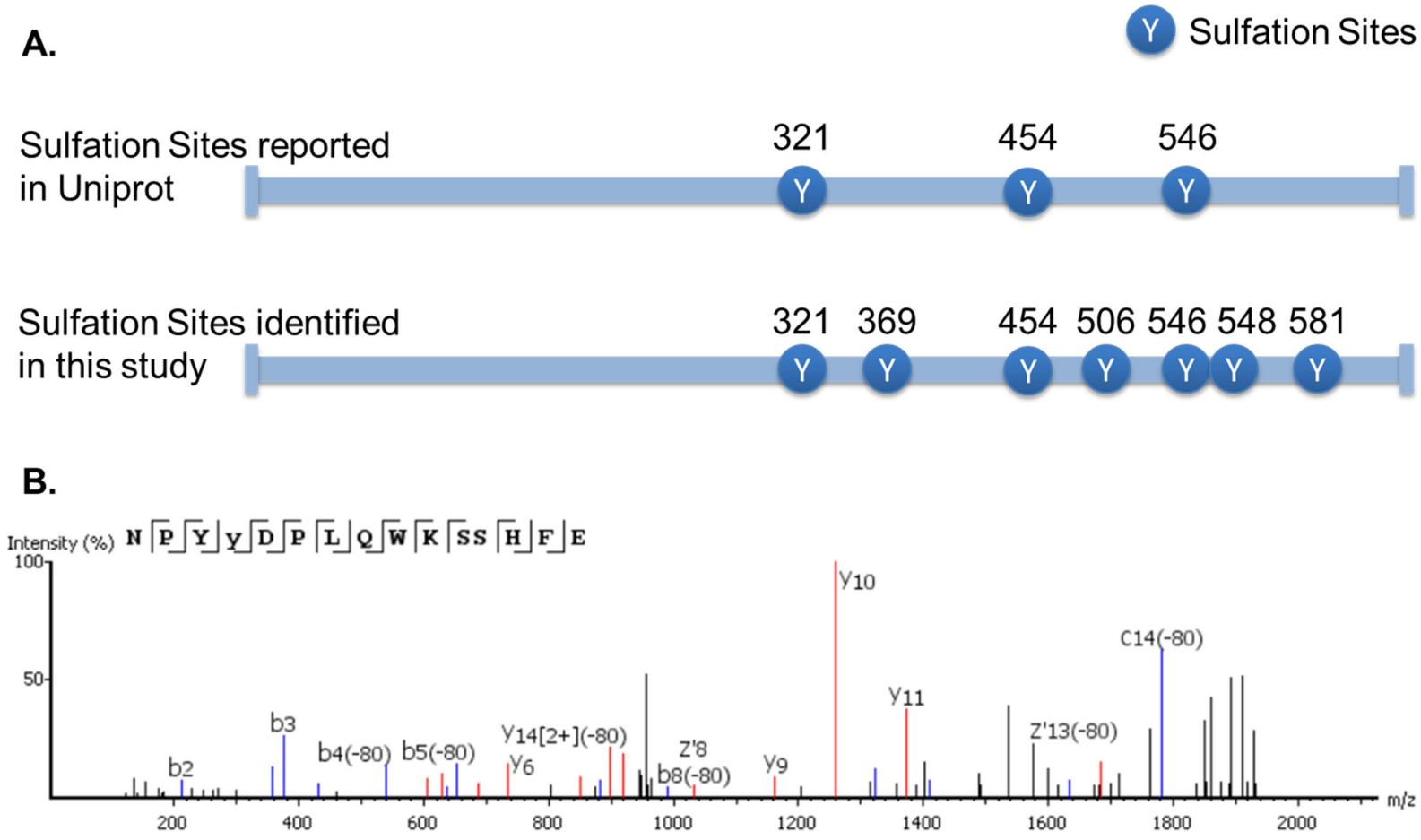


Figure 5. A: Novel sulfation sites identified in neuropeptide secretogranin-1 compare with Uniprot; **B:** Mass Spectra of sulfotyrosine peptide Y.Y(+79.96)DPLQWKSSHFE.R from Secretogranin-1.

SUPPLEMENTARY INFORMATION

Table S1. List of 39 neuropeptides shared by all four human pituitary tumors.

Preprohormone(Uniprot)	Neuropeptide ID	Avg. Mass	-10lgP	Coverage (%)
P05408	NP00012	20887	437.82	92
Q8N6N7	NP00026	9790	94.07	10
P07108	NP00034	9913	231.82	71
P07492	NP00790	2860	190.36	100
P08949	NP00793	3674	86.53	59
Q16568	NP00881	9942	98.13	16
Q6UW01	NP00913	18256	63.07	12
P10645	NP00984	48918	499.74	89
P05060	NP00999	76326	460.95	73
P13521	NP01000	67829	408.18	80
Q8WXD2	NP01002	50994	268.29	31
P01042	NP01097	69897	109.31	6
P22466	NP02025	3157	275.51	100
P22466	NP02026	6672	176.16	73
Q9UBC7	NP02040	6500	90.26	45
P06307	NP02242	10750	91.75	25
P01275	NP02465	8101	62.91	9
P01275	NP02471	3766	53.09	27
O43555	NP02577	10516	96.57	32
P41159	NP02870	16026	75.05	23
Q15848	NP03617	24544	79.52	13
P43490	NP03624	55521	86.57	4
O95158	NP03717	30326	113.72	11
Q8N729	NP03747	3543	55.78	43
P30990	NP03782	14303	54.89	11
Q02818	NP04021	51146	230.87	21
P80303	NP04026	47328	149.29	12
P01210	NP04136	4587	48.09	24
P01189	NP04893	4541	390.69	100
P01189	NP04895	9806	518.05	100
P01189	NP04901	3006	337.2	100
Q9UHG2	NP04921	24034	497.41	96
P08185	NP05299	42639	91.44	3
Q86U17	NP05300	44982	139.05	14

Q8IW75	NP05301	45108	109.81	8
O75830	NP05302	44060	69.89	7
P01019	NP05334	49761	110.33	11
P01236	NP05514	22898	76.93	21
O15240	NP05915	65000	564.76	90

The green highlight neuropeptides are sulfated neuropeptides.

Table S2. List of 27 identified sulfotyrosine peptides.

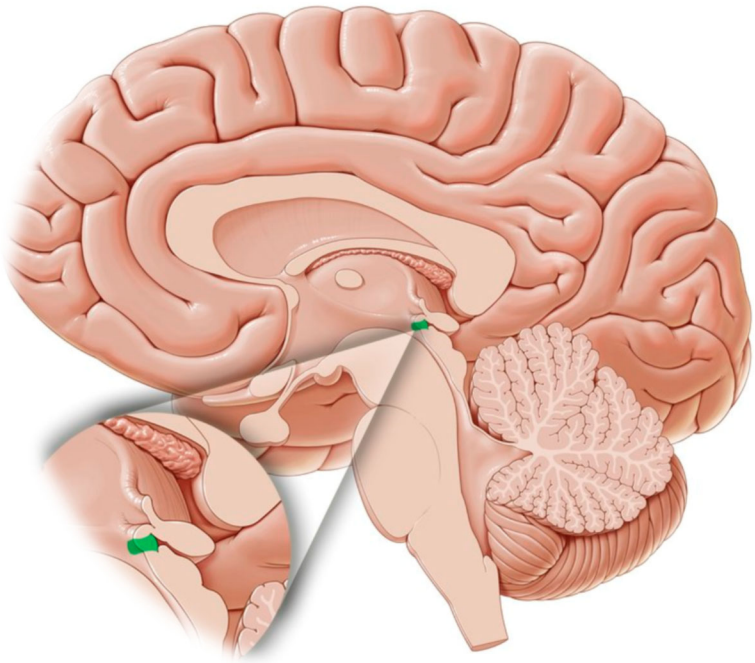
Neuropeptide	Sulfotyrosine Peptide	PTM	Sulfotyrosine Site
Neuroendocrine protein 7B2	D.PEHDY(+79.96)PGLGKW.N*	Sulfation	131
Neuroendocrine protein 7B2	D.PEHDY(+79.96)PGLGKWN.K*	Sulfation	131
Neuroendocrine protein 7B2	F.Q(-17.03)LHQHLFDPEHDY(+79.96)P.G*	Pyro-glu from Q; Sulfation	131
Neuroendocrine protein 7B2	F.Q(-17.03)LHQHLFDPEHDY(+79.96)PGLGKWN.K*	Pyro-glu from Q; Sulfation	131
Neuroendocrine protein 7B2	H.LFDPEHDY(+79.96)PGLGKWN.K*	Sulfation	131
Neuroendocrine protein 7B2	H.Q(-17.03)HLFDPEHDY(+79.96)PGLGKWN.K*	Pyro-glu from Q; Sulfation	131
Neuroendocrine protein 7B2	L.FDPEHDY(+79.96)PGLGKW.N*	Sulfation	131
Secretogranin-1	N.Y(+79.96)DWWEKKP.F*	Sulfation	548
Secretogranin-1	Y.NY(+79.96)DWWEKKP.F*	Sulfation	548
Secretogranin-1	Y.Y(+79.96)DPLQW.K*	Sulfation	506
Secretogranin-1	Y.Y(+79.96)DPLQWKS.S	Sulfation	506
Secretogranin-1	Y.Y(+79.96)DPLQWKSSH.F	Sulfation	506
Secretogranin-1	Y.Y(+79.96)DPLQWKSSHFE.R	Sulfation	506
Secretogranin-1	F.NPYY(+79.96)DPLQWKSSHFE.R	Sulfation	506
Secretogranin-1	E.KNFFPEY(+79.96)N.Y*	Sulfation	546
Secretogranin-1	L.NY(+79.96)GEEGAPGKWQ.Q*	Sulfation	454
Secretogranin-1	E.EEPEY(+79.96)GEEIKGYPGVQAPEDLEWERYR(-.98).G	Sulfation; Amidation	321

Secretogranin-1	R.ASEEEPEY(+79.96)GEEIKGYPGVQAPEDLEWERYR(-.98).G	Sulfation; Amidation	321
Secretogranin-1	R.NY(+79.96)PSLELDKM.A	Sulfation	369
Secretogranin-1	R.Q(-17.03)Y(+79.96)DRVAQLDQLLH.Y*	Pyro-glu from Q; Sulfation	581
Secretogranin-1	R.Q(-17.03)YDRVAQLDQLLHY(+79.96).R	Pyro-glu from Q; Sulfation	581
Secretogranin-2	R.E(-18.01)RMDEEQKLY(+79.96).T*	Pyro-glu from E; Sulfation	196
Lipotropin beta	K.N(+42.01)AIIKNAY(+79.96)KKGE*	Acetylation (N-term); Sulfation	85
ProSAAS	L.RPRPPVY(+79.96).D*	Sulfation	138
ProSAAS	L.RPRPPVY(+79.96)D.D*	Sulfation	138
ProSAAS	L.VPAPVPAAALRPRPPVY(+79.96)DDGPAGPDAEE.A*	Sulfation	138
ProSAAS	R.PRPPVY(+79.96)DDGPAGPDAEE.A*	Sulfation	138

* Peptides with Ascore (PTM localization confidence score) of 1000.

Chapter 5

Neuropeptidomic Study of the Mammalian Subcommissural Organ (SCO) by High-Resolution Mass Spectrometry



Adapted from **Wei P**, Ma F, Ge W, Li L. "Neuropeptidomic Study of the Mammalian Subcommissural Organ (SCO) by High-Resolution Mass Spectrometry." *To be submitted.* **Wei P**, focus on the data analysis and manuscript preparation under the supervision of LL; Ma F conducted the sample preparation and instrument analysis, Ge W provide the SCO sample and provided insights in SCO.

ABSTRACT

The mammalian subcommissural organ (SCO) is a circumventricular organ (CVO) of ependymal (glial) origin. The brain could utilize neuroactive messengers to control the secretory process in SCO. Although the functionality of the SCO is not fully understood, it has been implicated in guiding the development of the brain and spinal cord during embryonic development. Since neuropeptides are neuronal signaling molecules that coordinate the activities of the brain and whole body, to elucidate the functions of the SCO at the molecular level requires knowledge of its neuropeptidome, along with their critical post-translational modifications (PTMs). In this work, liquid chromatography (LC) coupled with Orbitrap tandem mass spectrometry (MS) was used to uncover the neuropeptide profiles of the SCO. With the assistance of bioinformatics, a total of 12 neuropeptides were detected from the SCO. Base on their neuro functions, the crucial role of SCO in embryonic development was supported. Also, novel SCO functions, such as stress response, cold response and dietary excess response, and reproduction, were first identified in SCO. The identification of rare PTMs, such as citrullination, Hexose (NSY), and the discovery of unique O-glycosylated neuropeptide helped better understand the roles of the SCO in different biological processes. Furthermore, novel neuropeptide, DVGSYQEKVDVVLGPIQLQSPSKE, was uniquely identified in SCO. Together, these data represent the first neuropeptidomics study of SCO. Information on the neuropeptide content and PTM modulation can be used as a basic understanding of the roles of SCO in different biological processes and enable a range of follow-up studies.

KEYWORDS:

Subcommissural Organ (SCO), neuropeptidomics, mass spectrometry, post-translational modifications (PTMs), glyco-neuropeptide, preprohormone

INTRODUCTION

The mammalian subcommissural organ (SCO) is a circumventricular organ (CVO) located in the dorsocaudal region of the third ventricle (1). Cells of the subcommissural organ, which are specialized in the secretion of glycoproteins, are arranged into two layers: a superficial layer called the ependyma and an underlying layer called the hypendyma (2). Both hypendymal cells and ependymal cells are secretory in nature, and they secrete Reissner's fiber glycoprotein and many other proteins into the CSF (1, 3). The function of SCO is not known, although studies have shown preliminary evidence to support its role in embryonic development (4). Although there are no neuronal cell bodies within the SCO, it receives neuronal input from different neuropeptide and neurotransmitter systems. These neurochemical messengers may be responsible for changes in the secretory activity of the specialized SCO cells. Therefore, the neural function of SCO could be revealed by the study of the signaling molecules from SCO.

Neuropeptides and peptide hormones are the main classes of signaling molecules in nervous and endocrine systems that regulate physiological processes and behaviors (5). They are classified into different families based on the shared sequence motif. Neuropeptides are usually released into the circulating system after cellular enzymatic processes from inactive precursors and multiple post-translational modifications (PTMs). Low concentration and various PTMs are the main challenges for global neuropeptide characterization (6). The immunohistochemical method used to be applied to the SCO input system for neuropeptide identification (7). Although immunohistochemical is a sensitive method, it is not suitable for the large-scale discovery of neuropeptide study. Mass spectrometry has evolved to be a sensitive and unbiased method for the untargeted neuropeptide characterization (8). Combined with the advanced bioinformatic software, both neuropeptide

structure and PTMs could be confidently elucidated.

PTMs are of great significance in determining or altering peptide structures and biological activities. Glycosylation is one of the most common yet the most challenging PTMs, due to the complexity of attached glycans and the site occupancy on the backbones (9, 10). Among different types of glycosylation, two categories are widely studied, namely, N- and O-linked glycosylation. Specifically, an oligosaccharide attached to the side chain amide nitrogen of asparagine (Asn) is recognized as N-linked glycans with the Asn-X-Ser/Thr sequon, where X could be any amino acid except proline. O-linked glycosylation usually occurs on the hydroxyl group of serine and threonine when the carbohydrate is attached to the oxygen (11). To deal with the high complexity and low abundance of the endogenous glycosylated peptides, MS-based approaches have emerged as powerful tools. Especially with the instrumentation developments in ionization sources, mass analyzers and the fragmentation modes, significant advancements in glycopeptide analysis enable revealing peptide backbone structures, site-specific information, and glycan types with less instrument time (12).

Due to the limitation of the neuropeptide database, as well as the limited knowledge about the SCO neural function, novel neuropeptides may exist in SCO neuropeptidomics study. Therefore, besides the mature neuropeptide database, a custom-built preprohormone database, which including all the proteins with signal peptides, was developed and applied to the SCO novel neuropeptide investigation. Herein, we performed the first neuropeptidomic analysis of the SCO, revealing unique information on the neuropeptide content and PTM modulation of the SCO, which helps to understand better the roles of the SCO in different biological processes.

EXPERIMENTAL SECTION

SCO Isolation

All animal studies conform to the Guide for the Care and Use of Laboratory Animals (National Institutes of Health Publication No. 85 – 23, 1996 revision) and protocols approved by the Institutional Animal Care and Use Committee at the University of Texas Southwestern Medical Center. 12-week old male C57BL/6J mice were used for this study (Jackson Lab, Bar Harbor, MA). Forty mice were euthanized with isoflurane, and the brains were quickly removed after fast decapitation. The entire brain samples were placed on dry ice immediately after isolation. Then SCOs were carefully isolated pooled together as the SCO sample. A brain slice without SCO was used as the control sample. Both SCO and control samples were stored at -80 ℃ until use.

Endogenous Peptide Extraction

The isolated SCO sample and control sample were dissolved in water/methanol/acetic acid (1:90:9, v/v/v) buffer based on the ratio of buffer volume to tissue weight, 10μL/1mg. After homogenizing by Sonic Dismembrator (8 secs on, 15 secs off, three cycles in total, Fisher Scientific Model FB120), the mixture was centrifuged at (20,000 *xg*, 20 min, 4°C). Then clarified supernatant (200μL) was aliquoted and loaded on the pre-rinsed ((i) 0.1M sodium hydroxide, 200 μl (ii) water/methanol/acetonitrile solution (50:30:20, v/v/v), 200 μL, repeat one time (iii) water/methanol/acetic acid (1:90:9, v/v/v) buffer, 200μL, 15,000 *xg*, 5min, 4°C) 0.5ml molecular weight cut off (MWCO) filter (30 kDa, Millipore Amicon Ultra, Burlington, MA). The fluid that went through centrifugal filters during centrifugation (15,000 *xg*, 30 min, 4 °C) was collected as the neuropeptide samples (<30 kDa). The centrifugal filters were then rinsed twice with extraction buffer (200 μL; same centrifugation speed and time), and the flow-through was combined with the

previously collected peptide samples (total volume ~ 600 µL). All the samples were then aliquoted and lyophilized. C18 Zip tips (Millipore Sigma, St. Louis, MO) was used for desalting. After lyophilization, the peptide samples were stored at -80 °C until LC-MS analysis.

LC–Orbitrap MS Analysis

The peptides extracted from the SCO and control samples were analyzed using a nanoLC system (Dionex UltiMate 3000, Thermo Fisher Scientific) coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). The samples were reconstituted in loading solvent of 3% acetonitrile (ACN), 0.1% formic acid (v/v). The analytical column was self-made with an integrated emitter tip and dimensions of 75 µm inner diameter × 17 cm length, caped with 3.0 µm, packed with 1.7 µm, 150 Å, BEH C_{18} material (Waters, Milford, MA). H_2O with 0.1% FA, and ACN with 0.1% FA, were used as solvents A and B, respectively. The flow rate was 0.3 µL/min, and the gradient for the peptide elution range was as follows: 0-16 min, 3% solvent B; 16-20 min, 3-25% B; 20-30 min, 25-45% B; 30-50 min, 45-70% B; 50-56 min, 70-95% B; 56-60 min 95% B; 60-60.5 min, 95-3% B; 60.5-70 min, 3% B. Data acquisition was accomplished in top speed data-dependent mode. Other parameters included: precursor scan automatic gain control (AGC), 1e+6; MS/MS scan AGC, 5e+4; isolation window, 1 m/z; normalized collision energy, 30%; with both higher-energy C-trap dissociation (HCD) and electron-transfer/higher-energy collision dissociation (EThcD) tandem mass spectrometry analysis.

Data Analysis

Raw data were searched against the mouse neuropeptide database combined Neuropep and SwePep (337 entries) with PEAKS 8.5 for neuropeptide identification. A mass tolerance of ±10

ppm was used for precursors, monoisotopic mass tolerance was set to ±0.02 Da for product ions. HCD fragmentation type and EthcD fragmentation were selected individually. An advanced setting that searched against 313 build-in modifications was used for more accurate PTM searching. Parameters for confident neuropeptide identification were Ascore (PTM site confidence) higher than 20, FDR lower than 1%, and the presence of at least one unique peptide. Besides the mature neuropeptide database, a house-made candidate preprohormone database was also used to discover novel neuropeptides, as previously described (13).

Raw data were also searched against 309 mammalian N-glycans and 78 O-glycans modifications with Byonic^{TM} for glyco-neuropeptide study. Our post-processing steps included manual filtering to 1% false discovery rate (FDR) at the peptide spectral match level using the 2D-FDR score. Then removing identifications that had a Byonic score below 150 and retaining glyco PSMs that had |logProb| value above 1 (which is the absolute value of the log base 10 of the protein p-value). A further filtering step was added that only allowed for identifications with a DeltaMod score of 10 or higher to remove all decoy hits (14).

RESULTS AND DISCUSSION

Neuropeptide Identification:

From **Table 1 and Fig. 1A**, a total of 12 neuropeptides were detected in the SCO, with nine of them exhibiting 100% sequence coverage. Notably, neuropeptide manserin, which could regulate the neuroendocrine system, was identified in mice for the first time (15). And neuropeptide somatostatin, which has been identified as an SCO input neuropeptide with the immunohistochemical method, has also been identified with our MS method. Most of the identified

neuropeptides were secreted neuropeptides **(Fig 1B)**, while Thymosin beta 4 (Tβ4) and Thymosin beta 10 (Tβ10), belonging to the cytoskeleton, and Acyl-CoA-binding protein (ACBP) located at the endoplasmic reticulum and Golgi apparatus. Their preprohormone belongs to the granin (chromogranin/secretogranin) family, thymosin beta family, MCH family, opioid family, somatostatin family, and other families. While in the control sample, 62 neuropeptides were identified, and 34 of them were exhibiting 100% sequence coverage (**Fig 1A**). All the 12 neuropeptides identified in SCO were also identified in the control samples.

PTM Identification:

To identify more PTM modified neuropeptides in the SCO, an advanced setting that searched raw data against 313 built-in modifications was used. Besides the common peptide PTMs, acetylation (N-terminal), amidation, and oxidation (M), rare PTMs, citrullination, and hexose (NSY) were also identified from the SCO neuropeptides. For the two rare PTMs, citrullination is specifically modified on the arginine (site 39) of neuropeptide Tβ10, and hexose is modified on the serine (site 1 and site 15) of neuropeptide Tβ4. Mass spectra of two modifications could be found in **Fig 2.** The same modified neuropeptides, Tβ10 and Tβ4 were also found in the control sample.

After searching raw data against both 309 mammalian N-Glycan and 78 mammalian O-glycan modification lists with Byonics, nine O-glycosylated peptides (**Table 2**) were found belonging to the SCO neuropeptides. One O-glycosylated peptide from neuropeptide Tβ10 and eight from Tβ4. When comparing with the O-glycosylated peptides identified from the control sample, four O-glycosylated peptides from neuropeptide Tβ4 were uniquely detected in SCO (**Table 2**)**.**

Novel Neuropeptide Identification:

Besides the mature neuropeptides database, the preprohormone database that includes all the mouse proteins with signal peptide sequences was also used for the novel neuropeptide exploration. From the searching result of the preprohormone database, 15 preprohormones were identified from the SCO. When comparing with the control sample, 7 of them were overlap with the control sample, while 8 of them were uniquely identified in SCO (**Fig. 3**). From the Gene Ontology (GO) analysis shown in **Fig. 4B**, extracellular region, extracellular space, and extracellular exosome are the top cell component of the preprohormones from SCO, which are consistent with the cell component with neuropeptides.

After entering those 15 preprohormones identified in SCO into NeuroPred for the novel neuropeptide investigation, five peptides (**Table 3**) were identified from the preprohormone. According to the cleavage probability prediction, Arginine at protein site 2002 shows a cleavage probability of 0.9906, which is much higher than the neuropeptide cleavage probability requirement value of 0.8. Based on the five detected peptides and predicted cleavage sites, the sequence of novel neuropeptide is DVGSYQEKVDVVLGPIQLQSPSKE, which has been identified as a bioactive peptide in rat plasma after hemorrhagic shock. Notably, those five peptides were only identified in SCO when compared with the control sample, which shows the unique function of this candidate neuropeptide in SCO.

Discussion:

From neuropeptide content to infer SCO neuronal function

The SCO is a secretory gland that receives some innervation from outside neurons. By studying

those input neuropeptides in SCO, which brings the neuroactive messengers to control the secretory process, we could know more about the neural functions of SCO. Both Tβ10 and Tβ4 belong to the beta-thymosin family and exhibit diverse physiological functions such as tissue development and regeneration, anti-inflammatory effects, and induction of insulin secretion (16–20). Especially Tβ4 could bind to actin and promote cell migration, including the mobilization, migration, and differentiation of stem/progenitor cells, which form new blood vessels and regenerate the tissue. Previous study has suggested that Tβ4 is a crucial factor in cardiac development, growth, disease, epicardial integrity, and blood vessel formation during embryonic development (21). Besides peptides Tβ4 and Tβ10, the other two neuropeptides, cerebellin-4 and ACBP, were also identified related to embryonic development. Neuropeptide cerebellin-4 plays a vital role in the formation and maintenance of inhibitory GABAergic synapses, and neuropeptide ACBP modulates the action of the GABA receptor (22). GABA receptor has been identified in the SCO in the previous study with the immunochemistry method (7). Although GABA is the chief inhibitory neurotransmitter in the developmentally mature mammalian central nervous system, its actions were thought to be primarily excitatory in the developing brain (23). Therefore, GABA is a crucial neurotransmitter during embryonic development, and its function was regulated by the two neuropeptides, cerebellin-4 (24) and ACBP (22, 25), identified in the SCO. Although the functionality of the SCO is not fully understood, it has been implicated in guiding the development of the brain and spinal cord. These identified neuropeptides further support SCO's function during embryonic development.

Besides potential role during embryonic development, the SCO may also be involved in stress response. Three identified neuropeptides, Met-enkephalin-Arg-Phe (MERF),

PENK_MOUSE_198-209, and Somatostatin-28 have been implicated in various biological processes related to stress response (26–29). Among them, MERF and PENK_MOUSE_198-209 were from the opioid neuropeptide precursor family, which is well known for their roles in pain perception and responses to stress. Somatostatin-28 could markedly influence several components of the stress response by activation of somatostatin receptor receptors. Besides the stress response, the three neuropeptides from chromogranin/secretogranin protein family, Manserin, Secretogranin-3, and SCG1_MOUSE_588-597 may indicate the role of SCO in regulated secretion (30). Neuropeptide Big SAAS is a serine-type endopeptidase inhibitor and response to cold and dietary excess biological process or relevance to modulation of the auditory and vestibular system. Finally, Neuropeptide- glutamic acid isoleucine (NEI) is a peptide related to reproduction (31, 32).

For the rare PTMs, citrullination could increase the hydrophobicity of the protein, which can lead to changes in protein folding, affecting the structure and function. Therefore, it has a vital role in the insulation of neurons and the plasticity of the central nervous system (33). The mechanism and biological function of hexosylation remain unclear, might be a kind of glycation. Notably, the four O-glycosylated peptides from neuropeptide Tβ4 were uniquely expressed in the SCO when compared with the control sample. Unlike other CVOs, the blood-brain barrier (BBB) is present in the SCO. The neuropeptides after glycosylation could be more easily and effectively penetrate the BBB, which shows the essential role of neuropeptide Tβ4 on SCO.

CONCLUSIONS

SCO, a CVO of ependymal (glial) origin, utilizes neuroactive messengers, possibly controlling the secretory process. This was the first MS-based neuropeptidomic study of SCO, revealing twelve

neuropeptides that may suggest potential neuronal function of the SCO. The crucial role of SCO in embryonic development was supported. Also, novel SCO functions, such as stress response, cold and dietary excess response, and reproduction, were implicated in the SCO for the first time. PTM modulation, including glycosylation, helped to understand better the roles of the SCO in different biological processes. Furthermore, one candidate neuropeptide was found in the SCO, whose neural function will be explored and verified in the follow-up study.

ACKNOWLEDGMENTS

This research was supported in part by NIH R01 DK071801 (to L.L.). The Orbitrap Fusion Lumos Tribrid instrument was purchased through the support of an NIH shared instrument grant NIH-NCRR S10RR029531 (L.L.) and Office of the Vice-Chancellor for Research and Graduate Education at the University of Wisconsin-Madison.

REFERENCES

1. Guerra MM, González C, Caprile T, Jara M, Vío K, Muñoz RI, Rodríguez S, Rodríguez EM. 2015. Understanding how the subcommissural organ and other periventricular secretory structures contribute via the cerebrospinal fluid to neurogenesis. Front Cell Neurosci. Frontiers Media S.A.
2. Rodríguez EM, Rodríguez S, Hein S. 1998. The subcommissural organ. Microsc Res Tech 41:98–123.
3. Vio K, Rodríguez S, Yulis CR, Oliver C, Rodríguez EM. 2008. The subcommissural organ of the rat secretes Reissner's fiber glycoproteins and CSF-soluble proteins reaching the internal and external CSF compartments. Cerebrospinal Fluid Res 5:3.

4. Hoyo-Becerra C, López-Avalos MD, Alcaide-Gavilán M, Gómez-Roldán MC, Pérez J, Fernández-Llebrez P, Grondona JM. 2005. Reissner's fiber formation depends on developmentally regulated factors extrinsic to the subcommissural organ. Cell Tissue Res 321:429–441.

5. Wei P, Keller C, Li L. 2020. Neuropeptides in gut-brain axis and their influence on host immunity and stress. Comput Struct Biotechnol J. Elsevier B.V.

6. Li L, Sweedler J V. 2008. Peptides in the Brain: Mass Spectrometry–Based Measurement Approaches and Challenges. Annu Rev Anal Chem 1:451–483.

7. Nürnberger F, Schöniger S. 2001. Presence and functional significance of neuropeptide and neurotransmitter receptors in subcommissural organ cells. Microsc Res Tech 52:534–540.

8. Yang N, Anapindi KDB, Rubakhin SS, Wei P, Yu Q, Li L, Kenny PJ, Sweedler J V. 2018. Neuropeptidomics of the Rat Habenular Nuclei. J Proteome Res 17:1463–1473.

9. Polt R, Dhanasekaran M, Keyari CM. 2005. Glycosylated neuropeptides: A new vista for neuropsychopharmacology? Med Res Rev 25:557–585.

10. Cao Q, Wang Y, Chen B, Ma F, Hao L, Li G, Ouyang C, Li L. 2019. Visualization and Identification of Neurotransmitters in Crustacean Brain via Multifaceted Mass Spectrometric Approaches. ACS Chem Neurosci 10:32.

11. Liu Y, Cao Q, Li L. 2019. Isolation and characterization of glycosylated neuropeptides, p. 147–202. *In* Methods in Enzymology. Academic Press Inc.

12. Yu Q, Wang B, Chen Z, Urabe G, Glover MS, Shi X, Guo LW, Kent KC, Li L. 2017. Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. J Am Soc Mass Spectrom 28:1751–1764.

13. Pingli W, Qing Y, Haidan S, Fengfei M, P Vaishali B, Wei S, Zhi Z, Chun Z, Lingjun L.

Development of an Optimal Mass Spectrometry Approach for Comprehensive Neuropeptide Characterization and its Application to Analysis of Human Pituitary Tumor.

14. Riley NM, Hebert AS, Westphall MS, Coon JJ. 2019. Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. Nat Commun 10:1–13.

15. Narita; AYIMMN. 2004. Manserin, a novel peptide from secretogranin Ii in the neuroendocrine system. Neuroreport 15:1755–1759.

16. Zhang X, Ren D, Guo L, Wang L, Wu S, Lin C, Ye L, Zhu J, Li J, Song L, Lin H, He Z. 2017. Thymosin beta 10 is a key regulator of tumorigenesis and metastasis and a novel serum marker in breast cancer. Breast Cancer Res 19:15.

17. Nemolato S, Ekstrom J, Cabras T, Gerosa C, Fanni D, Di Felice E, Locci A, Messana I, Castagnola M, Faa G. 2013. Immunoreactivity for thymosin beta 4 and thymosin beta 10 in the adult rat oro-gastrointestinal tract. Eur J Histochem 57:106–111.

18. Hall AK, Hempstead J, Morgan JI. 1990. Thymosin β10 levels in developing human brain and its regulation by retinoic acid in the HTB-10 neuroblastoma. Mol Brain Res 8:129–135.

19. Gao X, Liang H, Hou F, Zhang Z, Nuo M, Guo X, Liu D. 2015. Thymosin beta-4 induces mouse hair growth. PLoS One 10.

20. Gealekman O, Gurav K, Chouinard M, Straubhaar J, Thompson M, Malkani S, Hartigan C, Corvera S. 2014. Control of adipose tissue expandability in response to high fat diet by the insulin-like growth factor-binding protein-4. J Biol Chem 289:18327–18338.

21. Banerjee I, Zhang J, Moore-Morris T, Lange S, Shen T, Dalton ND, Gu Y, Peterson KL, Evans SM, Chen J. 2012. Thymosin beta 4 is dispensable for murine cardiac development and function. Circ Res 110:456–464.

22. Burton M, Rose TM, Færgeman NJ, Knudsen J. 2005. Evolution of the acyl-CoA binding

protein (ACBP). Biochem J 392:299–307.

23. Li K, Xu E. 2008. The role and the mechanism of γ-aminobutyric acid during central nervous system development. Neurosci Bull. Springer.

24. Chacón PJ, del Marco Á, Arévalo Á, Domínguez-Giménez P, García-Segura LM, Rodríguez-Tébar A. 2015. Cerebellin 4, a synaptic protein, enhances inhibitory activity and resistance of neurons to amyloid-β toxicity. Neurobiol Aging 36:1057–1071.

25. Knudsen J, Mandrup S, Rasmussen JT, Andreasen PH, Poulsen F, Kristiansen K. 1993. The function of acyl-CoA-binding protein (ACBP)/Diazepam binding inhibitor (DBI). Mol Cell Biochem 123:129–138.

26. Rossier J, Audigier Y, Ling N, Cros J, Udenfriend S. 1980. Met-enkephalin-Arg6-Phe7, present in high amounts in brain of rat, cattle and man, is an opioid agonist. Nature 288:88–90.

27. Vats ID, Chaudhary S, Karar J, Nath M, Pasha Q, Pasha S. 2009. Endogenous peptide: Met-enkephalin-Arg-Phe, differently regulate expression of opioid receptors on chronic treatment. Neuropeptides 43:355–362.

28. Stengel A, Rivier J, Tache Y. 2012. Central Actions of Somatostatin-28 and Oligosomatostatin Agonists to Prevent Components of the Endocrine, Autonomic and Visceral Responses to Stress Through Interaction with Different Somatostatin Receptor Subtypes. Curr Pharm Des 19:98–105.

29. D'Alessio DA, Sieber C, Beglinger C, Ensinck JW. 1989. A physiologic role for somatostatin 28 as a regulator of insulin secretion. J Clin Invest 84:857–862.

30. Secretogranin II binds to secretogranin III and forms secretory granules with orexin, neuropeptide Y, and POMC in: Journal of Endocrinology Volume 202 Issue 1 (2009).

31. Fujimoto M, Fukuda S, Sakamoto H, Takata J, Sawamura S. 2017. Neuropeptide glutamic acid-isoleucine (NEI)-induced paradoxical sleep in rats. Peptides 87:28–33.

32. Ester M. 2013. Influence of Neuropeptide – Glutamic Acid-Isoleucine (NEI) on LH RegulationGonadotropin. InTech.

33. Baka Z, György B, Géher P, Buzás EI, Falus A, Nagy G. 2012. Citrullination under physiological and pathological conditions. Jt Bone Spine.

TABLES AND FIGURES

Table 1. A list of identified neuropeptides in the SCO

Precursor Family	Neuropeptide	Coverage (%)	Subcellular Location	Neurofunction
thymosin beta	Tβ4	100	Cytoskeleton	embryonic development
	Tβ10	100	Cytoskeleton	embryonic development
CBLN	Cerebellin-4	7	Secreted	inhibitory GABAergic synapses
ACBP	ACBP	29	Endoplasmic reticulum, Golgi apparatus	GABA receptor
opioid	MERF	100	Secreted	responses to stress.
	PENK_MOUSE_198-209	100	Secreted	responses to stress.
somatostatin	Somatostatin-28	100	Secreted	responses to stress.
ProSAAS	Big_SAAS	100	Secreted	response to cold and dietary excess
granin	Manserin	100	Secreted	regulated secretory proteins
	Secretogranin-3	3	Secreted	
	SCG1_MOUSE_588-597	100	Secreted	
MCH	NEI	100	Secreted	reproduction.

Table 2. O-glycosylated peptides identified from the SCO. The green highlighted ones are only identified in SCO.

Neuropeptide	Peptide	Glycan Modification	Byonic score	Delta Mod	|Log Prob|
Tβ4	M.AEIEKFDKS[+892.317]KLKKTETQEKNPLPSKETIEQEKQAGES.-	HexNAc(2)Hex(3)	403.70	21.19	3.69
	-.S[+203.079]DKPDMAEIEKFDKSKLKKT[+162.053]ETQEKNPLPSKETIEQEKQAGES.-	HexNAc(1)	381.39	34.15	5.35
	E.IEKFDKS[+933.344]KLKKTETQEKNPLPSKETIEQEKQAGES.-	HexNAc(3)Hex(2)	365.30	143.08	4.41
	D.[+42.011]KPDMAEIEKFDKS[+203.079]KLKK.T	HexNAc(1)	338.22	338.22	3.82
	M.AEIEKFDKS[+892.317]KLKKTETQEKNPLPSKETIEQEKQAGES.-	HexNAc(2)Hex(3)	333.60	27.42	3.58
	D.KPDM[+15.995]AEIEKFDKS[+406.159]KLKKTETQEKNPLPSKETIEQEKQAGES.-	HexNAc(2)	303.78	134.15	3.60
	D.[+42.011]KPDMAEIEKFDKS[+203.079]KLK.K	HexNAc(1)	300.55	300.55	4.10
	M.AEIEKFDKS[+1038.375]KLKKTETQEKNPLPSKET[+162.053]IEQEKQAGES.-	HexNAc(2)Hex(3)Fuc(1)	198.82	25.55	2.76
Tβ10	G.EIAS[+162.053]FDKAKLKKT[+755.296]ETQEKNTLPT[+162.053]KETIEQEKRSEIS.-	HexNAc(3)Fuc(1)	281.29	18.72	3.10

Table 3. Peptides identified from candidate neuropeptide DVGSYQEKVDVVLGPIQLQSPSKE.

Neuropeptide	Cleavage Site	Cleavage Probability	Peptide	Mass	ppm	Length	Start	End	Unique
Deleted in malignant brain tumors 1 protein	R2002	0.9906	R.DVGSYQEKV.D	1023.4872	0	9	2003	2011	Y
			R.DVGSYQEKVD.V	1138.5142	0.7	10	2003	2012	Y
			R.DVGSYQEKVDV.V	1237.5826	0.1	11	2003	2013	Y
			R.DVGSYQEKVDVV.L	1336.6510	0.1	12	2003	2014	Y
			R.DVGSYQEKVDVVLGPIQ.L	1844.9519	1.2	17	2003	2019	Y

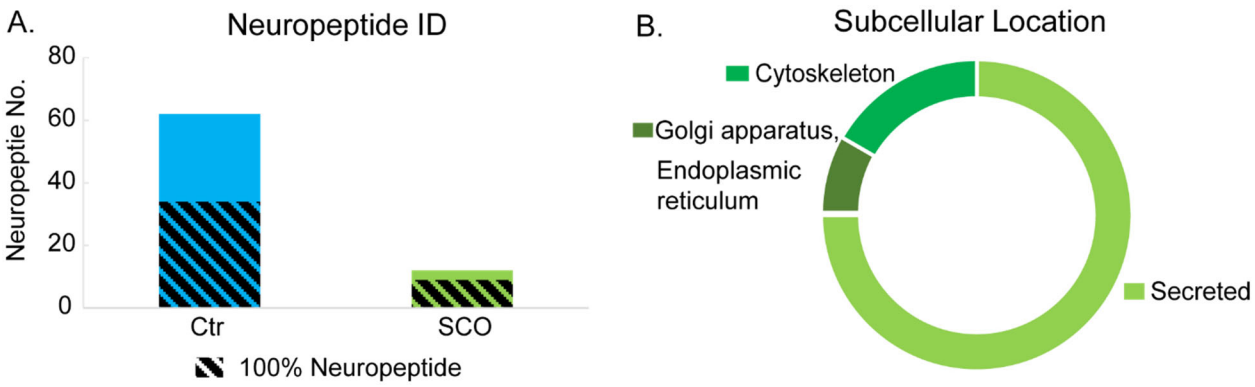
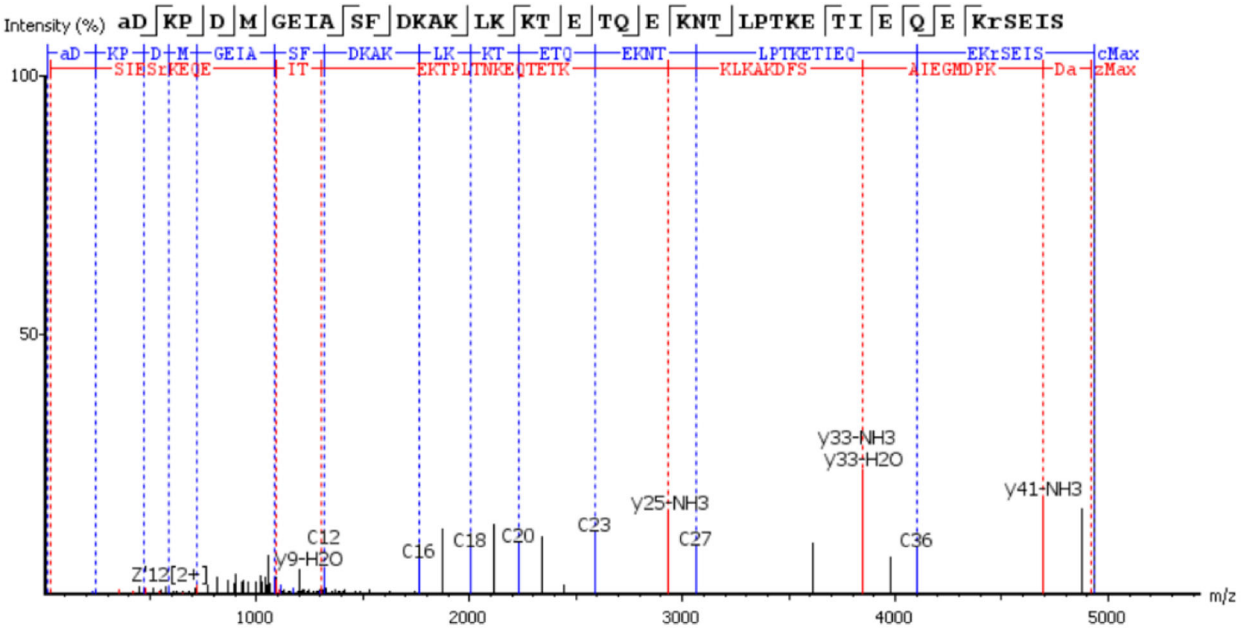


Figure 1. Neuropeptide identification result. A: Neuropeptide identification in the SCO compared with the control sample. **B:** Subcellular location of SCO neuropeptides.

A. Citrullinated Neuropeptide Thymosin beta 10



B. Hexdose Neuropeptide Thymosin beta 4

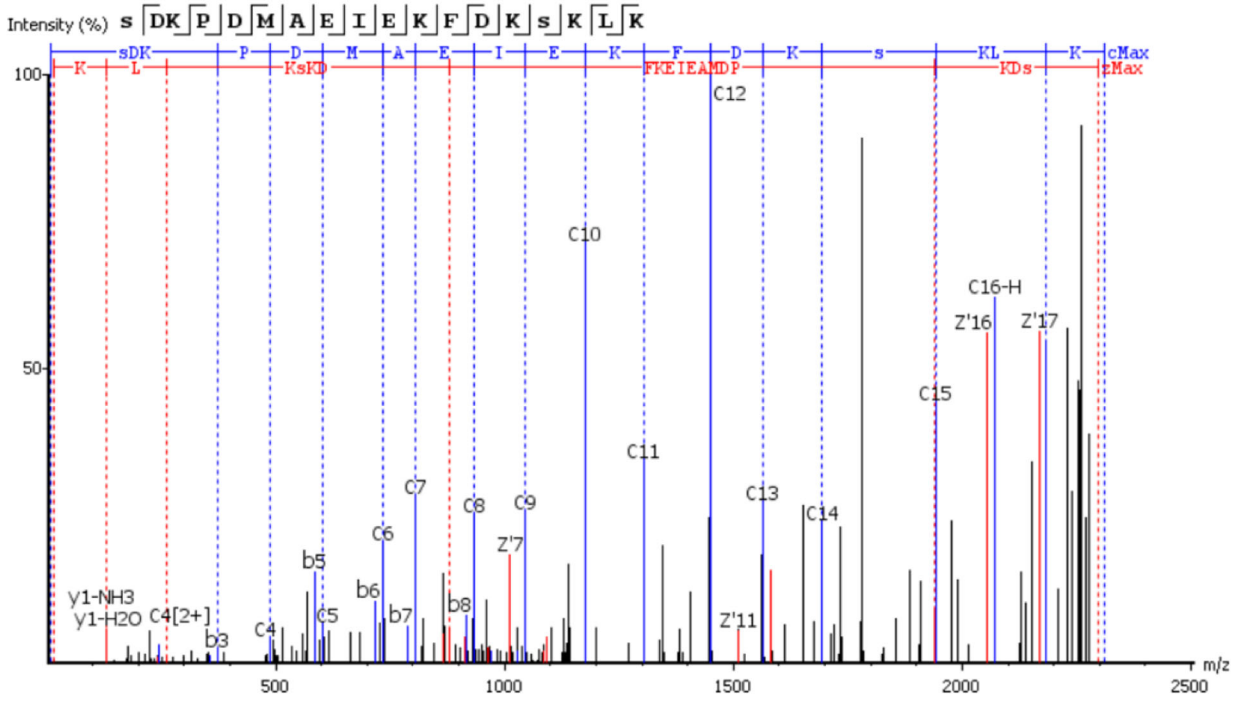


Figure 2. MS Spectra of rare PTM A: Example spectrum of citrullination neuropeptide thymosin beta 10. **B:** Example spectra of Hexdose neuropeptide thymosin beta 4.

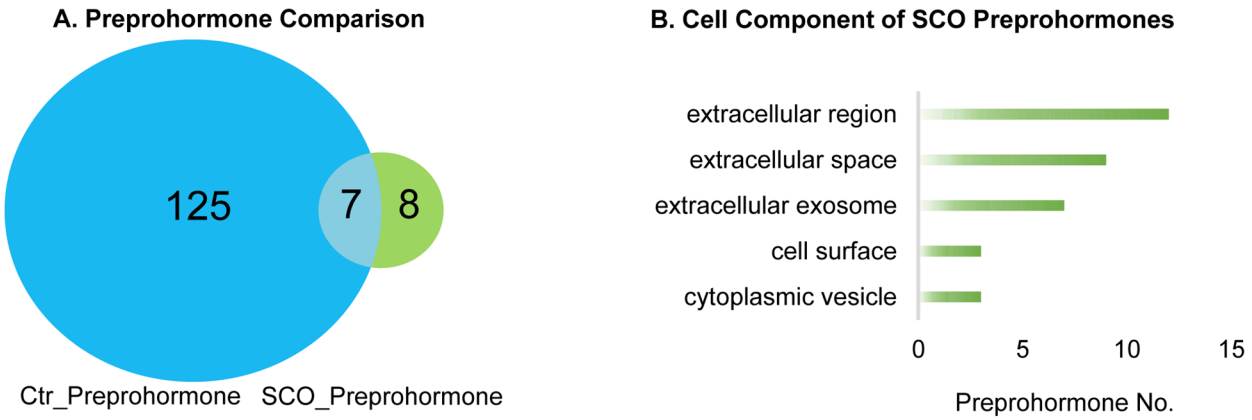
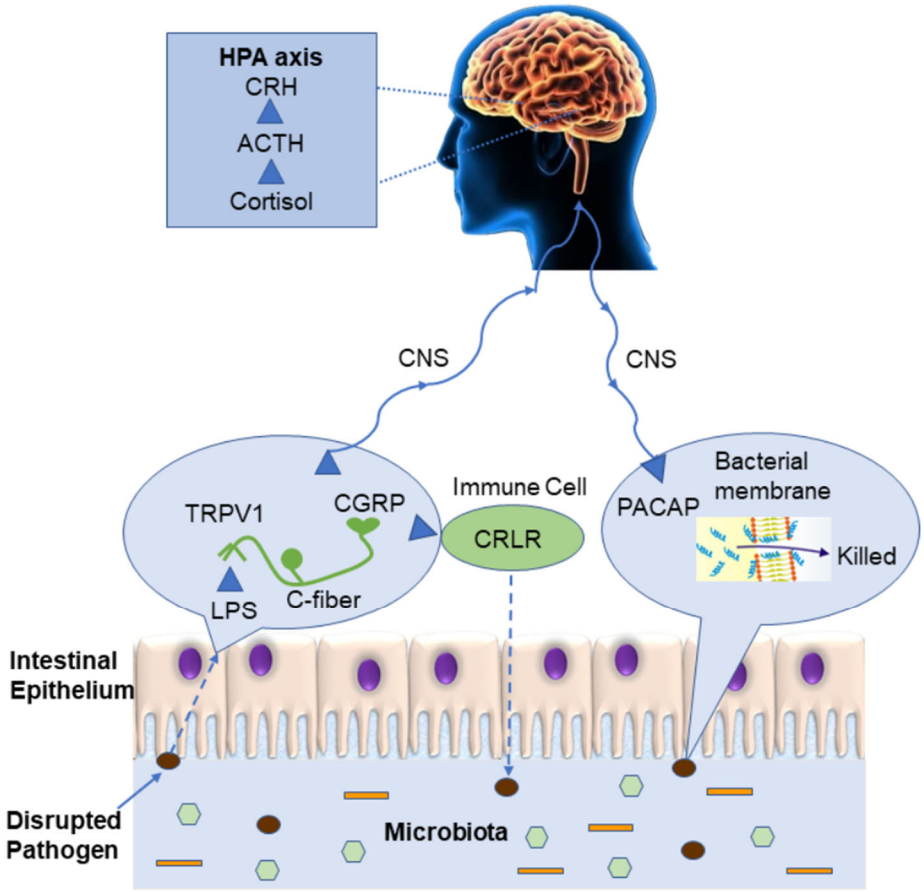


Figure 3. Preprohormone identification result. A: Preprohormones identification in SCO compared with the control sample. **B:** Cell component analysis of SCO preprohormones (Top 5).

Chapter 6

Neuropeptides in Gut-Brain Axis and Their Influence on Host Immunity and Stress



Adapted from **Wei P**, Keller C, Li L. 2020. "Neuropeptides in gut-brain axis and their influence on host immunity and stress. " *Comput Struct Biotechnol Journal.* J 18:843–851. **Wei P** designed and conducted this review under the supervision of LL, Keller C provided review on the stres part.

Abstract

In recent decades, neuropeptides have been found to play a major role in communication along the gut-brain axis. Various neuropeptides are expressed in the central and peripheral nervous systems, where they facilitate the crosstalk between the nervous systems and other major body systems. In addition to being critical to communication from the brain in the nervous systems, neuropeptides actively regulate immune functions in the gut in both direct and indirect ways, allowing for communication between the immune and nervous systems. In this mini review, we discuss the role of several neuropeptides, including calcitonin gene-related peptide (CGRP), pituitary adenylate cyclase-activating polypeptide (PACAP), corticotropin-releasing hormone (CRH) and phoenixin (PNX), in the gut-brain axis and summarize their functions in immunity and stress. We choose these neuropeptides to highlight the diversity of peptide communication in the gut-brain axis.

Keywords:

Neuropeptide, gut-brain axis, immunity, antimicrobial peptides, stress, hypothalamic-pituitary-adrenal axis.

Abbreviations:

NPY, neuropeptide Y; SP, substance P; α-MSH, α-melanocyte-stimulating hormone; VIP, vasoactive intestinal peptide; LPS, lipopolysaccharides; TRPV1, transient receptor potential vanilloid receptor-1; CGRP, calcitonin gene-related peptide; CNS, central nervous system; CRLR, calcitonin receptor like receptor; RAMP1, receptor activity-modifying protein1; PACAP, pituitary adenylate cyclase-activating polypeptide; CRH, corticotropin-releasing hormone;

ACTH, adrenocorticotrophic hormone; HPA axis, hypothalamic-pituitary-adrenal axis; PNX, phoenixin.

1. Introduction

Trillions of bacteria colonize the human gut and the gut microbiota is essential for human health [1]–[4]. While commensal bacteria reside in the host, providing key benefits, including protection against invasion, opportunistic or virulent bacteria are eliminated from the gut by the local innate immune system [5]. Increasing evidence points to appropriate gut microbiota not only making critical contribution to the immune system, but also having a profound impact on brain function [6]–[8]. When virulent pathogens invade, metabolic products such as lipopolysaccharides (LPS) can directly affect the function of enteric neurons, spinal sensory neurons and the vagus nerve through activation of Toll-like receptors or translocation and release of neuropeptides and hormones [9]. Gut permeability is perhaps the most important factor in initiating microbial interaction with the rest of the body [10].

The gut is the largest immune competent organ in our body, and contains close to 100 million neurons [11], [12]. Most of the nerve structures belong to the enteric nervous system (ENS), which regulates gut functions autonomously. The remaining extrinsic nerves connect the central nervous system with the gut and belong either to the afferent gut-brain or the efferent brain-gut axis [13], [14]. Bidirectional communication between the brain and the gut has long been recognized [7], [8]. To summarize the concept of the bidirectional deep interaction between central and peripheral nervous systems and the immune system, one can note that most of the neuropeptides, produced by neurons during immune response versus infectious agents or malignant cells, have

neuroendocrine-like activity that can influence both brain and gut functions. The neuropeptides calcitonin gene-related peptide (CGRP) and pituitary adenylate cyclase-activating polypeptide (PACAP) have been studied within this framework.

Neuropeptides are important due to their ability to regulate a range of diverse biological activities. As neurotransmitters, neuropeptides are components of the autonomic nervous system and act locally at peripheral sites; as neuromodulators, neuropeptides could act on central regulatory centers; and as neurohormones and hormones, neuropeptides could reach the immune system, peripheral vessels, organs and glands through the circulatory system [18]. It is reported that neuropeptides show close relationships and deep interactions between the neuropeptidergic and immunological systems within the host immune homeostasis. Neuropeptides, such as substance P (SP) [16], vasoactive intestinal peptide (VIP) [15], and α-melanocyte-stimulating hormone (α-MSH), released by nervous fibers in the intestine, exhibit a variety of proinflammatory or anti-inflammatory effects that are required for the modulation of innate and adaptive immune response [19], [20].

The influence of the gut microbiota on several aspects of central nervous system (CNS) function is increasingly supported by a growing body of experimental data [21]–[23]. The mechanism of this influence is complex and involves multiple direct and indirect pathways. The direct link between the microbiota and hypothalamic-pituitary-adrenal (HPA) axis association with stress shows the significant role of microbiota in CNS function [24]–[27]. While inflammatory diseases, such as inflammatory bowel disease (IBD), could lead to mental disorder, stress could also induce inflammation through increasing intestinal permeability [28]. In this article, we will review the

current evidence in the literature that points towards roles of neuropeptides in gut-brain axis and their influence on host immune system and psychiatric disorders. We will also review the possible mechanisms through which gut microbiota might be involved in the pathogenesis of these disorders with example neuropeptides. While there are numerous neuropeptides involved in communication along the gut-brain axis, many of which are mentioned in this mini review, a couple select neuropeptides are chosen for in depth analysis to demonstrate the diversity of gut-brain axis communications, as a full description of all gut-brain axis neuropeptides is outside the scope of this mini-review.

2. Role of Gut-Brain Axis in Immune Response

Starting from the study on capsaicin activity, the role of neuropeptides in the connection between the neuroendocrine and immune system has been receiving increasing attention [29], [30]. Neuropeptides from the gut-brain axis have two key roles on the immune system: enhance innate host defense and direct antimicrobial function [31], [32]. In initiating microbial interactions with the intestine, metabolic products such as lipopolysaccharide (LPS) created by pathogenic microorganisms can increase the gut permeability and alter the activity of the ENS and CNS [33]. The intestinal barrier, which could act against this invasion, consists of multiple layers that includes gut flora and external mucus layer, epithelial layer, and lamina propria [34]. Immune cells such as lymphocytes, macrophages, plasma cells, antigen presenting cells, and mast cells are mainly concentrated on the epithelial layer and lamina propria [33].

Many neurotransmitters and neuropeptides bind directly to their receptors expressed in human T cells (also termed T lymphocytes), and also in various other immune cells, including B cells,

dendritic cells, macrophages, and microglia, and subsequently induce various very potent immune effects [35], [36]. Recent studies show that neuropeptides such as neuropeptide Y, Somatostatin, GnRH-I, GnRH-II and CGRP could bind to their receptors in normal peripheral human T cells and trigger or elevate significantly a kaleidoscope of T cell functions and features crucial for health-keeping and disease-fighting tasks [37], [38].

Conventionally neuropeptides are considered as signaling molecules but have recently been shown to be pleiotropic molecules that are integral components of the nervous and immune system [39]. Common characteristics between some neuropeptides and antimicrobial peptides (AMPs), such as shared signal sequence, similarities in size, cationic charge or amphipathic design, suggest that neuropeptides might also serve an additional function in antimicrobial immunity [40]. **Table1** shows some significant neuropeptides that have been proven to have antimicrobial activity, such as substance P (SP), neuropeptide Y (NPY), and vasoactive intestinal peptide (VIP).

The negatively charged bacterial membranes make it a target of antimicrobial neuropeptides compared to the membranes of plants and animals, which have no net charge. After binding to the bacterial membrane, antimicrobial neuropeptides could displace membrane lipids to depolarize the normally energized bacterial membrane, causing fatal problems for the bacteria by altering the membrane structure to create physical holes that cause cellular contents to leak out, or even by entering into the interior of the target bacterial cell to activate deadly processes, such as induction of hydrolases that degrade the cell wall. While conventional antibiotics are easier for the bacteria to become drug-resistant, it is almost impossible for bacteria to become resistant against a neuropeptide antibiotic as its target is the bacterial membrane, the principle cell structure of

bacteria [41]. Since the increased emergence of multi-resistant human pathogenic bacteria have become a worldwide problem, the antimicrobial neuropeptide opens new avenues for future therapeutic application of antibiotic-resistant infection.

2.1 Neuroimmune Connector, Calcitonin Gene-related Peptide (CGRP)

Neuropeptide calcitonin gene-related peptide (CGRP) is a member of the calcitonin peptide family that is expressed in both peripheral and central neurons. It is released from afferent fibers at the site of stimulation when gut environment is changing [42]. As shown in **Fig. 1**, once infection stimulates transient receptor potential vanilloid receptor-1 (TRPV1) on the surface of C fibers, CGRP is secreted from sensory nerves, which are distributed in C fibers and upstream. After release, CGRP actives host defense and immune response at different sites by binding to its receptor, calcitonin receptor like receptor (CRLR), and its receptor activity-modifying protein1 (RAMP1) found on T and B lymphocytes, macrophages, mast cells, and dendritic cells among others [43], [44]. Since CGRP is released at the site of stimulation, regulating innate immune activation and mediating information flow to the rest of the nervous system, it is exemplified as a neuroimmune connector [42].

CGRP has a regulatory effect on both dendritic cell and T-cell functions, down-regulating pro-inflammatory cytokine tumor necrosis factor α (TNF- α) and promoting anti-inflammatory cytokine, interleukin 10 (IL-10) [45]. Dendritic cells are known to express the CGRP receptor CalcR and respond to CGRP in an anti-inflammatory manner [35]. Recently, a study shows that CGRP treatment significantly reduced TNF- α release while upregulating IL-10 release in LPS-activated dendritic cells. This data suggest that TRPV1 activation of dendritic cells plays a role in

their homeostasis and regulation through the release of CGRP [46]. The role of CGRP might also be different in the acute and later stages of inflammation. In acute stages of inflammation, CGRP that is present in peripheral tissues is known to enhance inflammatory responses by dilation of blood vessels, extravasation of inflammatory cells, and activation of secretion of inflammatory cytokines, while in the later stage of inflammation, the proinflammatory activities of macrophages and lymphocytes can be inhibited by the action of CGRP, leading to the suppression of inflammatory responses [47].

The CGRP signaling has been implicated in the cAMP, PKC, ERK, and p38 signal transduction pathways [43]. A recent study has shown that CGRP may play an important role in biological defenses including infection and inflammation by integrating the nervous system, hematopoiesis, and immunity. On the one hand, after initiation of inflammation, CGRP stimulates hematopoiesis in the bone marrow cell, which may compensate proliferating hematopoietic cells including monocytes, which are recruited to the local inflammatory tissues. On the other hand, in chronic inflammation, expression of the CGRP receptor CRLR may be reduced and the loss of CGRP function may inhibit hematopoiesis with reduction of myeloid cells, which could terminate the inflammatory responses by reduction of tissue-invaded macrophages [48].

CGRP is one of the main neurotransmitters involved in immune function and is a key responder to tissue damage that is perceived as “pain” [49]. A recent study with Drosophila shows that diuretic hormone 31 (DH31) and CGRP display a similar activity that triggers muscle contractions during bacterial infection which may cause intestinal pain. The traditional treatment may prescribe drugs that inhibit the visceral spasms which probably slows down the elimination of the pathogens

that are responsible for the discomfort. Therefore, a deeper understanding of the CGRP physiological mechanism would be helpful to design novel drugs and adapt medical practices to treat the visceral pain and diarrhea associated with bacterial infection [50].

2.2 Antibacterial neuropeptide, Pituitary Adenylate Cyclase-activating Peptide (PACAP)

Pituitary adenylate cyclase-activating polypeptide (PACAP) is a member of the secretin/glucagon superfamily, with potent anti-inflammatory and potent cytoprotective properties [51], [52]. It is most abundant in the brain, but there are significant levels in other organs, including the thymus, spleen, lymph nodes, and duodenal mucosa [53]. In the current work, analysis of *C. gariepinus* PACAP-38 primary structure reveals the high cationic nature of the peptide, which displays a net charge of +9 at physiological pH and provides evidence of antimicrobial activity of PACAP against a wide spectrum of Gram-negative and Gram-positive bacteria and fungi (**Fig. 1**) [54]. The basic structure of PACAP consists of 38 amino acids. The C-terminal domain of PACAP-38 is involved in the stabilization of the α-helix, and the N-terminal domain of PACAP plays an essential role for binding affinity and biological activity [55]. It has a significant component of hydrophobic residues with a structurally amphipathic arrangement, both hallmarks of canonical AMP. Its sequence has been remarkably conserved throughout evolution, from fish to mammals, suggesting that this peptide fulfills important biological functions in a broad spectrum of organisms [55]. PACAP-27 is a native isoform of PACAP-38. They both had antimicrobial activity against the Gram-negative bacteria *E. coli* in the radial diffusion assay, with PACAP-38 having the highest potency, i.e. the lowest minimum inhibitory concentration (MIC). While PACAP-38 has potent sterilizing activity against *P. aeruginosa* in the more stringent broth dilution assay and antimicrobial activity against the Gram-positive bacteria *S. aureus* in the radial diffusion assay,

PACAP-27 did not have any detectable effect against *S. aureus*. The difference of antimicrobial activity between PACAP-38 and PACAP-27, shows that while the N-terminal of PACAP plays an essential role in the bacterial binding process, the α-helix on the C-terminal end might also play an important role [56].

Although PACAP has highly potent antimicrobial activity against a wide spectrum of bacteria and fungi, as a drug, PACAP38 is limited by its short half-life [57]. After studying the microbial activity of PACAP and its analogs (PACAP3-38), the result shows that they all have potent antimicrobial activity against a panel of Gram-positive and Gram-negative pathogens. While PACAP-38 and PACAP-27 were almost completely cleaved by DPP IV in less than 30 min, there was no indication that any analog tested was cleaved at all by DPP IV even after 72 h. Additionally, they have low toxicity against human RBCs. The potency and species selectivity of the antimicrobial activity of PACAP analogs also improved. PACAP (9-38), which adopts a π-helix conformation rather than an α-helical conformation like PACAP38, exhibits an increased specificity toward *Burkholderia cenocepacia* J2315 compared to other tested bacteria [58]. Besides the mammals, the antimicrobial activities of PACAP have also been explored in the non-mammalian vertebrate immune system. Current work provides evidence of antimicrobial activity of *Clarias gariepinus* PACAP against a wide spectrum of Gram-negative and Gram-positive bacteria and fungi of interest for human medicine and aquaculture, in which computational prediction studies support the putative PACAP therapeutic activity [54]. Overall, the current study contributes to a better understanding of PACAP and its function on the immune system and stimulates renewed interest in PACAP as a new therapeutic agent for the treatment of microbial species.

3. Role of Gut-Brain Axis in Stress Response and Psychological Disorders

While many neuropeptides are involved in communication along the gut-brain axis in the immune response, neuropeptides also are key players in the response to stress. From neuropeptides and neurotransmitters secreted from the brain to gut peptides secreted from enteroendocrine (EEC) cells in the gastrointestinal (GI) tract, neuropeptides and peptide hormones have critical functions in bidirectional communication in the stress response. EEC cells, which are a small portion of epithelial cells in the gut, are regulated through the gut microbiome [60], [61], allowing for bidirectional communication between the brain and the microbiome. Many neuropeptides, such as corticotropin-releasing hormone (CRH) and neuropeptide Y, among others, are involved in communication along the gut-brain axis, including in the response to harmful circumstances, namely stress [62]. Furthermore, the microbiome and signaling along the gut-brain axis can also affect behaviors related to psychological disorders, such as anxiety and depression [63]–[65]. Many endogenous peptides, such as neurotensin, oxytocin, amylin, neuropeptide Y, and cholecystokinin-8, are well studied in psychological disorders [62], [66].

3.1 Connection Between the Gut and Psychological Disorders

Diseases like obesity and inflammatory bowel disease (IBD), are commonly studied due to the involvement of the microbiome and bidirectional communication along the gut-brain axis in these diseases. As the connection between the gut-brain axis and obesity has recently been reviewed [67], this mini review will focus on IBD. IBD affects a diverse population of patients world-wide, especially in western countries. Although IBD is not considered fatal, its effects and associated psychological disorders, such as anxiety and depression, further decrease the quality of life [68], [69]. Studies have shown that these psychological disorders are more prevalent in IBD individuals

compared to healthy controls [70], [71]. Furthermore, stress was correlated with symptoms of IBD, indicating a sensitivity to stress in IBD [72]. Thus, the bi-directional communication between the gut and the brain in IBD influences symptoms of the disease, including the effect of psychological diseases and stress on the gastrointestinal tract. Communication in the gut-brain axis in stress involves a wide range of molecules participating in either brain to gut or gut to brain communication. Here, the focus will be on brain to gut communication during stress via neuropeptides.

Communication during stress in the gut-brain axis commonly occurs through the hypothalamic-pituitary-adrenal (HPA) axis. HPA is the main pathway for stress response and is involved in psychological disorders [73]. This pathway starts with release of corticotropin-releasing hormone (CRH) from the hypothalamus, which then initiates the release of adrenocorticotrophic hormone (ACTH) from the pituitary. The circulation of ACTH releases glucocorticoid hormones, namely cortisol and corticosterone from the adrenal glands [73]. A discussion about the role of CRH and ACTH peptide hormones in the stress response will be provided in the following section, and in the second section a relatively new peptide with potential anxiolytic effects will be discussed in detail. Here, we will provide a short discussion about other neuropeptide changes due to stress.

3.2 Overview of Neuropeptide Changes in Stress

Neuropeptide Y (NPY) is a key neuropeptide in the enteric nervous system, and has anti-stress properties [74]. The role of NPY in inflammatory responses, pain, emotion, mood, cognition, stress, ingestion, and energy homeostasis has previously been reviewed [75], so only a brief discussion will be presented here. NPY levels have been found to decrease in patients with post-

traumatic stress disorder and increase in people exposed to trauma who do not develop or recover from the trauma [75]. In a study in humans with IBD, NPY levels in venous blood samples were found to increase in IBD patients [76]. In rodent models of stress and IBD, NPY levels were also found to be altered. In a dextran sulfate sodium (DSS)-induced colitis model in mice, increased levels of NPY were found in plasma as well as increased NPY expression in the hypothalamus [77], [78]. Another study investigated the effect of electro-acupuncture on two rat models of stress, namely the chronic and acute stress model and the senna gavage and chronic model. This study revealed decreased levels of NPY in the distal colon, hypothalamus, and spinal cord with the two models of stress, but electro-acupuncture restored NPY and other neurotransmitter levels to that of the healthy control group [79]. Finally, a study of minimal traumatic brain injury in rats found that the NPY levels were initially decreased 48 hours after trauma, but then were elevated 30 days after trauma [80]. Additional in-depth reviews of the role of the NPY neuropeptide family in stress can be found elsewhere [81], [82].

A variety of neuropeptides are potentially involved in the gut-brain axis's response to stress. Thyrotropin-releasing hormone was increased in water avoidance stress in both serum and mucosa, and a receptor for the hormone was expressed in the colon mucosal epithelium and myenteric plexus neurons [83]. The role of endogenous opioid signaling has also been investigated, where it was found that there was a switch in the opioid signaling to an excitation effect in dorsal root ganglia [84]. Additionally, apelin and cholecystokinin (CCK) levels were studied during acute restraint stress in rats [85]. Endogenous apelin mediated release of CCK from enteric neurons in stressed rats [85]. Further study of these neuropeptides may provide a deeper understanding of the role of the gut-brain axis in stress response.

3.3 Hypothalamic-Pituitary-Adrenal Axis Communication in the Stress Response

The response to stress along the hypothalamic-pituitary-adrenal axis involves peptide hormones CRH and ACTH. CRH in mammals is a 41 amino acid long peptide. CRH is secreted from the hypothalamus [86], stimulating the release of ACTH, a 39 amino acid peptide, from the pituitary. There are two CRH receptors (CRH-R1 and CRH-R2), which are G protein-coupled receptors from the B1 family. CRH-R1 is highly expressed in the brain, with more limited expression in peripheral tissues, while CRH-R2 has more limited expression in the brain and is more highly expressed in peripheral tissues [87].

In humans, the role of the hypothalamic-pituitary-adrenal axis in stress and IBD is extensively studied, and recent studies in humans examined the effect of administered CRH under different conditions. CRH administration in humans reproduced the effect of public speaking psychological stress, namely increased small intestine permeability [88]. The administration of CRH increased ACTH levels in plasma, which is expected as the function of CRH is to promote ACTH release. Additionally, exogenous administered CRH affects the brain regions associated with emotional-arousal circuitry and pain differently during colorectal distention in control individuals than IBD patients [89]. A follow up study found that hypothalamic-pituitary-adrenal-sympathoadrenal responses to CRH injection were different in control and IBD patients during colorectal distention, indicating that changes in adrenal gland activity in response to ACTH stimulation may be involved with IBD [90].

In rodent studies, different models of IBD and stress are combined to investigate the role of the hypothalamic-pituitary-adrenal axis, including CRH and ACTH. The chronic restraint stress model is one model commonly used in rodents. An investigation of germ-free mice and specific pathogen infected mice under restraint stress revealed higher levels of hypothalamic-pituitary-adrenal axis related compounds, including CRH and ACTH, in germ-free mice [91]. Overall, specific pathogen infected mice exhibited more anxiety behaviors in restraint stress, indicating changes in intestinal microbiota can influence the hypothalamic-pituitary-adrenal axis response to stress [91]. Another study investigated probiotic treatment during restraint stress, revealing that treatment with *Lactobacillus helveticus* NS8 restored ACTH levels and lessened the anxiety and depression symptoms caused by restraint stress, again revealing the importance of the gut microbiome composition in stress [92]. Maternal separation stress early in development and CRH injection both resulted in changes in colorectal mobility and gut microbiome content by the stressors [93]. Furthermore, buserelin treatment, which was expected to cause enteric neuronal loss, resulted in up to 50% loss of neurons and an increase in CRH immunoreactive neurons in the colon. Despite these physical changes, however, the physiology and function of the gut is still well preserved with regards to the stress response and hypothalamic-pituitary-adrenal axis [94].

In a mouse model of IBD, DDS-induced colitis, the effect of stress, via a water avoidance stress model of mild psychological stress, was studied. This revealed changes in CRH, CRH-R1, NPY, and NPY receptor Y1, among others, revealing altered brain-gut signaling during DDS induced colitis [78]. The role of the hypothalamic-pituitary-adrenal axis in visceral nociception was investigated in rats using a colorectal distention protocol, which demonstrated that CRH in the central nucleus of the amygdala was involved in signaling for visceral nociception through release

of noradrenaline after CRH-R1 binding [95]. Thus, changes in the HPA axis occur due to a variety of stressors, and when combined with the important role of the gut microbiome composition in stress, a dynamic gut-brain axis signaling process is revealed.

3.4 Phoenixin – A Neuropeptide with Multiple Roles

In contrast to the well-studied peptides along the HPA axis, novel neuropeptides are still being discovered. Relatively recently, the neuropeptide phoenixin (PNX) was discovered with bioinformatics approach, revealing a 14 amino acid peptide and a 20 amino acid peptide, both of which are amidated [96]. The peptides are highly conserved across species. Phoenixin is primarily localized to the hypothalamus, with peripheral localization to other organs, including the heart, thymus, and stomach [96]. Originally, in studies in rats, phoenixin was found to be involved in reproduction as administration of small interfering RNA (siRNA) against phoenixin *in vivo* delayed estrus and reduced gonadotrophin-releasing hormone (GnRH) receptor expression, which is involved in regulation of the reproductive system [96]. The G protein-coupled receptor Gpr173 is the putative receptor for PNX binding [97], [98]. While the original function of the peptide was in reproduction, with multiple studies focusing on this, there has been a range of other proposed functions for PNX, including roles in psychological disorders, which will be discussed following description of the pathway and localization of PNX.

In vitro studies have provided insight into the pathway for PNX involvement in reproduction. GnRH is involved in regulation of release of reproductive hormones, such as luteinizing hormone (LH). Evidence suggests that PNX mediates GnRH secretion by regulating the expression of the GnRH receptor [96] through activation of the cyclic AMP/ protein kinase A pathway after PNX

binds to the Gpr 173 receptor [98]. Additional research has examined the ability of PNX to affect levels of other molecules, including vasopressin [99], as well as identifying potential regulators of PNX and the PNX receptor [100]–[102]. Interestingly, in women with polycystic ovary syndrome, PNX-14 levels, along with LH and androgen, were increased in the serum of patients compared to controls, indicating a potential role of PNX-14 in polycystic ovary syndrome [103].

The specific immunohistological localization and expression analysis revealed numerous localization patterns in the brain, suggesting that phoenixin has a wide range of functions [96], [104], [105]. A study of the localization of pheonixin-14 (as opposed to the initial studies that used an antibody that bound to both the 14 and 20 amino-acid residue peptides with C-terminal amidation) revealed pheonixin-14 was localized to the medial division of the central amygdaloid nucleus in the brain, spinal trigeminal tract, spinocerebellar tract, and cells between crypts of duodenum, jejunum, and ileum [106]. PNX has been implicated in feeding behavior as studies in rats have shown high co-expression with the neuropeptide nesfatin-1, which is involved in food intake, energy expenditure, and glucose homeostasis [105], along with an increase in food intake upon injection of PNX [107], [108]. The peptide has also been implicated in the suppression of visceral pain [104] as well as signaling of itch sensation [109]. The effect of PNX in the heart has also been studied [110].

Furthermore, PNX levels have been investigated in psychological disorders, namely anxiety, and neurocognitive disorders. In mice, PNX-14 was shown to reduce anxiety in the open field and elevated maze test in a dose dependent manner [111]. In humans, a study in obese men revealed a negative association of PNX and anxiety. Interestingly, when the group was divided into minimal

and moderate anxiety subgroups, the minimal anxiety group had higher PNX levels, although the p-value was just above the level needed for statistical significance [112]. As depression and perceived stress were not correlated with PNX, the PNX anxiety correlation appears to be specific, but the study in obese men revealed only a moderate correlation between PNX and anxiety, potentially due to other factors [112], [113]. For neurocognitive disorders, PNX improved memory formation and retention through GnRH receptor activation in mice, indicating a potential ability of the neuropeptide to enhance memory [114]. Also, PNX was shown to potentially reverses the effects of Alzheimer's Disease as it lessened the effects of Aβ1-42 and scopolamine, which cause memory impairment [114]. PNX levels were tested in humans with either mild cognitive impairment, subjective memory complaints, or Alzheimer's disease, and no significant differences were observed between the groups, although there was no healthy control for comparison [115]. The reported positive effects of PNX on anxiety and cognitive function have the potential to be impactful for treating these disorders. Thus, it will be interesting to follow up with further studies on the role of PNX in anxiety and neurocognitive disorders.

4. Summary

Significant progress has been made over the past decades in recognizing the importance of neuropeptides in bidirectional communication between the brain and the gut and their influence on host immunity and stress. As stated earlier, neuropeptides such as CGRP are involved in host monitoring of the gut environment and have an important function that connects nervous and immune systems during infection. Neuropeptides such as PACAP could directly act on the bacteria membrane to kill the bacteria. The effective defense and low bacterial resistance probability make the antimicrobial neuropeptides an attractive new class of antibiotics. Key findings show that stress

influences the composition of the gut microbiota and that bidirectional communication between microbiota and the CNS influences stress reactivity and responses. Neuropeptide CRH acts as a common mediator during this dynamic gut-brain axis signaling process. The positive effects of PNX on anxiety and cognitive function have the potential to be impactful for development of therapeutic strategies to treat these disorders.

We are just beginning to understand the meaning of gut-to-brain microbiome interactions and what role neuropeptides ultimately play for host homeostasis including immunity and stress. Although their implications are described in selected models, some of the precise mechanisms and overall effects are not fully understood. Further progress in understanding the various processes involved in neuropeptide modulation of the interactions between the gut microbiome and the central and peripheral nervous systems is essential to develop effective treatments for immune and stress disorders with neurogenic components. Because our understanding of the interactions of the gut-brain axis continues to expand, novel therapies will be developed to treat gut microbiome-mediated immune and stress diseases. Neuropeptides, their receptors, and the proteases that degrade the same neuropeptides may become the special target of new specific pharmacologic approaches.

Acknowledgement

Preparation of this manuscript was supported in part by National Science Foundation (CHE-1710140) and National Institutes of Health grant R01DK071801. LL acknowledges a Vilas Distinguished Achievement Professorship and a Charles Melbourne Johnson Distinguished Chair Professorship, and a UW2020 grant with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

TABLES AND FIGURES

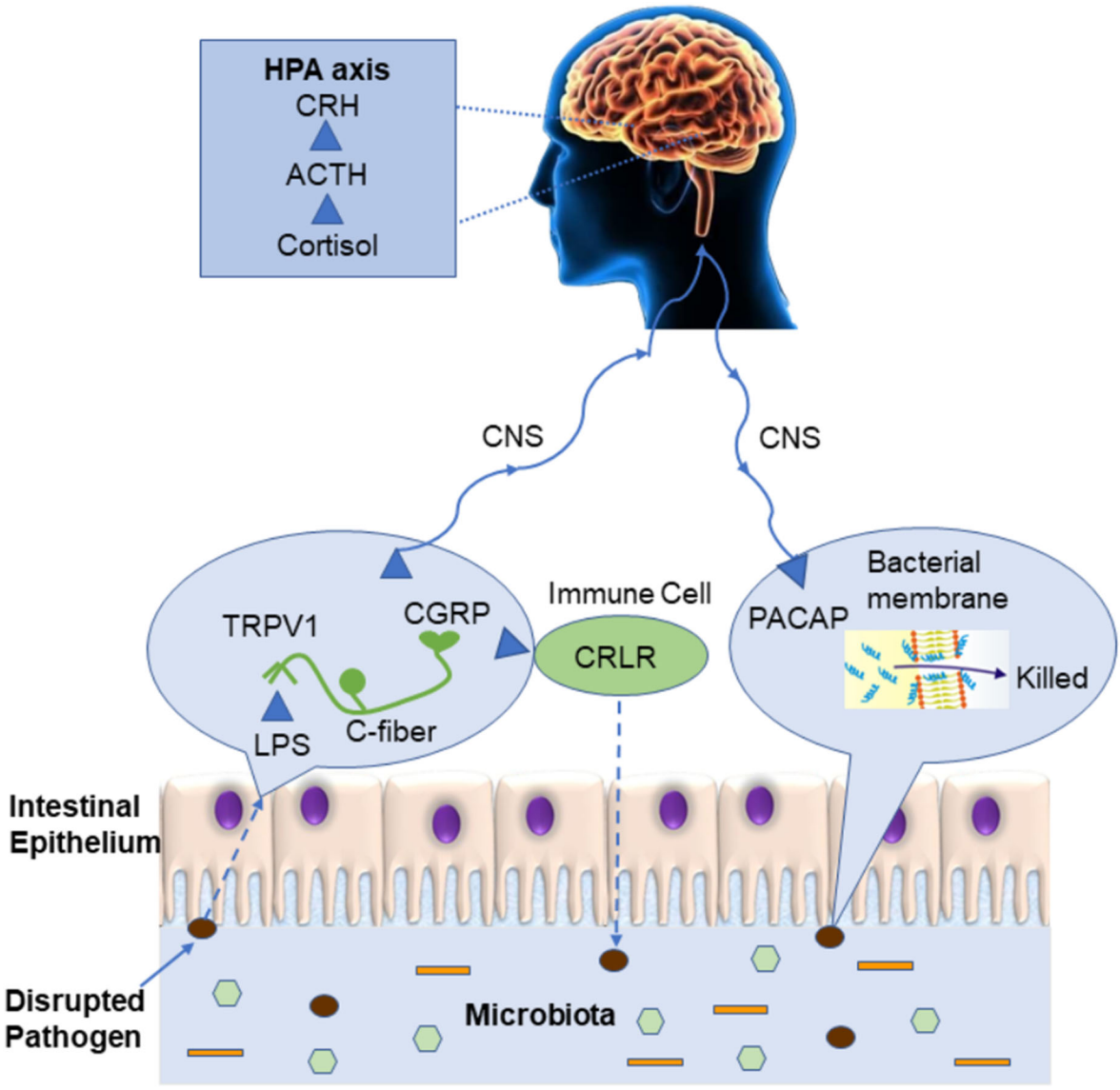


Figure 1. The bi-directional communication between the gut and the brain during pathogen infection and the potential function of neuropeptides on the host immunity and stress. LPS, lipopolysaccharides; TRPV1, transient receptor potential vanilloid receptor-1; CGRP, calcitonin gene-related peptide; CNS, central nervous system; CRLR, calcitonin receptor like receptor; PACAP, pituitary adenylate cyclase-activating polypeptide; CRH, corticotropin-releasing hormone; ACTH, adrenocorticotrophic hormone; HPA axis, hypothalamic-pituitary-adrenal axis.

Table 1 Common Antimicrobial Neuropeptides

Neuropeptide	Origin	Sequence	Antimicrobial Activity
Neuropeptide Y (NPY)	Human	Pro-neuropeptide Y (30-64)	Against Gram-positive bacteria, Gram-negative bacteria, fungi and parasites [116].
Calcitonin gene-related peptide (CGRP)	Human	Calcitonin gene-related peptide 1 (83-119)	Against Gram-positive bacteria, Gram-negative bacteria and fungi [32].
Substance P (SP)	Human	SP (1-11)	Against Gram-positive bacteria, Gram-negative bacteria and fungi [116], [118].
Vasoactive intestinal peptide (VIP)	Human	VIP peptides (125-152)	Against Gram-positive bacteria, Gram-negative bacteria, fungi and parasites [119].
α-melanocyte-stimulating hormone (αMSH)	Human	α-MSH (1–13)	Against Gram-positive bacteria, Gram-negative bacteria, fungi and parasites [120].
Pituitary adenylate cyclase-activating peptide (PACAP)	Human/Catfish	PACAP (131-169, PACAP-38) / PACAP (131-158, PACAP-27)	Against Gram-positive bacteria, Gram-negative bacteria and fungi [54].
Adrenomedullin (AM)	Human	Adrenomedullin (1-52)	Against Gram-positive bacteria, Gram-negative bacteria and parasites [39], [122].
NDA-1	*hydra*	NDA-1 (1-71)	Against Gram-positive bacteria, Gram-negative bacteria [123].
Enkelytin	Human	proenkephalin-A (209–237)	Against Gram-positive bacteria [124].

References

[1] G. K. John and G. E. Mullin, “The Gut Microbiome and Obesity,” *Curr. Oncol. Rep.*, vol. 18, no. 7, p. 45, Jul. 2016.

[2] S. Ghaisas, J. Maher, and A. Kanthasamy, “Gut microbiome in health and disease: Linking the microbiome–gut–brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases,” *Pharmacol. Ther.*, vol. 158, pp. 52–62, Feb. 2016.

[3] C. M. Guinane and P. D. Cotter, “Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ.,” *Therap. Adv. Gastroenterol.*, vol. 6, no. 4, pp. 295–308, Jul. 2013.

[4] A. B. Shreiner, J. Y. Kao, and V. B. Young, “The gut microbiome in health and in disease,” *Curr. Opin. Gastroenterol.*, vol. 31, no. 1, p. 69, Jan. 2015.

[5] D. Artis, “Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut,” *Nat. Rev. Immunol.*, vol. 8, no. 6, pp. 411–420, Jun. 2008.

[6] M. P. Sherman, H. Zaghouani, and V. Niklas, “Gut microbiota, the immune system, and diet influence the neonatal gut–brain axis,” *Pediatr. Res.*, vol. 77, no. 1–2, pp. 127–135, Jan. 2015.

[7] J. A. Foster and K.-A. McVey Neufeld, “Gut–brain axis: how the microbiome influences anxiety and depression,” *Trends Neurosci.*, vol. 36, no. 5, pp. 305–312, May 2013.

[8] A. I. Petra, S. Panagiotidou, E. Hatziagelaki, J. M. Stewart, P. Conti, and T. C. Theoharides, “Gut-Microbiota-Brain Axis and Its Effect on Neuropsychiatric Disorders With Suspected Immune Dysregulation,” *Clin. Ther.*, vol. 37, no. 5, pp. 984–995, May 2015.

[9] R.-B. Yang *et al.*, “Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling,” *Nature*, vol. 395, no. 6699, pp. 284–288, Sep. 1998.

[10] T. H. Frazier, J. K. DiBaise, and C. J. McClain, “Gut Microbiota, Intestinal Permeability, Obesity-Induced Inflammation, and Liver Injury,” *J. Parenter. Enter. Nutr.*, vol. 35, no. 5_suppl, pp. 14S-20S, Sep. 2011.

[11] S. Buhner *et al.*, “Calcium imaging of nerve-mast cell signaling in the human intestine,” *Front. Physiol.*, vol. 8, no. NOV, pp. 1–12, 2017.

[12] A. K. Rajvanshi, “The three minds of the body-Brain, heart and gut,” 2011.

[13] S. H. Rhee, C. Pothoulakis, and E. A. Mayer, “Principles and clinical implications of the brain–gut–enteric microbiota axis,” *Nat. Rev. Gastroenterol. Hepatol.*, vol. 6, no. 5, pp. 306–314, May 2009.

[14] M. Carabotti, A. Scirocco, M. A. Maselli, and C. Severi, “The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems,” *Ann. Gastroenterol. Q. Publ. Hell. Soc. Gastroenterol.*, vol. 28, no. 2, p. 203, 2015.

[15] K. Ichihara, J. Eng, and R. S. Yalow, “Ontogeny of immunoreactive CCK, VIP and secretin in rat brain and gut,” *Biochem. Biophys. Res. Commun.*, vol. 112, no. 3, pp. 891–898, May 1983.

[16] W. D. Barber and T. F. Burks, “Brain-gut interactions: brain stem neuronal response to local gastric effects of substance P.,” *Am. J. Physiol.*, vol. 253, no. 3 Pt 1, pp. G369-77, Sep. 1987.

[18] T. Lotti, A. M. D’Erme, and J. Hercogová, “The role of neuropeptides in the control of regional immunity,” *Clin. Dermatol.*, vol. 32, no. 5, pp. 633–645, Sep. 2014.

[19] F. Sabatino, A. Di Zazzo, L. De Simone, and S. Bonini, “The Intriguing Role of Neuropeptides at the Ocular Surface,” *Ocul. Surf.*, vol. 15, no. 1, pp. 2–14, 2017.

[20] N. Y. Lai, K. Mills, I. M. Chiu, and I. Chiu, “Sensory neuron regulation of gastrointestinal

inflammation and bacterial host defense HHS Public Access," *J Intern Med*, vol. 282, no. 1, pp. 5–23, 2017.

[21] L. E. Goehler, R. P. A. Gaykema, N. Opitz, R. Reddaway, N. Badr, and M. Lyte, "Activation in vagal afferents and central autonomic pathways: Early responses to intestinal infection with Campylobacter jejuni," *Brain. Behav. Immun.*, vol. 19, no. 4, pp. 334–344, Jul. 2005.

[22] P. Bercik *et al.*, "The Intestinal Microbiota Affect Central Levels of Brain-Derived Neurotropic Factor and Behavior in Mice," *Gastroenterology*, vol. 141, no. 2, pp. 599-609.e3, Aug. 2011.

[23] N. Sudo *et al.*, "Postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice.," *J. Physiol.*, vol. 558, no. Pt 1, pp. 263–75, Jul. 2004.

[24] J. A. Foster, L. Rinaman, and J. F. Cryan, "Stress & the gut-brain axis: Regulation by the microbiome," *Neurobiol. Stress*, vol. 7, pp. 124–136, Dec. 2017.

[25] G. De Palma *et al.*, "Microbiota and host determinants of behavioural phenotype in maternally separated mice," *Nat. Commun.*, vol. 6, no. 1, p. 7735, Nov. 2015.

[26] A. V. Golubeva *et al.*, "Prenatal stress-induced alterations in major physiological systems correlate with gut microbiota composition in adulthood," *Psychoneuroendocrinology*, vol. 60, pp. 58–74, Oct. 2015.

[27] J. Barouei, M. Moussavi, and D. M. Hodgson, "Effect of Maternal Probiotic Intervention on HPA Axis, Immunity and Gut Microbiota in a Rat Model of Irritable Bowel Syndrome," *PLoS One*, vol. 7, no. 10, p. e46051, Oct. 2012.

[28] N. Moussaoui *et al.*, "Chronic Early-life Stress in Rat Pups Alters Basal Corticosterone, Intestinal Permeability, and Fecal Microbiota at Weaning: Influence of Sex.," *J.*

Neurogastroenterol. Motil., vol. 23, no. 1, pp. 135–143, Jan. 2017.

[29] T. Jin, H. Wu, Y. Wang, and H. Peng, "Capsaicin induces immunogenic cell death in human osteosarcoma cells," *Exp. Ther. Med.*, vol. 12, no. 2, pp. 765–770, Aug. 2016.

[30] M. Granato, M. S. Gilardini Montani, M. Filardi, A. Faggioni, and M. Cirone, "Capsaicin triggers immunogenic PEL cell death, stimulates DCs and reverts PEL-induced immune suppression.," *Oncotarget*, vol. 6, no. 30, pp. 29543–54, Oct. 2015.

[31] J. E. Morley, N. E. Kay, G. F. Solomon, and N. P. Plotnikoff, "Neuropeptides: Conductors of the immune orchestra," *Life Sci.*, vol. 41, no. 5, pp. 527–544, Aug. 1987.

[32] I. A. El Karim, G. J. Linden, D. F. Orr, and F. T. Lundy, "Antimicrobial activity of neuropeptides against a range of micro-organisms from skin, oral, respiratory and gastrointestinal tract sites," *J. Neuroimmunol.*, vol. 200, no. 1–2, pp. 11–16, Aug. 2008.

[33] X. Zhu, Y. Han, J. Du, R. Liu, K. Jin, and W. Yi, "Microbiota-gut-brain axis and the central nervous system.," *Oncotarget*, vol. 8, no. 32, pp. 53829–53838, Aug. 2017.

[34] S. S. Yarandi, D. A. Peterson, G. J. Treisman, T. H. Moran, and P. J. Pasricha, "Modulatory effects of gut microbiota on the central nervous system: How gut could play a role in neuropsychiatric health and diseases," *J. Neurogastroenterol. Motil.*, vol. 22, no. 2, pp. 201–212, 2016.

[35] B. N. Lambrecht, "Immunologists getting nervous: neuropeptides, dendritic cells and T cell activation," *Respir. Res.*, vol. 2, no. 3, p. 133, Apr. 2001.

[36] M. P. Stewart, C. Cabanas, N. Hogg, L. Steinman, and O. Lider, "T cell adhesion to intercellular adhesion molecule-1 (ICAM-1) is controlled by cell spreading and the activation of integrin LFA-1.," *J. Immunol.*, vol. 156, no. 5, pp. 1810–7, Mar. 1996.

[37] A. Chen, Y. Ganor, S. Rahimipour, N. Ben-Aroya, Y. Koch, and M. Levite, "The

neuropeptides GnRH-II and GnRH-I are produced by human T cells and trigger laminin receptor gene expression, adhesion, chemotaxis and homing to specific organs," *Nat. Med.*, vol. 8, no. 12, pp. 1421–1426, Dec. 2002.

[38] J.-P. Timmermans, D. W. Scheuermann, W. Stach, D. Adriaensen, and M. H. A. De Groodt-Lasseel, "Distinct distribution of CGRP-, enkephalin-, galanin-, neuromedin U-, neuropeptide Y-, somatostatin-, substance P-, VIP- and serotonin-containing neurons in the two submucosal ganglionic neural networks of the porcine small intestine," *Cell Tissue Res.*, vol. 260, no. 2, pp. 367–379, May 1990.

[39] D. Augustyniak, J. Nowak, and F. T. Lundy, "Direct and indirect antimicrobial activities of neuropeptides and their therapeutic potential.," *Curr. Protein Pept. Sci.*, vol. 13, no. 8, pp. 723–38, Dec. 2012.

[40] T. Ganz, "Defensins: antimicrobial peptides of innate immunity," *Nat. Rev. Immunol.*, vol. 3, no. 9, pp. 710–720, Sep. 2003.

[41] M. Zasloff, "Antimicrobial peptides of multicellular organisms," vol. 415, no. January, pp. 389–395, 2002.

[42] R. Uddman, L. Edvinsson, E. Ekblad, R. Håkanson, and F. Sundler, "Calcitonin gene-related peptide (CGRP): perivascular distribution and vasodilatory effects," *Regul. Pept.*, vol. 15, no. 1, pp. 1–23, Aug. 1986.

[43] F. A. Russell, R. King, S.-J. Smillie, X. Kodji, and S. D. Brain, "Calcitonin Gene-Related Peptide: Physiology and Pathophysiology," *Physiol. Rev.*, vol. 94, no. 4, pp. 1099–1142, 2014.

[44] B. M. Assas, J. I. Pennock, and J. A. Miyan, "Calcitonin gene-related peptide is a key neurotransmitter in the neuro-immune axis," *Front. Neurosci.*, vol. 8, no. 8 FEB, pp. 1–9,

2014.

[45] R. Yaraee, M. Ebtekar, A. Ahmadiani, and F. Sabahi, “Neuropeptides (SP and CGRP) augment pro-inflammatory cytokine production in HSV-infected macrophages,” *Int. Immunopharmacol.*, vol. 3, no. 13–14, pp. 1883–1887, Dec. 2003.

[46] B. M. Assas, M. H. Wakid, H. A. Zakai, J. A. Miyan, and J. L. Pennock, “Transient receptor potential vanilloid 1 expression and function in splenic dendritic cells: a potential role in immune homeostasis.,” *Immunology*, vol. 147, no. 3, pp. 292–304, Mar. 2016.

[47] U. Hanesch, U. Pfrommer, B. D. Grubb, and H.-G. Schaible, “Acute and Chronic Phases of Unilateral Inflammation in Rat’s Ankle are Associated with an Increase in the Proportion of Calcitonin Gene-related Peptide-immunoreactive Dorsal Root Ganglion Cells,” *Eur. J. Neurosci.*, vol. 5, no. 2, pp. 154–161, Feb. 1993.

[48] A. Suekane *et al.*, “CGRP-CRLR/RAMP1 signal is important for stress-induced hematopoiesis,” *Sci. Rep.*, vol. 9, no. 1, p. 429, Dec. 2019.

[49] S. Benemei, P. Nicoletti, J. G. Capone, and P. Geppetti, “CGRP receptors in the control of pain and inflammation,” *Curr. Opin. Pharmacol.*, vol. 9, no. 1, pp. 9–14, Feb. 2009.

[50] O. Benguettat *et al.*, “The DH31/CGRP enteroendocrine peptide triggers intestinal contractions favoring the elimination of opportunistic bacteria,” *PLoS Pathog.*, vol. 14, no. 9, pp. 1–26, 2018.

[51] H. Ji *et al.*, “Neuropeptide PACAP in Mouse Liver Ischemia and Reperfusion Injury: Immunomodulation by the cAMP-PKA Pathway,” *HEPATOLOGY*, vol. 57, pp. 1225–1237, 2013.

[52] D. Reglodi *et al.*, “PACAP is an Endogenous Protective Factor—Insights from PACAP-Deficient Mice,” *J. Mol. Neurosci.*, vol. 48, no. 3, pp. 482–492, Nov. 2012.

[53] M. DELGADO *et al.*, “PACAP in Immunity and Inflammation,” *Ann. N. Y. Acad. Sci.*, vol. 992, no. 1, pp. 141–157, May 2003.

[54] J. M. Lugo *et al.*, “Evidence for antimicrobial and anticancer activity of pituitary adenylate cyclase-activating polypeptide (PACAP) from North African catfish (Clarias gariepinus): Its potential use as novel therapeutic agent in fish and humans,” *Fish Shellfish Immunol.*, vol. 86, pp. 559–570, Mar. 2019.

[55] S. Debbabi *et al.*, “Antibacterial properties of the pituitary adenylate cyclase-activating polypeptide: A new human antimicrobial peptide.,” *PLoS One*, vol. 13, no. 11, p. e0207366, 2018.

[56] N. D. Doan *et al.*, “Design and characterization of novel cell-penetrating peptides from pituitary adenylate cyclase-activating polypeptide,” *J. Control. Release*, vol. 163, no. 2, pp. 256–265, 2012.

[57] B. D. Green, N. Irwin, and P. R. Flatt, “Pituitary adenylate cyclase-activating peptide (PACAP): Assessment of dipeptidyl peptidase IV degradation, insulin-releasing activity and antidiabetic potential,” *Peptides*, vol. 27, no. 6, pp. 1349–1358, Jun. 2006.

[58] S. Debbabi *et al.*, “Antibacterial properties of the pituitary adenylate cyclase-activating polypeptide: A new human antimicrobial peptide,” *PLoS One*, vol. 13, no. 11, pp. 1–15, 2018.

[60] G. J. Dockray, “Enteroendocrine cell signalling via the vagus nerve,” *Curr. Opin. Pharmacol.*, vol. 13, no. 6, pp. 954–958, Dec. 2013.

[61] L. J. Cohen *et al.*, “Commensal bacteria make GPCR ligands that mimic human signalling molecules,” *Nature*, vol. 549, no. 7670, pp. 48–53, Sep. 2017.

[62] G. Lach, H. Schellekens, T. G. Dinan, and J. F. Cryan, “Anxiety, Depression, and the

Microbiome: A Role for Gut Peptides," *Neurotherapeutics*, vol. 15, no. 1, pp. 36–59, Jan. 2018.

[63] A. Burokas *et al.*, "Targeting the Microbiota-Gut-Brain Axis: Prebiotics Have Anxiolytic and Antidepressant-like Effects and Reverse the Impact of Chronic Stress in Mice," *Biol. Psychiatry*, vol. 82, no. 7, pp. 472–487, Oct. 2017.

[64] Diaz Heijtz R *et al.*, "Normal gut microbiota modulates brain development and behavior," *Proc. Natl. Acad. Sci.*, vol. 108, no. 7, pp. 3047–3052, Feb. 2011.

[65] M.-L. Wong *et al.*, "Inflammasome signaling affects anxiety- and depressive-like behavior and gut microbiome composition," *Mol. Psychiatry*, vol. 21, no. 6, pp. 797–805, Jun. 2016.

[66] P. McGonigle, "Peptide therapeutics for CNS indications," *Biochem. Pharmacol.*, vol. 83, no. 5, pp. 559–566, Mar. 2012.

[67] E. Niccolai, F. Boem, E. Russo, and A. Amedei, "The gut–brain axis in the neuropsychological disease model of obesity: A classical movie revised by the emerging director 'microbiome,'" *Nutrients*, vol. 11, no. 1. MDPI AG, 01-Jan-2019.

[68] E. Guthrie, J. Jackson, J. Shaffer, D. Thompson, B. Tomenson, and F. Creed, "Psychological disorder and severity of inflammatory bowel disease predict health-related quality of life in ulcerative colitis and Crohn's disease," *Am. J. Gastroenterol.*, vol. 97, no. 8, pp. 1994–1999, Aug. 2002.

[69] K. Nordin, L. Påhlman, K. Larsson, M. Sundberg-Hjelm, and L. Lööf, "Health-related quality of life and psychological distress in a population-based sample of Swedish patients with inflammatory bowel disease.," *Scand. J. Gastroenterol.*, vol. 37, no. 4, pp. 450–7, Apr. 2002.

[70] J. R. Walker *et al.*, "The Manitoba IBD Cohort Study: A Population-Based Study of the

Prevalence of Lifetime and 12-Month Anxiety and Mood Disorders," *Am. J. Gastroenterol.*, vol. 103, no. 8, pp. 1989–1997, Aug. 2008.

[71] L. A. Graff *et al.*, "Stress Coping, Distress and Health Perceptions in Inflammatory Bowel Disease and Community Controls," *Am. J. Gastroenterol.*, vol. 104, no. 12, pp. 2959–2969, Dec. 2009.

[72] Whitehead WE, Crowell MD, Robinson JC, Heller BR, and Schuster MM, "Effects of stressful life events on bowel symptoms: subjects with irritable bowel syndrome compared with subjects without bowel dysfunction.," *Gut*, vol. 33, no. 6, pp. 825–30, Jun. 1992.

[73] R. L. Spencer and T. Deak, "A users guide to HPA axis research," *Physiol. Behav.*, vol. 178, pp. 43–65, Sep. 2017.

[74] A. Thorsell, "Brain neuropeptide Y and corticotropin-releasing hormone in mediating stress and anxiety," *Exp. Biol. Med.*, vol. 235, no. 10, pp. 1163–1167, Oct. 2010.

[75] P. Holzer and A. Farzi, "Neuropeptides and the Microbiota-Gut-Brain Axis," in *Advances in experimental medicine and biology*, vol. 817, 2014, pp. 195–219.

[76] C. Stasi *et al.*, "Neuroendocrine Dysregulation in Irritable Bowel Syndrome Patients: A Pilot Study," *J. Neurogastroenterol. Motil.*, vol. 23, no. 3, pp. 428–434, Jul. 2017.

[77] A. M. Hassan *et al.*, "Repeated predictable stress causes resilience against colitis-induced behavioral changes in mice," *Front. Behav. Neurosci.*, vol. 8, p. 386, Nov. 2014.

[78] F. Reichmann, A. M. Hassan, A. Farzi, P. Jain, R. Schuligoi, and P. Holzer, "Dextran sulfate sodium-induced colitis alters stress-associated behaviour and neuropeptide gene expression in the amygdala-hippocampus network of mice," *Sci. Rep.*, vol. 5, no. 1, p. 9970, Sep. 2015.

[79] J. Sun *et al.*, "Electro-acupuncture decreases 5-HT, CGRP and increases NPY in the brain-gut axis in two rat models of Diarrhea-predominant irritable bowel syndrome(D-IBS),"

BMC Complement. Altern. Med., vol. 15, no. 1, p. 340, Dec. 2015.

[80] S. Sagarkar, S. Mahajan, A. G. Choudhary, C. D. Borkar, D. M. Kokare, and A. J. Sakharkar, "Traumatic stress-induced persistent changes in DNA methylation regulate neuropeptide Y expression in rat jejunum," *Neurogastroenterol. Motil.*, vol. 29, no. 9, p. e13074, Sep. 2017.

[81] A. Farzi, F. Reichmann, and P. Holzer, "The homeostatic role of neuropeptide Y in immune function and its impact on mood and behaviour," *Acta Physiol.*, vol. 213, no. 3, pp. 603–627, Mar. 2015.

[82] F. Reichmann and P. Holzer, "Neuropeptide Y: A stressful review," *Neuropeptides*, vol. 55, pp. 99–109, Feb. 2016.

[83] Y. Zhang, C. Wang, and L. Zhang, "The potential role of thyrotropin-releasing hormone in colonic dysmotility induced by water avoidance stress in rats," *Neuropeptides*, vol. 70, pp. 47–54, Aug. 2018.

[84] R. Guerrero-Alba *et al.*, "Stress activates pronociceptive endogenous opioid signalling in DRG neurons during chronic colitis," *Gut*, vol. 66, no. 12, pp. 2121–2131, Dec. 2017.

[85] M. Bülbül, O. Sinen, O. Bayramoğlu, and G. Akkoyunlu, "Acute restraint stress induces cholecystokinin release via enteric apelin," *Neuropeptides*, vol. 73, pp. 71–77, Feb. 2019.

[86] W. Vale, J. Spiess, C. Rivier, and J. Rivier, "Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and beta-endorphin.," *Science*, vol. 213, no. 4514, pp. 1394–7, Sep. 1981.

[87] K. D. Ketchesin, G. S. Stinnett, and A. F. Seasholtz, "Corticotropin-releasing hormone-binding protein and stress: from invertebrates to humans," *Stress*, vol. 20, no. 5, pp. 449–464, Sep. 2017.

[88] T. Vanuytsel *et al.*, “Psychological stress and corticotropin-releasing hormone increase intestinal permeability in humans by a mast cell-dependent mechanism,” *Gut*, vol. 63, no. 8, pp. 1293–1299, Aug. 2014.

[89] Y. Tanaka *et al.*, “Differential Activation in Amygdala and Plasma Noradrenaline during Colorectal Distention by Administration of Corticotropin-Releasing Hormone between Healthy Individuals and Patients with Irritable Bowel Syndrome,” *PLoS One*, vol. 11, no. 7, p. e0157347, Jul. 2016.

[90] Y. Tanaka, M. Kanazawa, M. Kano, M. Tashiro, and S. Fukudo, “Relationship between sympathoadrenal and pituitary-adrenal response during colorectal distention in the presence of corticotropin-releasing hormone in patients with irritable bowel syndrome and healthy controls,” *PLoS One*, vol. 13, no. 7, p. e0199698, Jul. 2018.

[91] R. Huo *et al.*, “Microbiota Modulate Anxiety-Like Behavior and Endocrine Abnormalities in Hypothalamic-Pituitary-Adrenal Axis,” *Front. Cell. Infect. Microbiol.*, vol. 7, p. 489, Nov. 2017.

[92] S. Liang *et al.*, “Administration of Lactobacillus helveticus NS8 improves behavioral, cognitive, and biochemical aberrations caused by chronic restraint stress,” *Neuroscience*, vol. 310, pp. 561–577, Dec. 2015.

[93] T. Murakami *et al.*, “Changes in Intestinal Motility and Gut Microbiota Composition in a Rat Stress Model,” *Digestion*, vol. 95, no. 1, pp. 55–60, 2017.

[94] E. Sand *et al.*, “Buserelin treatment to rats causes enteric neurodegeneration with moderate effects on CRF-immunoreactive neurons and Enterobacteriaceae in colon, and in acetylcholine-mediated permeability in ileum,” *BMC Res. Notes*, vol. 8, no. 1, p. 824, Dec. 2015.

[95] J. Su, Y. Tanaka, T. Muratsubaki, M. Kano, M. Kanazawa, and S. Fukudo, “Injection of corticotropin-releasing hormone into the amygdala aggravates visceral nociception and induces noradrenaline release in rats,” *Neurogastroenterol. Motil.*, vol. 27, no. 1, pp. 30–39, Jan. 2015.

[96] G. L. C. Yosten *et al.*, “A Novel Reproductive Peptide, Phoenixin,” *J. Neuroendocrinol.*, vol. 25, no. 2, pp. 206–215, Feb. 2013.

[97] L. M. Stein *et al.*, “Hypothalamic action of phoenixin to control reproductive hormone secretion in females: importance of the orphan G protein-coupled receptor Gpr173,” *Am. J. Physiol. Integr. Comp. Physiol.*, vol. 311, no. 3, pp. R489–R496, Sep. 2016.

[98] A. K. Treen, V. Luo, and D. D. Belsham, “Phoenixin Activates Immortalized GnRH and Kisspeptin Neurons Through the Novel Receptor GPR173,” *Mol. Endocrinol.*, vol. 30, no. 8, pp. 872–888, Aug. 2016.

[99] S. Gasparini *et al.*, “Novel regulator of vasopressin secretion: phoenixin,” *Am. J. Physiol. Integr. Comp. Physiol.*, vol. 314, no. 4, pp. R623–R628, Apr. 2018.

[100] E. K. McIlwraith, N. Loganathan, and D. D. Belsham, “Phoenixin Expression Is Regulated by the Fatty Acids Palmitate, Docosahexaenoic Acid and Oleate, and the Endocrine Disrupting Chemical Bisphenol A in Immortalized Hypothalamic Neurons,” *Front. Neurosci.*, vol. 12, p. 838, Nov. 2018.

[101] A. Suszka-Świtek *et al.*, “The Gn RH analogues affect novel neuropeptide SMIM 20/phoenixin and GPR 173 receptor expressions in the female rat hypothalamic–pituitary–gonadal (HPG) axis,” *Clin. Exp. Pharmacol. Physiol.*, vol. 46, no. 4, pp. 350–359, Apr. 2019.

[102] E. K. McIlwraith, N. Loganathan, and D. D. Belsham, “Regulation of Gpr173 expression,

a putative phoenixin receptor, by saturated fatty acid palmitate and endocrine-disrupting chemical bisphenol A through a p38-mediated mechanism in immortalized hypothalamic neurons," *Mol. Cell. Endocrinol.*, vol. 485, pp. 54–60, Apr. 2019.

[103] K. Ullah *et al.*, "Phoenixin-14 concentrations are increased in association with luteinizing hormone and nesfatin-1 concentrations in women with polycystic ovary syndrome," *Clin. Chim. Acta*, vol. 471, pp. 243–247, Aug. 2017.

[104] R.-M. Lyu *et al.*, "Phoenixin: a novel peptide in rodent sensory ganglia.," *Neuroscience*, vol. 250, pp. 622–31, Oct. 2013.

[105] A. Pałasz, E. Rojczyk, K. Bogus, J. J. Worthington, and R. Wiaderkiewicz, "The novel neuropeptide phoenixin is highly co-expressed with nesfatin-1 in the rat hypothalamus, an immunohistochemical study," *Neurosci. Lett.*, vol. 592, pp. 17–21, Apr. 2015.

[106] P. Prinz *et al.*, "Central and peripheral expression sites of phoenixin-14 immunoreactivity in rats," *Biochem. Biophys. Res. Commun.*, vol. 493, no. 1, pp. 195–201, Nov. 2017.

[107] M. Schalla *et al.*, "Phoenixin-14 injected intracerebroventricularly but not intraperitoneally stimulates food intake in rats," *Peptides*, vol. 96, pp. 53–60, Oct. 2017.

[108] T. Friedrich *et al.*, "Intracerebroventricular injection of phoenixin alters feeding behavior and activates nesfatin-1 immunoreactive neurons in rats," *Brain Res.*, vol. 1715, pp. 188–195, Jul. 2019.

[109] A. Cowan, R.-M. Lyu, Y.-H. Chen, S. L. Dun, J.-K. Chang, and N. J. Dun, "Phoenixin: A candidate pruritogen in the mouse.," *Neuroscience*, vol. 310, pp. 541–8, Dec. 2015.

[110] C. Rocca *et al.*, "Phoenixin-14: detection and novel physiological implications in cardiac modulation and cardioprotection," *Cell. Mol. Life Sci.*, vol. 75, no. 4, pp. 743–756, Feb. 2018.

[111] J. H. Jiang *et al.*, “Effects of Phoenixin-14 on anxiolytic-like behavior in mice,” *Behav. Brain Res.*, vol. 286, pp. 39–48, Jun. 2015.

[112] T. Hofmann *et al.*, “Phoenixin is negatively associated with anxiety in obese men,” *Peptides*, vol. 88, pp. 32–36, Feb. 2017.

[113] M. A. Schalla and A. Stengel, “The role of phoenixin in behavior and food intake,” *Peptides*, vol. 114, pp. 38–43, Apr. 2019.

[114] J. H. Jiang *et al.*, “Phoenixin-14 enhances memory and mitigates memory impairment induced by Aβ1-42 and scopolamine in mice,” *Brain Res.*, vol. 1629, pp. 298–308, Dec. 2015.

[115] M. Yuruyen *et al.*, “Does plasma phoenixin level associate with cognition? Comparison between subjective memory complaint, mild cognitive impairment, and mild Alzheimer’s disease,” *Int. Psychogeriatrics*, vol. 29, no. 9, pp. 1543–1550, Sep. 2017.

[116] C. J. Hansen, K. K. Burnell, and K. A. Brogden, “Antimicrobial activity of Substance P and Neuropeptide Y against laboratory strains of bacteria and oral microorganisms,” *J. Neuroimmunol.*, vol. 177, no. 1–2, pp. 215–218, Aug. 2006.

[118] K. Kowalska, D. B. Carr, and A. W. Lipkowski, “Direct antimicrobial properties of substance P,” *Life Sci.*, vol. 71, no. 7, pp. 747–750, Jul. 2002.

[119] J. Campos-Salinas *et al.*, “Therapeutic efficacy of stable analogues of vasoactive intestinal peptide against pathogens.,” *J. Biol. Chem.*, vol. 289, no. 21, pp. 14583–99, May 2014.

[120] M. Cutuli, S. Cristiani, J. M. Lipton, and A. Catania, “Antimicrobial effects of α-MSH peptides,” *J. Leukoc. Biol.*, vol. 67, no. 2, pp. 233–239, Feb. 2000.

[122] R. P. Allaker and S. Kapas, “Adrenomedullin and mucosal defence: interaction between host and microorganism,” *Regul. Pept.*, vol. 112, no. 1–3, pp. 147–152, Apr. 2003.

[123] R. Augustin *et al.*, “A secreted antibacterial neuropeptide shapes the microbiome of Hydra,” *Nat. Commun.*, vol. 8, no. 1, p. 698, Dec. 2017.

[124] Y. Goumon *et al.*, “Characterization of antibacterial COOH-terminal proenkephalin-A-derived peptides (PEAP) in infectious fluids. Importance of enkelytin, the antibacterial PEAP209-237 secreted by stimulated chromaffin cells.,” *J. Biol. Chem.*, vol. 273, no. 45, pp. 29847–56, Nov. 1998.

Chapter 7

Integrated, Multi-Omics Strategy to Study the Host Immune Response Induced by Human Microbiome during *Salmonella enterica* Typhimurium and *Candida albicans* Infection

Adapted from **Wei P**, Keller C, Bratburd JR, Liu R, Vivas E, Gemperline E, Rey FE, Currie CR, Li LL. "Integrated , Multi-Omics Strategy to Study the Host Immune Response Induced by Human Microbiome during Salmonella enterica Typhimurium and Candida albicans Infection. " *To be submitted*. **Wei P**, designed and conducted the peptidomics and proteomics study under the supervision of Li LL; Keller C conducted the metabolomics study, Bratburd JR provided the samples, Buchberger AR systhesized the DiLeu tags, Liu R, Vivas E, Gemperline E, Rey FE, Currie CR provided insights in gut microbiome and gut-brain axis.

ABSTRACT

The gut microbiome performs many critical roles in human health, including host metabolism, immune system modulation, and even brain function. While the disruptive pathogen invaded, the "good" microbiome may help the host in both direct and indirect ways. They may produce naturally produced against the invasive pathogen, and they may also enhance the host immune system to defend. How does this process work, and what biomolecules are involved remain unclear. Therefore, the goal of this study is to investigate how the host immune system is shaped by human microbiome during infections. Herein, the potential metabolites produced by the microbiome is investigated, the host proteome changes during infection are also studied; furthermore, neuropeptides from the host were identified to explore the gut-brain axis response during infection. Mass spectrometry is a sensitive and robust technique to profile not just small molecule changes produced by the microbiome due to infection, but also protein and neuropeptide changes due to infection. Thus, a combination of the metabolome, proteome, and peptidome-based approach was used to profile the effect of the microbiome on the host during infection by both *Salmonella enterica* Typhimurium and *Candida albicans*. The results from the metabolomics and proteomics studies reveal molecular changes that highlight the role of the microbiome in the immune and inflammatory response to infection—analysis of three omics from the same sample. The neuropeptides found in the gut, specifically induced by the microbiome under infection, could facilitate further understanding of the bidirectional communication between gut microbiota, brain, and immune system.

KEYWORDS: LC-MS; Metabolomics; Peptidomics; Proteomics; Microbiome; Infection; Multiomics

INTRODUCTION

In the gastrointestinal tract, a large and dynamic community of microbes, the gut microbiome, reside, where three domains of life, namely archaea, bacteria, and eukaryote, are represented.[1, 2] These microbes have an essential role in human health. They affect the nutrition of the host as they carry out various digestive processes. They could also modulate immune function.[3, 4] Additionally, disruptions to the normal composition of the microbiome are associated with various diseases, including obesity,[5, 6] inflammatory bowel disease,[7] and cardiovascular disease.[8, 9] Consequently, a variety of "omic" techniques have been employed to study the microbiome. For example, there are metagenomic and metaproteomic studies on the gut microbiome as well as studies on how the microbiome influences metabolites in other systems of the body, such as the blood and cerebrospinal fluid.[10-13]

Microbial metabolism produces both primary and secondary metabolites through the gut microbiome metagenome that influence health and disease.[4] The role of metabolites produced by the microbiome is not limited to nutrition and metabolism. Still, it can also provide defense against pathogens, and immune regulation among many other functions.[14, 15] Furthermore, the ability of secondary metabolites synthesized by the microbiome to protect the host against pathogens could potentially be used for discovery of new natural products, which is of interest due to the rise of antibiotic resistance.[16] Mass spectrometry (MS) is a powerful tool to study metabolites in a sample, and in conjunction with nuclear magnetic resonance (NMR), can aid in the natural product in discovery.[17] For example, metabolomic changes of the microbiome in response to disturbances to the microbiome has been profiled with MS techniques.[18, 19]

Metaproteomics studies on the microbiome provide valuable information on expressed proteins involved in various biological processes. For example, metaproteomics has been used to study the metabolism of the gut microbiome in premature infants,[20] and the proteome changes in the microbiome due to disease, including Chrohn's disease and inflammatory bowel disease.[21, 22] Additionally, metaproteomics is often coupled with metabolomics, as well as other "omics" for multi-omics analysis of the biological processes occurring in the gut.[23] For example, metagenomics and metabolomics are often combined to study how different disruptions to the gut microbiome affect microbiome composition and metabolism.[19, 24] Metaproteomics studies can also be used in with metagenomic and metabolomics techniques to provide a comprehensive analysis of active processes in the gut microbiome or the effect of stressors on the microbiome.[25]

Peptidomics has risen to importance due to its ability to identify and profile changes in endogenous peptides, especially neuropeptides.[26, 27] Neuropeptides are an important class of signaling molecules that help to regulate many biological processes.[28, 29] The gut-brain axis is a critical system whereby microbes in the gut can communicate with the host brain and vise versa. Bacteria in the gut can produce neurotransmitters, while signaling from the brain can be accomplished through cytokine molecules.[30] As neuropeptides, and other bioactive peptides, are critical signaling molecules, peptidomics analysis of the gut microbiome can provide key insights into communication in the host-microbiome relationship and the gut-brain axis.[31]

Here a multi-omics approach will be applied to study the response of a model human gut microbiome to infection by *Salmonella enterica* Typhimurium and *Candida albicans* in mice. Previously a metagenomics and metabolomics approach was utilized to investigate microbiome

community and small molecule changes in the gut microbiome in response to infection.[32] To further understand the host and microbiome's response to infection, another multi-omics approach combining metabolomics, peptidomics, and proteomics was investigated here. The combination of peptidomics with metabolomics and proteomics is less commonly employed but will provide a great depth of information, especially in regard to communication in the gut-brain axis. Peptidomics analysis resulted in the identification of multiple neuropeptides in the gut microbiome. The combined metabolomics, peptidomics, and proteomics analysis of the gut microbiome resulted in the discovery of immune system changes in response to infection.

MATERIALS AND METHODS

Sample Collection and Extraction

Mouse experiments, including inoculation with approximately 90 strains representing a model human microbiome and infection with either *Salmonella enterica* Typhimurium (Salmonella) or *Candida albicans* (Candida), were performed as previously described.[32] A methanol chloroform water extraction was performed, and a 3 kDa molecular weight cut-off performed on the aqueous and organic fractions.[32] The flow-through was collected as the metabolite fractions (below 3 kDa). The 3 kDa molecular weight cut-off device was equilibrated with 0.4 mL of 50% methanol and then flipped over and centrifuged (14,000 x g, 2 min) to collect content above three kDa. A 30 kDa MWCO was performed on the contents above 3 kDa to separate the extract into peptidomics (three kDa to 30kDa) and proteomics (above 30kDa) fractions.

Metabolomics Acquisition and Analysis

Aqueous metabolite fractions (below 3 kDa) were resuspended at 10 mg/mL in optima grade water with 0.1% formic acid. Liquid chromatography-mass spectrometry (LC-MS) for label-free quantitation was performed as previously described.[32] LC-MS/MS (tandem MS) was performed on samples with the same gradient as the MS1 data acquisition using a top 5 DDA method with inclusion lists for targets up-regulated in each sample. Full MS parameters were 70 k resolution, 1E6 AGC target, 100 ms maximum injection time, and 200 to 1700 *m/z*. MS/MS parameters were 70 k resolution, 5E5 AGC target, 100 ms maximum injection time, and 1.0 isolation width. A 10 s dynamic exclusion was employed. Three technical replicates were collected for MS/MS runs, and each technical replicate had different collision energy (25, 30, and 40 were used).

MS1 and MS/MS data was analyzed in Compound Discoverer 2.1 software. Initially, LC-MS data were analyzed. Retention time alignment was performed using an adaptive curve with 5 ppm mass accuracy and 2.0 min time shift. All following steps utilized 5 ppm mass accuracy and 0.2 retention time error unless otherwise stated. The workflow nodes were as follows: detect unknown compounds (30% intensity tolerance, 3 S/N ratio, and 1000000 min peak intensity), group unknown compounds, fill gasps (1.5 S/N threshold), normalize areas (constant sum), mark background compounds, predict compositions, differential analysis, and assign compound annotations. In addition to group ratios, the six biological replicates (each with two technical replicates) were also used to calculate biological replicate ratios. Salmonella targets were selected based upon a 2-fold down-regulation of the HumUn/HumSal and MonoSal/HumSal ratio with a p-value less than 0.05. Additionally, both the HumUn/HumSal and MonoSal/HumSal ratios needed to be less than 0.5 in 4 out of the six biological replicate ratios. Candida targets were selected based upon the same workflow but using the HumUn/HumCand and the

MonoCand/HumCand ratios. LC-MS/MS data was analyzed in Compound Discoverer 2.1 software with the same parameters as the LC-MS data, except search mzCloud and search Chemspider nodes were added. The mzCloud search was conducted against all compound classes and activation types with a 20 CE activation energy tolerance and 10 ppm parent and fragment mass tolerance. Identifications in mzCloud were manually verified by checking fragmentation patterns of the top hits. Additional identifications were made by searching the Metlin spectral database with 5 ppm mass accuracy against target molecular weights. MS/MS database spectra were manually compared against experimental spectra. LC-MS normalized peak areas data from Compound Discoverer was imported into Metaboanalyst[33, 34] with log transformation to generate principal component analysis results.

Proteomics Acquisition and Analysis

The protein fractions from the supernatant contents above 30 kDa (aqueous and organic fractions) and the pellets were combined. Ice-cold PBS was used to dissolved protein mixture samples to remove debris from the pellet at a low centrifuge speed (300 x g, 4℃ for 5 mins, three times). The supernatant was then centrifuged at 20,000 x g for 10 min to pellet the bacterial cells and host cells and then lysed with 4% SDS and 6 M urea Lysis buffer with sonication (On 15 sec, Off 30 sec, three cycles). Acetone protein precipitation was used to remove SDS from the protein sample, and the total protein concentration of each pellet was determined by the BCA assay. 100 µg proteins were aliquoted and digested with Trypsin/Lys-C mixture overnight. Then, the digested proteins were desalted, dried down in a speed vac, and saved in the -80℃ until labeling. Twelve-plex isobaric DiLeu reagents were synthesized and used for labeling reaction as previously described.[35] The labeled samples were combined in equal ratios to form pooled 12-plex samples (Labeling map

in Table S1). Excess DiLeu reagents were removed from pooled 12-plex samples via SCX Zip tips (OMIX-SCX, Agilent, Santa Clara, CA) and high pH RPLC fractionation approach was used to improve protein identification coverage. Four fractions were collected from each pooled sample and dried down in a speed vac and saved in the -80℃ until LC-MS/MS analysis.

Fusion Lumos Orbitrap MS was adopted here for protein fraction analysis. Full MS scans were acquired from m/z 300 to 1500 at a resolution of 60 K, AGC at 2E5, and maximum injection time (IT) of 100 ms. The top 20 precursors were then selected for higher-energy C-trap dissociation tandem mass spectrometry (HCD MS2) analysis with an isolation window of 1 m/z, a normalized collision energy (NCE) of 30, a resolving power of 60 K, an AGC target of 5 × 104, a maximum injection time of 100 ms, and a lower mass limit of 110 m/z. Both PEAKS Studio and Proteome Discoverer software packages were used for protein identification and quantification. Trypsin with D&P enzyme was selected for this bottom-up proteomics study, and at most two missed cleavages were allowed. Searches were performed with a precursor mass tolerance of 25 ppm and a fragment mass tolerance of 0.02 Da. The static modifications consisted of DiLeu labels on peptide N-termini (+145.12801 Da) and carbamidomethylation of cysteine residues (+57.02146 Da). The dynamic modifications consisted of DiLeu labels on lysine residues, oxidation of methionine residues (+15.99492 Da). Peptide spectral matches (PSMs) were validated based on 1% FDR and quantitation was performed with a reporter ion integration tolerance of 20 ppm for the most confident centroid. Significantly changed proteins were selected based upon a 2-fold down-regulation of the HumUn/HumSal and MonoSal/HumSal ratio with 1% FDR.

Neuropeptidomics Acquisition and Analysis

The sample preparation, data acquisition, and data analysis of peptide fraction are similar, as previously described. Briefly, Sep-Pak C18 Cartridges were used for peptide desalting. Then peptide samples were dried down in a speed vac and resuspended in optima grade water with 3% acetonitrile and 0.1% formic acid. UPLC-MS/MS analysis was conducted using a Thermo Dionex UltiMateTM 3000 nanoLC system coupled to a Thermo Fusion Lumos Orbitrap MS. Especially, a decision tree-driven MS scheme was used here to allow improved peptide sequencing by alternating between higher-energy collisional dissociation (HCD), electron transfer dissociation (ETD) and electron transfer higher energy dissociation (EThcD) for each specific charge state (Charge 2, combined HCD and EThcD, Charge 3-6 combined HCD and EThcD). The .raw data files from the Orbitrap MS analysis were searched against the mouse neuropeptide database using PEAKS 8.5 software, and a precursor tolerance of 10 ppm and a fragment mass tolerance of 0.02 Da were allowed. Acetylation (N-term), amidation, oxidation (M), pyro-glu from E, pyro-glu from Q, and sulfation (STY), were set as rare dynamic modifications and three maximum variable PTMs were allowed per peptide. Confident peptide identification was threshold with Ascore (PTM site confidence) higher than 20, FDR lower than 1%, and at least one unique peptide was found.

RESULTS

Metabolomics

Label-free metabolomics was performed to investigate small molecule changes by the microbiome in response to both Salmonella and Candida infections. Humanized, salmonella infected mice (HumSal) were compared to humanized, uninfected mice (HumUn) and germ-free mice monocolonized with Salmonella (MonoSal). Similarly, humanized, candida infected mice (HumCand) were compared to humanized, uninfected mice (HumUn) and germ-free mice

monocolonized with Candida (MonoCand). Volcano plots were used to analyze the differences in the data by fold change and p-value. Additionally, potential targets were filtered based upon 4/6 biological replicates having ratios below the 2-fold threshold. The HumSal vs. HumUn and HumSal vs. MonoSal comparisons were analyzed separately with volcano plots as shown in **Figure 1(A)** and **Figure 1(B)**. Then, the overlap in the 2-fold downregulated compounds between two infections was compared with a Venn diagram, resulting in 384 small molecule targets for compounds up-regulated in the humanized, salmonella infected group (**Figure 1(C)**). For the candida infection, the same analysis procedure resulted in 83 targets up-regulated in the humanized, candida infected mice (**Figure 1(D-F)**). The salmonella targets are listed in **Table S2,** and the candida targets are listed in **Table S3**. Of the targets, 13 were in both the Salmonella list and the Candida list.

To investigate trends in an unsupervised manner, principal component analysis (PCA) was utilized. PCA linear transforms the data to display the properties and variance of the sample data along the coordinate axis. **Figure 2** shows the PCA results for the data. In **Figure 2(A)**, the separation of the samples along the principal component axis is shown. The HumSal, MonoSal, and MonoCand groups separate well from the HumUn and HumCand groups, which correlates to the fact that the Candida infection was not as severe as the Salmonella infection.[32] In **Figure 2(B)**, the PCA loadings plots are shown, where *m/z* retention time points are plotted. Points on the outside contribute most to the variation in the data. The points in the bottom right corner are where the compounds up-regulated in the HumSal group are located. **Figure 2(C-E)** provides examples of selected salmonella targets that are localized to this area of the PCA plot. The targets that were selected for both infections are slight to the left of the salmonella only targets, an example of which

is in **Figure 2(F)**. Thus, the target list is composed of compounds that correlate well to the compounds causing the variance in the data as determined by the PCA plot.

LC-MS/MS data, in combination with mzCloud and Metlin spectra libraries, was used for compound identification. **Table 1** provides identifications for compounds detected in this study, with the associated error. The compounds listed in italics were previously identified as being up-regulated in the presence of the microbiome after infection.[32] Interestingly, multiple additional putative annotations were for carnitine molecules. Additional adenosine related compounds were also observed. These compounds may be related to trends in carnitine and adenosine metabolism changes in response to infection when the microbiome is present.

Proteomics

For proteomics analysis, nanoLC-MS/MS with 12-plex DiLeu labeling for relative quantitation was employed.[35] Both PEAKS Studio and Proteome Discoverer software packages were used for data analysis of the 12-plex DiLeu labeled samples, with the two packages generating consistent protein identification results. From the PEAKS analysis, 1318 proteins were identified from 6821 peptide spectral matches (PSMs), while the Proteome Discoverer analysis identified 1228 proteins from 9253PSMs. Up-regulated proteins were selected based upon 1.5 fold changes in the HumSal or HumCand groups compared to their respective controls (HumUn and either MonoSal or MonoCand). While 14 up-regulated proteins were found in HumSal (**Figure 3A & 3B**), none of the proteins were found significantly up-regulated in the HumCand group, which was also correlated to the fact that the Candida infection was not as severe as the Salmonella infection. The gene ontology (GO) analysis of the up-regulated proteins in the HumSal group (**Figure 3C**)

displays the biological process of the up-regulated proteins. Notably, three proteins, Alpha-1-acid glycoprotein, Alpha-2-HS-glycoprotein, and Haptoglobin, were detected, which modulate the activity of the immune system during the acute-phase reaction.

Neuropeptidomics

The aqueous and organic fraction contents between 3 and 30 kDa were analyzed with nanoLC-MS/MS to determine endogenous peptides in the gut produced in response to Salmonella infection. For peptidomics analysis, the data were searched against the neuropeptide database using PEAKS 8.5 software. The fragmention MS spectrum of neuropeptide Fibrinopeptide B shown in **Figure 4A** proved the confidence of the neuropeptide identification. Overall, 38 neuropeptides were identified in the HumSal group, while 28 neuropeptides from the MonoSal group and 7 from the HumUn group. **Figure 4B** compared the neuropeptide identification overlap in HumSal, MonoSal, and HumUn groups, with more unique neuropeptides being identified in HumSal. Also, the identification of neuropeptides in the gut belonging to various neuropeptide preprohormones indicates communication between gut microbiota, brain, and immune system.

DISCUSSION

Metabolites identified that are up-regulated in response to infection and the presence of the model microbiome provide evidence of metabolism changes induced by the microbiome in response to infection. Here, we focused on *m/z* greater than 200 to limit the detection of common primary metabolites and instead focus on the larger secondary metabolites, which could be potential natural products. Previously, the importance of glutathione metabolism against potential oxidative stress was discussed due to the presence of glutathione disulfide and glutathione cysteine disulfide after

Salmonella infection.[32] Multiple carnitine molecules were identified in this study. Acetyl-L-carnitine is involved in energy metabolism; it increases acetyl-CoA uptake in the mitochondria in fatty acid oxidation, increases acetylcholine production, and is a free radical scavenger.[36-38] As acetyl-L carnitine can cross the blood-brain barrier easier compared to carnitine, and consequently its positive effects for various neurological disorders, including depression and Alzheimer's disease, has been observed.[39, 40] Acylcarnitines are involved in β-oxidation through transportation of long-chain fatty acids across the mitochondrial membrane.[41] Studies have focused on various biological and metabolic roles of acylcarnities,[42] including potential pro-inflammatory roles.[42, 43]

The proteomics results also display up-regulated proteins after infection that are involved in the immune response and regulation of the inflammatory response. **Figure 3C** shows among the top five biological processes that the up-regulated proteins involved, four of them were immune or inflammatory response related. In particular, the alpha-1-acid glycoprotein is an acute-phase protein that is involved in immune modulation and inflammatory response.[49] Therefore, proteomics results here prove that the host immune system is enhanced by the human microbiome during infection. Besides, the proteomics result also explains the bodyweight study with this body. While the mice's weight was decreased considerably in the Salmonella groups, the mice's weight did not change too much in the Candida groups. This consistent result proved that the human microbiome could adjust the response according to the situation. They would augment the host immune system when the host health is being threatened, such as severe infection caused by Salmonella.

From the peptidomics analysis, we could see that more neuropeptides were involved during infection. **Figure 4B** showed the presence of 38 neuropeptides in the HumSal group and 28 neuropeptides from the MonoSal group, while only 7 were from the HumUn group. When compared with the two Salmonella infection groups, more unique neuropeptides were identified in the HumSal group. Those 15 unique peptides may carry out neuronal function induced by the human microbiome under infection. To further analyze the biological process of those unique neuropeptides involved, nine of them were immune-related neuropeptides. Notably, a special antimicrobial neuropeptide, cathelin-related antimicrobial peptide, was identified among those unique neuropeptides. The antimicrobial neuropeptide could kill the gram-negative bacteria by attaching to their membrane. This kind of antibiotics is less prone to develop resistance, which could be further investigated for the Salmonella growth inhabition.

The other immune-related neuropeptides could enhance host immune defense in different ways. For example, Fibrinopeptide B is formed by cleavage of fibrinogen by thrombin during the coagulation process.[44] Previous studies have shown fibrinopeptide B involvement in the inflammatory processes.[44] Investigations with sepsis, rheumatoid arthritis, and pilocytic astrocytoma, which are pathogenic inflammatory conditions, found increases in fibrinogen degradation fragments and fibrinopeptide B.[45-47] Thus, Fibrinopeptide B and other immune-related neuropeptide stimulation by the human microbiome during infection could enhance cell-mediated immunity and influence brain and body activities, which facilitates further understanding of the bidirectional communication between gut microbiota, brain, and immune system.

CONCLUSION

Overall, metabolomics, proteomics, and peptidomics analysis reveal microbiome-mediated immune and inflammatory responses in response to Salmonella infection. Three omics analyses were achieved in every single sample. The up-regulated metabolites and proteins, and unique peptides indicate that the 92 microbial strains in the human microbiome enhanced the host immune system and could help the host defense during infection. In particular, the neuropeptidomics analysis provides insights into gut-brain axis communication in response to Salmonella infection to stimulate immune functions. The neuropeptide cathelin-related antimicrobial peptide is a potential novel antibiotics for the treatment of Salmonella infection and could be further evaluated.

ACKNOWLEDGMENTS

This work was supported by T32AI55397 and U19AI109673 from the National Institute of Health, and the Office of the Vice-Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation. This work was supported in part by grants NIH DK108259 (to F.E.R.). L. L. acknowledges a Vilas Distinguished Achievement Professorship, NIH DK071801, and S10RR029531. We would like to thank Kimberley Romano for assistance in experimental mouse procedures.

References

1. Hoffmann, C. et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PLoS One* **8**, e66019 (2013).
2. Scanlan, P.D. & Marchesi, J.R. Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. *Isme j* **2**, 1183-1193 (2008).
3. Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L. & Gordon, J.I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327-336 (2011).
4. Sharon, G. et al. Specialized metabolites from the microbiome in health and disease. *Cell Metab* **20**, 719-730 (2014).
5. Turnbaugh, P.J., Backhed, F., Fulton, L. & Gordon, J.I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213-223 (2008).
6. Rosenbaum, M., Knight, R. & Leibel, R.L. The gut microbiota in human energy homeostasis and obesity. *Trends Endocrinol Metab* **26**, 493-501 (2015).
7. Kostic, A.D., Xavier, R.J. & Gevers, D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* **146**, 1489-1499 (2014).
8. Tang, W.H., Kitai, T. & Hazen, S.L. Gut Microbiota in Cardiovascular Health and Disease. *Circ Res* **120**, 1183-1196 (2017).
9. Jie, Z. et al. The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun* **8**, 845 (2017).
10. Turnbaugh, P.J. et al. A core gut microbiome in obese and lean twins. *Nature* **457**, 480-U487 (2009).

11. Verberkmoes, N.C. et al. Shotgun metaproteomics of the human distal gut microbiota. *Isme Journal* **3**, 179-189 (2009).

12. Wikoff, W.R. et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 3698-3703 (2009).

13. Vogt, N.M. et al. The gut microbiota-derived metabolite trimethylamine N-oxide is elevated in Alzheimer's disease. *Alzheimer's Research & Therapy* **10**, 124 (2018).

14. Brestoff, J.R. & Artis, D. Commensal bacteria at the interface of host metabolism and the immune system. *Nat Immunol* **14**, 676-684 (2013).

15. Clarke, T.B. et al. Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. *Nat Med* **16**, 228-231 (2010).

16. Wilson, M.R., Zha, L. & Balskus, E.P. Natural product discovery from the human microbiome. *J Biol Chem* **292**, 8546-8552 (2017).

17. Wolfender, J.L., Marti, G., Thomas, A. & Bertrand, S. Current approaches and challenges for the metabolite profiling of complex natural extracts. *J Chromatogr A* **1382**, 136-164 (2015).

18. Zheng, X.J. et al. The Footprints of Gut Microbial-Mammalian Co-Metabolism. *Journal of Proteome Research* **10**, 5512-5522 (2011).

19. Lu, K. et al. Arsenic exposure perturbs the gut microbiome and its metabolic profile in mice: an integrated metagenomics and metabolomics analysis. *Environ Health Perspect* **122**, 284-291 (2014).

20. Xiong, W., Brown, C.T., Morowitz, M.J., Banfield, J.F. & Hettich, R.L. Genome-resolved metaproteomic characterization of preterm infant gut microbiota development reveals

species-specific metabolic shifts and variabilities during early life. *Microbiome* **5**, 72 (2017).

21. Erickson, A.R. et al. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS One* **7**, e49138 (2012).
22. Presley, L.L. et al. Host-microbe relationships in inflammatory bowel disease detected by bacterial and metaproteomic analysis of the mucosal-luminal interface. *Inflamm Bowel Dis* **18**, 409-417 (2012).
23. Mondot, S. & Lepage, P. The human gut microbiome and its dysfunctions through the meta-omics prism. *Ann N Y Acad Sci* **1372**, 9-19 (2016).
24. Zhang, Y., Zhao, F., Deng, Y., Zhao, Y. & Ren, H. Metagenomic and metabolomic analysis of the toxic effects of trichloroacetamide-induced gut microbiome and urine metabolome perturbations in mice. *J Proteome Res* **14**, 1752-1761 (2015).
25. Maier, T.V. et al. Impact of Dietary Resistant Starch on the Human Gut Microbiome, Metaproteome, and Metabolome. *MBio* **8** (2017).
26. Clynen, E. et al. Peptidomics of the pars intercerebralis-corpus cardiacum complex of the migratory locust, Locusta migratoria. *Eur J Biochem* **268**, 1929-1939 (2001).
27. Verhaert, P. et al. Matrix-assisted laser desorption/ionization quadrupole time-of-flight mass spectrometry: an elegant tool for peptidomics. *Proteomics* **1**, 118-131 (2001).
28. Schank, J.R., Ryabinin, A.E., Giardino, W.J., Ciccocioppo, R. & Heilig, M. Stress-related neuropeptides and addictive behaviors: beyond the usual suspects. *Neuron* **76**, 192-208 (2012).
29. An, S. et al. A neuropeptide speeds circadian entrainment by reducing intercellular synchrony. *Proc Natl Acad Sci U S A* **110**, E4355-4361 (2013).

30. Dinan, T.G. & Cryan, J.F. The Microbiome-Gut-Brain Axis in Health and Disease. *Gastroenterol Clin North Am* **46**, 77-89 (2017).

31. Holzer, P. & Farzi, A. Neuropeptides and the microbiota-gut-brain axis. *Adv Exp Med Biol* **817**, 195-219 (2014).

32. Bratburd, J.R. et al. Gut Microbial and Metabolic Responses to Salmonella enterica Serovar Typhimurium and Candida albicans. *MBio* **9** (2018).

33. Xia, J., Sinelnikov, I.V., Han, B. & Wishart, D.S. MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res* **43**, W251-257 (2015).

34. Xia, J. & Wishart, D.S. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Curr Protoc Bioinformatics* **55**, 14.10.11-14.10.91 (2016).

35. Frost, D.C., Greer, T. & Li, L. High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Anal Chem* **87**, 1646-1654 (2015).

36. Mansour, H.H. Protective role of carnitine ester against radiation-induced oxidative stress in rats. *Pharmacol Res* **54**, 165-171 (2006).

37. Di Cesare Mannelli, L., Ghelardini, C., Toscano, A., Pacini, A. & Bartolini, A. The neuropathy-protective agent acetyl-L-carnitine activates protein kinase C-gamma and MAPKs in a rat model of neuropathic pain. *Neuroscience* **165**, 1345-1352 (2010).

38. Imperato, A., Ramacci, M.T. & Angelucci, L. Acetyl-L-carnitine enhances acetylcholine release in the striatum and hippocampus of awake freely moving rats. *Neurosci Lett* **107**, 251-255 (1989).

39. Veronese, N. et al. Acetyl-L-Carnitine Supplementation and the Treatment of Depressive Symptoms: A Systematic Review and Meta-Analysis. *Psychosom Med* **80**, 154-159 (2018).

40. Pettegrew, J.W., Levine, J. & McClure, R.J. Acetyl-L-carnitine physical-chemical, metabolic, and therapeutic properties: relevance for its mode of action in Alzheimer's disease and geriatric depression. *Mol Psychiatry* **5**, 616-632 (2000).

41. Rinaldo, P., Matern, D. & Bennett, M.J. Fatty acid oxidation disorders. *Annu Rev Physiol* **64**, 477-502 (2002).

42. McCoin, C.S., Knotts, T.A. & Adams, S.H. Acylcarnitines--old actors auditioning for new roles in metabolic physiology. *Nat Rev Endocrinol* **11**, 617-625 (2015).

43. Adams, S.H. et al. Plasma acylcarnitine profiles suggest incomplete long-chain fatty acid beta-oxidation and altered tricarboxylic acid cycle activity in type 2 diabetic African-American women. *J Nutr* **139**, 1073-1081 (2009).

44. Jennewein, C. et al. Novel aspects of fibrin(ogen) fragments during inflammation. *Mol Med* **17**, 568-573 (2011).

45. Ichikawa, Y., Yamada, C., Horiki, T., Hoshina, Y. & Uchiyama, M. Serum matrix metalloproteinase-3 and fibrin degradation product levels correlate with clinical disease activity in rheumatoid arthritis. *Clin Exp Rheumatol* **16**, 533-540 (1998).

46. Iba, T. et al. Association between the severity of sepsis and the changes in hemostatic molecular markers and vascular endothelial damage markers. *Shock* **23**, 25-29 (2005).

47. Inserra, I. et al. Proteomic study of pilocytic astrocytoma pediatric brain tumor intracystic fluid. *J Proteome Res* **13**, 4594-4606 (2014).

48. King, R. & Tuthill, C. Immune Modulation with Thymosin Alpha 1 Treatment. *Vitam Horm* **102**, 151-178 (2016).

49. Hochepied, T., Berger, F.G., Baumann, H. & Libert, C. Alpha(1)-acid glycoprotein: an acute phase protein with inflammatory and immunomodulating properties. *Cytokine Growth Factor Rev* **14**, 25-34 (2003).

50. Pukhal'skii, A.L. et al. alpha1-acid glycoprotein possesses in vitro pro- and anti-inflammatory activities. *Bull Exp Biol Med* **131**, 479-481 (2001).

TABLES AND FIGURES

Table 1. Putative identifications from metabolomics analysis. Compounds from the previous study are shown in italics.

Molecular Weight	RT (min)	Compound Name	Infection	Literature Molecular Weight	ppm
203.11562	1.219	Acetyl-L-Carnitine	Salmonella	203.11575	-0.65
231.14701	6.577	Butyryl-L-carnitine, Isobutyryl carnitine	Salmonella	231.14706	-0.23
245.16276	12.672	2-Methylbutyroylcarnitine, pivaloylcarnitine	Salmonella	245.16270	0.23
247.14170	1.497/1.724	*hydroxybutyrylcarnitine*	Salmonella	247.14197	-1.08
259.17830	14.895	Hexanoylcarnitine	Salmonella	259.17835	-0.17
267.09656	2.413	Adenosine	Salmonella &Candida	267.09674	-0.68
347.06261	1.452	Adenosine 3’ monophosphate	Salmonella &Candida	347.06308	-1.35
347.06263	1.160	*similar to Adenosine monophosphate*	Salmonella	347.06308	-1.30
348.04670	1.188	*Inosine monophosphate*	Salmonella	348.04709	-1.12
426.08792	0.935	*cysteine glutathione disulfide*	Salmonella	426.08789	0.07
612.15125	1.881	*L-Glutathione oxidized*	Salmonella	612.15198	-1.19

Table 2. Unique neuropeptides in the HumSal group (Immune-related) from under the *Salmonella enterica* Typhimurium infection.

Neuropeptide	**Biology Process**
Fibrinopeptide_B	facilitate the immune response
Signal_transducer_CD24	positive regulation of inflammatory response
Cathelin-related_antimicrobial_peptide	antibacterial humoral response
Resistin-like_beta	response to bacterium
Prolactin-7B1	positive regulation of cell population proliferation
Prolactin-2C3	positive regulation of cell population proliferation
Parathyroid_hormone-related_protein	positive regulation of cell population proliferation
Prolactin-6A1	positive regulation of cell population proliferation
Prolactin-2A1	positive regulation of cell population proliferation

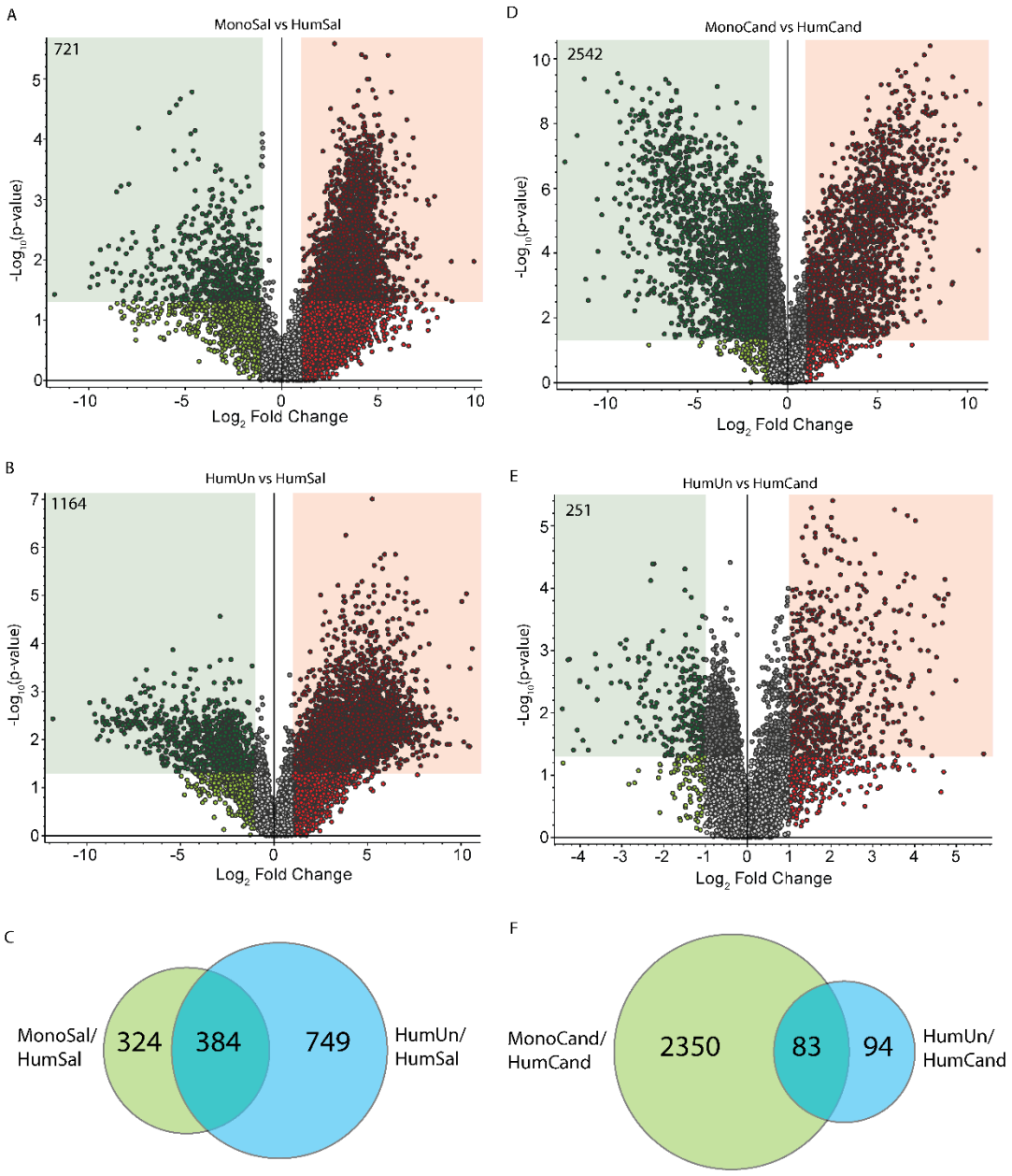


Figure 1. Target selection for metabolomics results. Volcano plots for MonoSal/HumSal (A), HumUn/HumSal (B), MonoCand (D), and HumUn/HumCand (E) selected statistically significant targets in each ratio. The overlap of the targets for the HumSal group (C) and HumCand group (F) were used to select final targets.

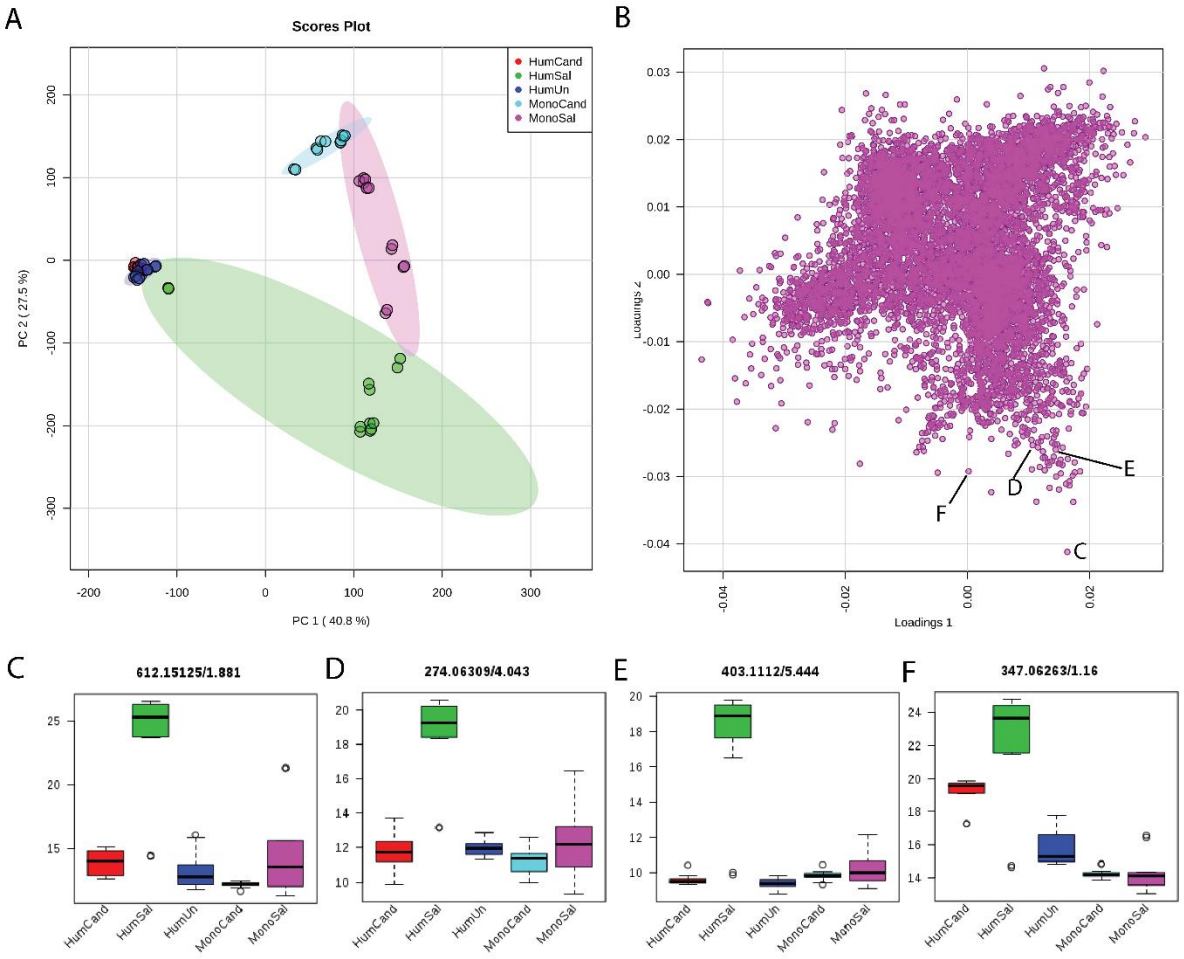


Figure 2. PCA results for the metabolomics data. **(A)** shows the PCA plot for separation of sample groups, while **(B)** shows the loadings plot for the various detected compounds. **(C-F)** show various compounds on the edge of the loadings plot as being up-regulated in the HumSal (and HumCand for **(F)**).

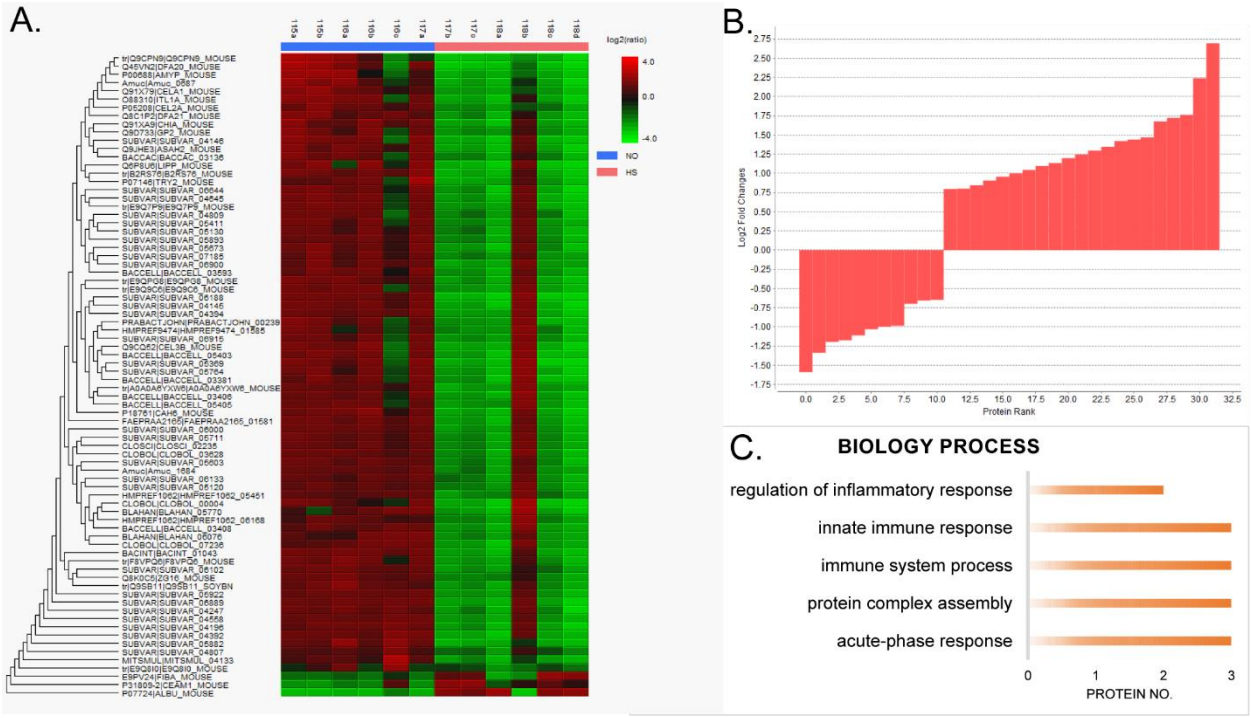


Figure 3. Quantification results for the proteomics data. **(A)** show for HumUn/HumSal; **(B)** shows for MonoSal/HumSal; **(C) shows the** gene ontology results for proteins up-regulated in the HumSal group. Four of the top five biological processes are related to inflammation response or immune response.

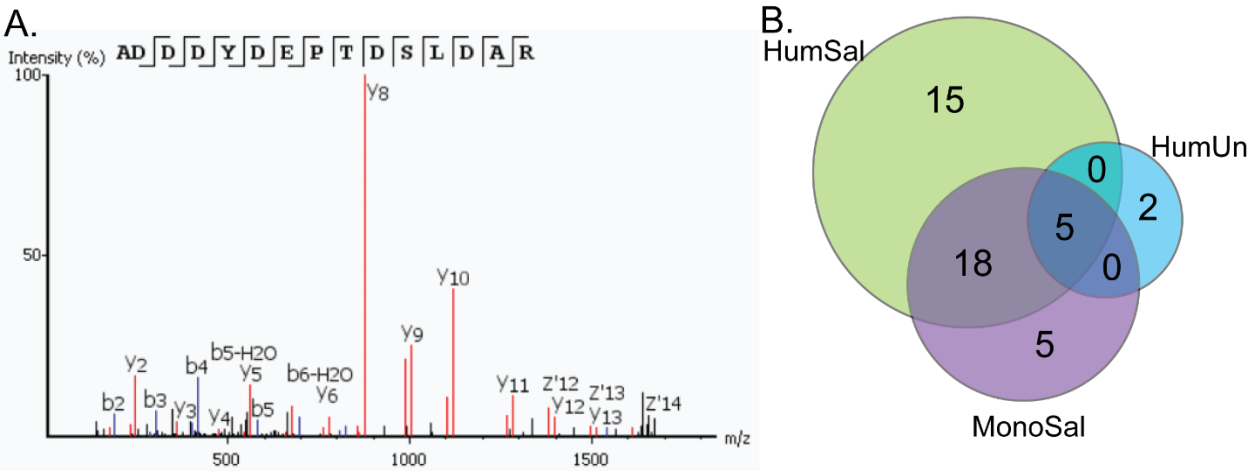


Figure 4. Identification results for the neuropeptidomics data. **(A)** shows the MS spectrum of neuropeptide Fibrinopeptide B; **(B)** compares the neuropeptides identified in HumSal, MonoSal, and HumUn groups.

SUPPLEMENTARY INFORMATION

Table S1. Labeling map for protein samples from HumUn (NO), HumSal (HS), and MonoSal (GS) groups.

Samples	HS1	HS2	HS3	HS4	HS5	HS6	GS1	GS2	GS3	GS4	GS5	GS6
DiLeu Labels	117b	117c	118a	118b	118c	118d	115a	115b	116a	116b	116c	117a
Samples							NO1	NO2	NO3	NO4	NO5	NO6
DiLeu Labels							115a	115b	116a	116b	116c	117a
2 Mixtures	HS&GS		HS&NO									

Table S2. Target Compounds produced in response to Salmonella infection.

Molecular Weight	RT [min]	Log2 Fold Change: (HumUn) / (HumSal)	Log2 Fold Change: (MonoSal) / (HumSal)	P-value: (HumUn) / (HumSal)	P-value: (MonoSal) / (HumSal)
203.11562	1.219	-4.41	-4.45	3.51E-04	7.27E-05
205.04819	0.854	-5.65	-6.98	3.02E-02	1.33E-02
206.01046	28.177	-2.49	-2.36	7.96E-03	4.71E-02
206.09148	18.084	-2.64	-3.45	1.92E-02	1.10E-02
211.13601	22.812	-2.54	-2.45	6.89E-03	4.28E-02
212.04077	0.883	-6.12	-6.97	9.05E-03	2.23E-02
216.09972	14.883	-4.39	-4.01	2.77E-03	2.89E-02
217.03840	0.961	-4.80	-3.23	3.46E-03	1.16E-02
222.08928	19.206	-1.90	-3.54	2.85E-02	1.29E-02
224.10196	15.413	-2.38	-4.03	3.75E-02	1.32E-02
225.00491	28.173	-2.63	-2.37	6.54E-03	4.78E-02
226.19290	22.922	-2.63	-2.97	1.73E-02	1.88E-02
227.00028	28.172	-2.55	-2.57	7.82E-03	4.46E-02
231.14701	6.577	-1.19	-4.85	4.38E-02	4.14E-03
231.88960	0.845	-3.41	-2.73	1.77E-03	3.75E-02
234.16187	22.151	-2.43	-2.32	5.84E-03	4.28E-02
236.96618	0.858	-2.20	-4.23	4.14E-02	3.75E-02
239.02031	0.954	-4.42	-3.86	4.24E-03	1.01E-02
242.11247	12.709	-2.27	-2.73	1.38E-02	3.74E-02
243.21976	21.301	-2.75	-2.31	6.89E-03	4.96E-02
244.12815	10.160	-2.99	-3.29	1.27E-02	2.76E-02
245.16276	12.672	-2.19	-1.99	2.61E-03	2.17E-02
247.14170	1.724	-7.40	-8.32	1.79E-03	6.14E-04
247.14175	1.497	-3.66	-5.57	4.41E-03	1.58E-04
247.86382	0.849	-2.07	-2.59	7.98E-03	4.21E-03
248.97206	8.038	-3.98	-3.86	1.86E-02	4.99E-02
254.15157	19.808	-1.33	-2.40	4.39E-02	2.84E-02
259.17830	14.895	-1.99	-3.95	1.83E-02	5.21E-03
261.03032	1.867	-3.47	-3.68	1.32E-02	4.39E-02
261.15741	14.883	-5.75	-4.15	3.15E-03	4.74E-02
262.01367	1.723	-6.25	-6.74	8.08E-03	8.13E-03
265.06713	1.155	-2.82	-3.84	8.24E-03	3.71E-02
266.16454	22.052	-2.39	-2.15	3.11E-03	3.68E-02
267.09656	2.413	-3.65	-3.55	5.18E-03	3.47E-02
268.12810	16.669	-1.95	-3.18	1.98E-02	2.10E-02
270.13037	0.921	-2.65	-3.29	8.40E-03	1.90E-02
270.14366	15.454	-2.95	-4.42	1.67E-02	1.09E-02
270.14386	14.877	-2.37	-4.31	3.27E-02	1.05E-02
270.90932	28.178	-2.46	-2.55	6.51E-03	2.47E-02
271.21468	22.026	-2.72	-2.61	1.44E-02	4.11E-02
271.25093	22.512	-2.59	-2.32	6.89E-03	3.95E-02
272.90979	28.176	-2.32	-2.36	6.68E-03	4.41E-02
274.06309	4.043	-7.12	-8.07	2.37E-03	2.65E-02
274.12737	1.373	-2.65	-3.23	1.30E-02	1.33E-03
274.13860	12.705	-2.38	-2.93	1.33E-02	3.17E-02
274.25072	23.413	-2.17	-2.27	9.99E-03	3.61E-02
278.15163	22.775	-2.53	-2.40	3.59E-03	1.50E-02
280.13075	21.019	-2.00	-2.56	1.18E-02	3.31E-02
280.24010	22.494	-3.17	-2.94	6.56E-04	1.66E-03
282.14367	18.510	-2.32	-3.95	1.73E-02	1.38E-02
282.14370	18.087	-2.38	-4.11	2.08E-02	1.24E-02
282.14378	16.668	-1.82	-3.34	1.99E-02	2.13E-02
283.32368	22.788	-3.12	-2.27	3.82E-02	3.13E-02
283.32368	22.598	-2.07	-2.16	7.91E-03	4.80E-02
285.11431	1.108	-6.84	-6.50	3.73E-03	4.88E-03
286.13869	15.420	-2.25	-3.41	3.00E-02	6.54E-03
286.88142	28.291	-2.58	-2.31	1.59E-02	4.56E-02
287.24576	21.124	-2.90	-3.21	1.24E-02	4.96E-02
288.88233	28.178	-2.56	-2.43	4.43E-03	4.08E-02
289.12702	1.420	-2.05	-2.46	3.53E-03	2.36E-02

289.99846	8.040	-2.94	-2.95	2.75E-02	3.89E-02
290.88296	28.183	-2.49	-2.44	1.72E-03	2.95E-02
292.08625	21.483	-1.88	-2.62	2.26E-02	2.55E-02
294.16173	22.781	-2.42	-2.23	5.66E-03	4.75E-02
294.16175	22.642	-2.36	-2.33	6.84E-03	4.40E-02
298.15676	21.409	-2.34	-2.59	1.45E-02	4.99E-02
299.28219	23.468	-2.32	-2.25	9.30E-03	4.02E-02
299.87702	0.846	-3.18	-3.10	2.68E-03	2.81E-02
299.89790	0.826	-2.90	-5.96	4.33E-02	1.46E-02
300.15435	15.915	-2.01	-3.59	2.69E-02	1.15E-02
300.15454	15.411	-2.28	-3.89	3.70E-02	1.17E-02
302.17006	14.878	-2.86	-4.78	2.50E-02	9.04E-03
303.88420	28.332	-2.66	-1.89	7.42E-03	3.45E-02
303.91471	28.183	-2.22	-2.35	1.12E-02	3.97E-02
306.09237	0.933	-5.08	-4.87	4.70E-02	1.86E-02
306.16166	22.932	-2.30	-2.06	9.73E-03	3.22E-02
307.09615	1.114	-9.49	-8.75	4.40E-03	1.59E-02
309.09181	1.108	-5.18	-4.09	5.51E-03	5.43E-03
311.22231	22.052	-2.46	-1.77	2.48E-03	3.90E-02
312.13588	22.614	-2.40	-2.35	9.73E-03	3.74E-02
314.17009	16.669	-1.86	-3.28	2.09E-02	2.21E-02
314.92232	28.180	-2.46	-2.34	8.17E-03	4.76E-02
316.26106	20.551	-3.03	-3.17	1.79E-02	4.85E-02
317.25637	22.950	-2.80	-2.57	6.23E-03	4.17E-02
317.27177	20.480	-2.38	-2.64	2.35E-02	3.95E-02
317.87950	28.218	-2.64	-2.58	6.62E-03	4.67E-02
318.16479	15.417	-2.31	-3.86	3.63E-02	1.37E-02
318.16494	14.442	-2.48	-3.47	1.74E-02	2.30E-02
318.27670	23.408	-2.29	-2.32	9.47E-03	3.39E-02
321.28769	22.952	-2.21	-1.90	8.66E-03	4.10E-02
322.10709	1.191	-5.65	-5.09	1.32E-02	1.31E-02
323.05147	1.189	-5.55	-6.15	5.32E-03	2.98E-02
323.05156	0.963	-4.88	-5.04	1.32E-02	4.04E-02
323.07018	1.109	-7.32	-6.43	4.81E-03	6.84E-03
323.20939	22.718	-2.95	-2.89	1.02E-02	3.35E-02
326.29308	23.314	-2.57	-2.27	7.07E-03	3.01E-02
326.91170	28.182	-2.40	-2.37	1.21E-02	4.39E-02
327.90807	28.314	-2.46	-2.31	2.29E-02	4.56E-02
329.90863	28.181	-2.59	-2.39	5.54E-03	4.22E-02
330.16487	16.668	-1.78	-2.92	2.66E-02	2.08E-02
331.06787	1.423	-5.34	-4.10	5.98E-03	2.11E-02
331.14169	21.402	-1.80	-2.70	3.17E-02	3.55E-02
331.90955	28.179	-2.59	-2.48	9.13E-04	2.37E-02
332.18065	15.413	-2.70	-4.19	2.25E-02	1.31E-02
333.90459	28.181	-2.51	-2.43	7.33E-03	4.35E-02
334.11785	22.607	-2.44	-2.41	1.09E-02	3.70E-02
334.15423	21.345	-3.24	-2.46	4.04E-03	3.76E-02
335.30321	23.408	-2.27	-2.30	9.71E-03	3.56E-02
337.14409	21.483	-2.08	-2.32	1.30E-02	3.32E-02
337.24626	20.458	-1.79	-2.89	1.79E-02	2.54E-02
337.90054	28.192	-2.76	-2.56	7.69E-03	4.50E-02
337.90312	28.179	-2.52	-2.41	7.20E-03	4.96E-02
340.30875	23.344	-2.39	-2.18	8.17E-03	4.30E-02
340.91667	28.186	-2.57	-2.39	6.13E-03	4.94E-02
343.30851	21.889	-2.19	-2.88	3.38E-02	4.54E-02
344.07956	18.793	-2.24	-2.48	9.78E-03	3.71E-02
344.18057	18.087	-2.14	-4.01	2.12E-02	1.17E-02
344.18069	16.669	-1.72	-3.26	2.66E-02	1.99E-02
344.18072	17.176	-1.74	-2.80	2.69E-02	2.65E-02
345.03346	0.968	-4.93	-4.41	4.41E-03	3.93E-02
346.11426	0.860	-2.34	-2.89	5.43E-03	3.44E-02
347.06261	1.452	-8.02	-9.37	7.12E-03	6.86E-03
347.06263	1.160	-7.34	-9.80	1.39E-02	1.16E-02
348.04670	1.188	-7.69	-8.04	5.67E-03	8.01E-03
348.16980	22.705	-4.32	-3.16	1.12E-03	4.66E-02
348.25100	21.770	-2.24	-2.87	2.59E-02	4.83E-02

349.28249	22.941	-1.94	-2.41	1.24E-02	6.70E-03
349.31893	22.955	-3.34	-2.90	8.58E-03	4.30E-02
349.31894	23.403	-2.69	-2.20	4.48E-03	2.96E-02
349.31894	23.493	-2.36	-2.03	7.29E-03	3.67E-02
350.08862	1.720	-4.98	-7.52	3.95E-02	5.19E-03
352.11324	23.513	-2.33	-2.18	1.53E-02	4.62E-02
354.23803	22.941	-1.79	-2.21	2.22E-02	7.77E-03
355.21445	22.932	-2.28	-2.31	1.62E-02	2.95E-02
357.23025	21.351	-2.61	-2.17	4.67E-03	3.29E-02
358.19620	18.514	-2.19	-3.70	1.65E-02	1.45E-02
358.19631	18.086	-2.29	-3.97	2.21E-02	1.16E-02
359.17285	21.402	-2.44	-1.81	3.57E-03	3.63E-02
359.33971	22.625	-2.76	-2.77	7.67E-03	2.52E-02
360.96415	8.035	-4.28	-4.36	2.26E-02	4.50E-02
361.31883	20.537	-2.99	-3.53	1.63E-02	3.31E-02
362.19111	15.916	-2.54	-3.97	2.00E-02	1.82E-02
362.19120	16.668	-1.74	-3.24	2.66E-02	2.06E-02
362.21251	22.586	-2.22	-2.07	9.82E-03	4.15E-02
362.30296	23.400	-2.31	-2.25	8.41E-03	3.31E-02
363.05745	1.201	-8.82	-9.04	4.89E-03	1.43E-02
363.05749	1.539	-6.91	-8.56	1.62E-02	1.18E-02
363.33444	23.281	-2.58	-2.22	3.82E-03	3.09E-02
363.33446	22.962	-3.72	-3.61	4.92E-03	3.75E-02
364.04170	1.258	-5.77	-6.66	2.49E-02	2.48E-02
365.05323	1.715	-8.13	-8.45	6.58E-03	6.80E-03
365.27753	21.772	-2.25	-2.88	2.68E-02	4.93E-02
365.31391	22.948	-2.24	-2.06	8.29E-03	3.86E-02
367.86476	0.845	-3.03	-3.41	3.65E-03	2.66E-02
369.28763	23.087	-2.79	-2.95	5.31E-03	3.99E-02
371.24574	22.735	-3.85	-4.07	1.10E-02	3.25E-02
371.24591	22.643	-2.52	-2.23	7.98E-03	2.74E-02
371.35092	23.315	-2.47	-2.30	8.72E-03	4.67E-02
373.31893	20.534	-3.00	-3.50	2.11E-02	4.77E-02
373.31913	21.315	-2.33	-3.12	3.56E-02	3.96E-02
376.20663	18.509	-2.21	-3.83	1.38E-02	1.41E-02
376.20672	18.087	-2.29	-3.98	2.19E-02	1.16E-02
376.20683	16.668	-1.97	-3.38	2.02E-02	2.35E-02
376.22160	23.854	-2.40	-2.14	7.25E-03	4.70E-02
376.28219	22.927	-2.41	-2.93	1.23E-02	1.43E-02
377.31387	22.941	-2.75	-2.87	8.67E-03	1.61E-02
377.35013	23.484	-2.24	-2.63	3.81E-03	2.75E-02
379.32943	23.400	-2.32	-2.26	8.57E-03	3.33E-02
379.32945	22.924	-2.29	-2.19	8.06E-03	4.91E-02
383.24585	23.125	-2.47	-2.31	1.01E-02	4.58E-02
383.24585	23.035	-2.58	-2.16	1.01E-02	4.70E-02
383.24595	22.985	-4.08	-3.64	9.26E-03	7.40E-03
383.30340	22.464	-2.75	-2.92	1.21E-02	3.84E-02
383.83925	0.849	-1.50	-3.04	1.17E-02	2.99E-03
387.20441	22.489	-2.80	-2.94	3.68E-02	2.16E-02
387.20452	22.421	-2.53	-2.65	1.94E-02	3.75E-02
388.20699	18.510	-2.20	-4.10	2.05E-02	1.22E-02
390.22231	18.088	-2.35	-4.02	2.04E-02	1.19E-02
392.07644	0.831	-2.03	-2.75	4.15E-02	3.98E-02
393.30875	22.921	-2.39	-2.94	5.75E-03	8.85E-03
393.30880	21.773	-2.22	-2.37	3.72E-03	1.68E-02
393.34507	23.352	-3.07	-2.61	2.31E-02	4.48E-02
393.34510	23.616	-2.14	-2.68	3.64E-03	2.17E-02
393.34517	23.481	-2.01	-2.45	9.84E-03	2.39E-02
394.17787	14.702	-4.10	-3.58	8.35E-03	4.33E-02
401.25645	21.799	-2.39	-1.94	7.05E-03	4.10E-02
401.25650	21.893	-2.59	-2.60	8.42E-03	4.91E-02
401.25658	21.980	-2.29	-2.17	7.96E-03	3.64E-02
402.29158	22.251	-2.98	-2.89	1.48E-02	1.69E-02
403.11120	5.444	-9.55	-8.93	2.98E-03	5.91E-03
406.04158	1.260	-5.05	-5.19	4.44E-03	4.20E-02
406.21738	18.514	-2.27	-4.05	1.88E-02	1.36E-02

406.21761	17.175	-2.05	-2.95	2.10E-02	2.82E-02
408.17576	24.492	-2.50	-2.40	1.74E-02	4.32E-02
409.30355	20.512	-3.38	-3.73	1.80E-02	4.60E-02
409.33998	22.930	-2.24	-1.99	9.27E-03	4.41E-02
412.20433	18.093	-3.00	-4.60	8.00E-03	1.51E-02
413.27749	20.628	-2.96	-3.34	2.67E-02	4.11E-02
414.20382	21.166	-2.88	-2.78	2.96E-03	1.63E-02
414.20384	20.657	-4.04	-4.66	5.45E-03	1.82E-02
414.20384	21.374	-2.59	-3.99	1.10E-02	2.67E-02
414.20389	21.020	-2.10	-2.53	1.36E-02	4.02E-02
416.30700	22.710	-5.22	-4.39	1.88E-02	2.87E-02
416.30713	22.752	-3.52	-3.53	2.12E-02	2.72E-02
418.23265	20.628	-3.31	-3.10	1.82E-02	4.79E-02
420.23298	18.514	-2.18	-3.85	1.83E-02	1.28E-02
420.30831	22.916	-2.72	-3.04	1.72E-02	2.50E-02
421.27916	23.859	-2.38	-2.29	4.28E-03	4.62E-02
421.33996	22.927	-2.81	-2.91	1.39E-02	3.29E-02
421.37627	23.486	-1.76	-2.78	3.10E-02	3.36E-02
421.37631	23.304	-1.87	-2.25	2.92E-03	4.96E-02
421.37651	23.244	-3.04	-3.72	1.88E-02	1.65E-03
423.35558	23.388	-2.33	-2.32	9.08E-03	3.14E-02
425.84380	0.836	-4.60	-6.55	2.10E-02	2.24E-02
426.08792	0.935	-6.39	-5.64	4.47E-03	4.40E-02
427.32967	22.461	-3.02	-3.44	1.42E-02	3.47E-02
431.23034	21.015	-1.99	-2.44	1.47E-02	3.98E-02
431.23046	21.377	-2.54	-3.98	1.98E-02	3.28E-02
431.23047	20.657	-3.95	-4.66	6.35E-03	1.74E-02
431.24354	20.891	-3.56	-3.87	7.93E-03	3.30E-02
433.16213	22.712	-2.67	-2.68	1.36E-02	3.70E-02
434.13253	21.400	-3.57	-2.84	1.53E-02	3.57E-02
434.13255	21.597	-3.46	-3.15	1.54E-02	2.49E-02
435.85203	0.845	-2.81	-3.60	9.06E-03	2.28E-02
436.18550	21.132	-3.40	-4.17	5.05E-03	3.16E-02
436.18557	21.025	-2.48	-2.61	1.02E-02	4.03E-02
436.18560	21.347	-2.80	-3.83	2.87E-03	1.66E-02
437.33500	22.916	-2.72	-3.04	1.72E-02	2.46E-02
437.33502	21.780	-2.65	-2.82	2.66E-03	7.87E-03
437.37119	23.728	-2.67	-2.49	9.10E-03	4.92E-02
437.37120	23.344	-3.31	-3.24	1.89E-02	2.72E-02
437.37126	23.468	-2.34	-2.77	1.28E-02	3.59E-02
438.37474	23.481	-2.61	-3.08	9.36E-03	2.17E-02
439.40232	22.669	-2.36	-2.28	7.42E-03	4.05E-02
440.10311	1.229	-6.36	-6.50	3.60E-03	3.57E-02
442.29020	22.916	-2.67	-3.01	1.82E-02	2.70E-02
443.07082	1.096	-5.29	-4.72	5.46E-03	9.76E-03
443.30108	22.491	-2.51	-2.30	1.22E-02	4.30E-02
443.32451	22.454	-2.32	-1.98	4.17E-03	4.36E-02
444.28037	16.102	-3.05	-3.06	1.05E-02	3.22E-02
447.28302	23.282	-2.32	-2.41	7.40E-03	4.29E-02
448.14852	22.267	-2.88	-2.59	1.65E-02	3.26E-02
450.24367	18.774	-2.11	-3.87	1.78E-02	1.46E-02
450.24379	18.911	-2.25	-3.81	1.48E-02	1.72E-02
452.15985	21.037	-2.21	-2.26	9.11E-03	4.28E-02
453.36608	22.907	-2.33	-2.32	6.07E-03	2.96E-02
454.18120	14.876	-4.61	-4.91	7.53E-03	3.05E-02
457.34002	23.072	-2.96	-2.38	8.82E-03	4.51E-02
459.26158	20.656	-4.17	-3.74	3.51E-03	2.30E-02
459.26162	21.759	-2.44	-3.67	4.11E-03	3.58E-03
459.26162	21.373	-3.54	-3.36	2.19E-03	2.40E-02
460.26482	21.375	-3.72	-3.25	2.48E-03	2.52E-02
460.26500	20.656	-4.29	-3.81	3.59E-03	2.32E-02
461.24148	21.020	-2.11	-2.82	1.03E-02	3.64E-02
462.12303	23.440	-1.96	-1.97	1.08E-02	3.85E-02
462.16372	22.707	-2.71	-2.85	1.08E-02	2.66E-02
464.33454	22.895	-2.83	-3.13	2.14E-02	3.44E-02
465.36619	22.916	-2.73	-3.27	1.59E-02	4.26E-02

465.40245	23.227	-3.92	-3.71	1.87E-02	3.39E-02
465.40263	23.413	-2.52	-3.43	3.35E-02	4.90E-02
466.30869	28.086	-2.58	-3.21	9.06E-03	1.30E-02
466.30882	28.052	-2.27	-2.35	7.99E-03	3.14E-02
467.38195	23.371	-2.51	-2.31	5.87E-03	4.03E-02
467.43341	23.300	-2.35	-2.30	7.39E-03	4.60E-02
472.09474	28.164	-2.38	-2.38	5.90E-03	4.32E-02
472.28957	16.032	-3.27	-2.37	1.64E-03	6.22E-03
475.25646	21.017	-2.32	-3.03	5.65E-03	3.94E-02
479.14957	23.441	-1.98	-1.96	1.14E-02	3.72E-02
479.15146	24.004	-2.28	-2.68	4.28E-02	3.80E-02
481.36114	22.896	-2.84	-3.13	2.15E-02	3.45E-02
481.36120	21.778	-2.64	-2.80	5.40E-03	1.52E-02
481.39736	23.339	-2.82	-2.63	1.15E-02	4.80E-02
481.39739	23.455	-2.56	-3.02	9.61E-03	2.08E-02
481.39740	23.586	-3.01	-3.02	9.37E-03	2.08E-02
483.10872	1.872	-7.21	-7.76	5.33E-03	2.40E-02
486.09259	24.167	-2.20	-2.68	2.88E-02	3.81E-02
486.31600	22.897	-2.78	-3.15	2.12E-02	3.34E-02
489.38476	28.052	-2.56	-2.42	6.37E-03	4.69E-02
493.83133	0.835	-4.29	-6.69	2.13E-02	1.75E-02
494.28378	1.494	-3.78	-5.51	4.51E-03	3.19E-04
501.36622	23.053	-2.62	-2.67	6.91E-03	4.70E-02
503.83960	0.844	-3.03	-3.81	9.42E-03	1.88E-02
508.36088	22.865	-2.84	-3.06	2.12E-02	3.76E-02
509.39232	22.896	-2.44	-2.60	3.58E-03	2.33E-02
509.42876	23.269	-2.70	-2.88	8.39E-03	3.79E-02
511.40776	22.875	-2.52	-2.37	8.33E-03	4.41E-02
511.40814	23.352	-2.58	-2.55	1.06E-02	4.95E-02
513.30871	21.020	-1.94	-2.45	1.50E-02	4.89E-02
513.35097	20.592	-2.50	-3.65	3.37E-02	4.69E-02
515.30667	14.510	-6.33	-6.89	4.37E-03	7.01E-03
520.31206	13.855	-6.36	-6.55	3.87E-03	3.42E-03
525.38738	21.770	-2.38	-2.52	2.57E-03	2.07E-02
525.38744	22.866	-2.84	-3.06	2.12E-02	3.75E-02
525.42358	23.437	-2.57	-2.62	1.41E-02	2.05E-02
525.42359	23.322	-2.00	-2.51	3.65E-02	3.61E-02
525.42375	23.567	-2.90	-4.83	1.47E-02	2.07E-02
529.32274	14.525	-3.44	-5.81	1.54E-03	3.65E-05
531.43187	27.250	-2.19	-1.81	3.29E-03	4.59E-03
531.43199	26.734	-1.93	-2.22	8.05E-03	4.05E-02
531.43204	27.142	-1.72	-1.80	1.83E-02	4.67E-02
532.37565	22.983	-4.11	-3.18	9.09E-03	1.91E-02
534.19309	2.412	-3.57	-3.38	5.87E-03	3.74E-02
537.32722	24.371	-2.07	-1.86	1.91E-02	4.87E-02
541.27495	13.989	-6.53	-5.29	4.34E-03	3.02E-03
549.11071	5.693	-9.08	-9.51	1.66E-03	2.61E-02
552.38139	28.161	-2.46	-2.72	3.97E-03	1.86E-02
552.38696	22.835	-2.85	-3.00	2.02E-02	3.79E-02
553.16846	23.884	-2.22	-2.09	8.89E-03	4.20E-02
553.45469	23.256	-2.33	-2.70	1.29E-02	3.97E-02
557.37699	20.607	-2.52	-3.71	3.38E-02	4.49E-02
559.46277	27.070	-2.44	-2.47	8.49E-03	4.24E-02
561.79806	0.846	-4.84	-3.89	1.64E-03	3.62E-02
563.12627	12.165	-6.55	-7.71	4.14E-03	3.55E-03
569.41353	22.835	-2.86	-3.00	2.03E-02	3.78E-02
569.44962	23.286	-2.07	-2.61	7.91E-03	1.87E-02
569.44976	23.421	-2.44	-3.00	1.06E-02	1.82E-02
571.82691	0.844	-3.01	-4.62	2.33E-02	1.93E-02
573.28653	1.993	-2.86	-7.42	2.25E-04	6.71E-05
586.15316	12.074	-7.55	-6.54	4.95E-03	4.48E-02
587.10269	2.257	-4.38	-8.55	7.37E-03	7.59E-04
587.13751	12.300	-7.68	-7.54	5.28E-03	4.13E-02
589.11674	10.136	-9.82	-8.17	1.74E-03	4.96E-02
589.41852	22.991	-2.63	-2.53	7.93E-03	4.81E-02
590.26534	1.221	-2.82	-3.88	3.42E-03	1.89E-03

596.26293	22.856	-2.90	-2.95	1.03E-02	4.21E-02
596.26341	22.693	-2.45	-2.53	1.13E-02	4.68E-02
597.23977	12.158	-2.40	-3.04	5.49E-04	1.05E-03
599.46011	23.308	-2.56	-2.27	7.68E-03	4.52E-02
600.30906	15.411	-2.26	-3.86	3.65E-02	1.12E-02
601.40315	20.619	-2.52	-3.69	3.23E-02	4.24E-02
603.13231	12.160	-7.42	-7.60	4.09E-03	1.61E-02
612.15104	2.017	-8.64	-9.90	3.69E-03	2.86E-02
612.15125	1.881	-11.80	-11.75	3.69E-03	3.79E-02
613.43969	22.805	-2.87	-2.97	2.00E-02	4.05E-02
613.47567	23.290	-2.49	-2.48	3.08E-03	2.17E-02
626.15955	12.291	-8.09	-6.76	5.07E-03	4.98E-02
627.18689	24.470	-2.16	-1.98	8.78E-03	4.56E-02
628.09287	3.234	-6.79	-7.34	1.95E-02	2.44E-02
629.78451	0.844	-4.73	-4.23	2.86E-03	3.52E-02
633.44475	22.956	-2.86	-2.62	8.81E-03	4.28E-02
635.37526	13.849	-9.14	-8.72	3.94E-03	2.33E-02
637.29439	20.125	-1.63	-2.21	5.85E-03	3.93E-02
639.81335	0.842	-2.88	-5.52	4.87E-02	2.16E-02
642.15453	12.122	-8.37	-7.37	4.66E-03	3.64E-02
645.42931	20.631	-2.68	-3.67	2.93E-02	4.10E-02
652.10374	5.751	-9.29	-9.82	8.47E-03	1.97E-02
657.46570	22.774	-2.83	-2.95	2.00E-02	3.92E-02
657.50174	23.285	-1.98	-2.40	6.35E-03	1.62E-02
660.31836	2.761	-7.37	-7.92	5.62E-04	5.61E-04
664.09245	1.275	-5.36	-5.22	4.31E-03	7.32E-03
665.32553	20.123	-2.64	-1.93	3.58E-02	4.74E-02
666.06998	1.888	-5.78	-6.52	4.82E-03	4.50E-03
668.09826	5.748	-7.83	-8.32	1.40E-02	1.70E-02
677.47048	22.928	-2.89	-2.82	1.09E-02	4.40E-02
691.82752	0.835	-4.07	-5.03	3.85E-02	3.96E-02
697.77088	0.844	-4.35	-4.54	5.36E-03	2.93E-02
701.49201	22.744	-2.75	-2.90	2.07E-02	4.27E-02
701.52782	23.226	-1.96	-1.92	1.02E-02	1.23E-02
707.50340	27.201	-2.78	-3.45	1.12E-02	1.62E-02
719.36329	13.108	-4.81	-5.59	1.21E-02	2.07E-02
721.49632	22.894	-2.89	-2.82	8.95E-03	3.73E-02
726.37009	13.401	-7.69	-7.23	4.18E-03	1.90E-02
727.39695	12.381	-3.99	-3.34	8.06E-03	2.35E-02
734.40276	20.539	-2.58	-4.57	2.40E-02	1.16E-02
759.81475	0.833	-4.12	-5.60	4.66E-02	4.73E-02
794.42888	13.703	-6.96	-6.56	5.51E-03	1.28E-02
803.39223	12.162	-3.52	-4.23	3.43E-02	2.56E-02
811.40724	11.820	-2.97	-3.90	6.80E-03	2.66E-02
850.38994	21.025	-1.61	-2.93	4.16E-02	2.75E-02
870.23685	1.890	-5.25	-5.58	8.44E-03	4.87E-03
873.18136	1.872	-7.78	-7.78	4.96E-03	1.79E-02
877.17912	11.247	-7.59	-7.55	2.56E-03	4.63E-02
878.27300	24.374	-2.60	-2.36	1.54E-02	4.63E-02
879.14672	11.764	-7.86	-7.28	5.72E-03	4.54E-02
888.56171	16.100	-3.08	-3.00	1.13E-02	4.68E-02
894.15731	11.400	-8.35	-8.23	4.67E-03	3.23E-02
1039.59072	14.691	-6.49	-5.86	5.77E-03	1.97E-02
1143.65445	0.843	-3.04	-4.62	2.42E-02	1.93E-02
1192.65652	14.069	-2.45	-3.16	1.38E-02	3.33E-02
1259.56968	0.843	-4.76	-4.61	2.76E-03	3.33E-02
1279.62761	0.841	-2.86	-5.42	4.26E-02	1.93E-02
1690.86495	15.055	-7.79	-7.77	3.97E-03	6.40E-03
1795.82207	14.031	-8.26	-8.24	4.36E-03	4.68E-03
1908.91261	14.475	-6.62	-7.30	1.03E-02	2.64E-02

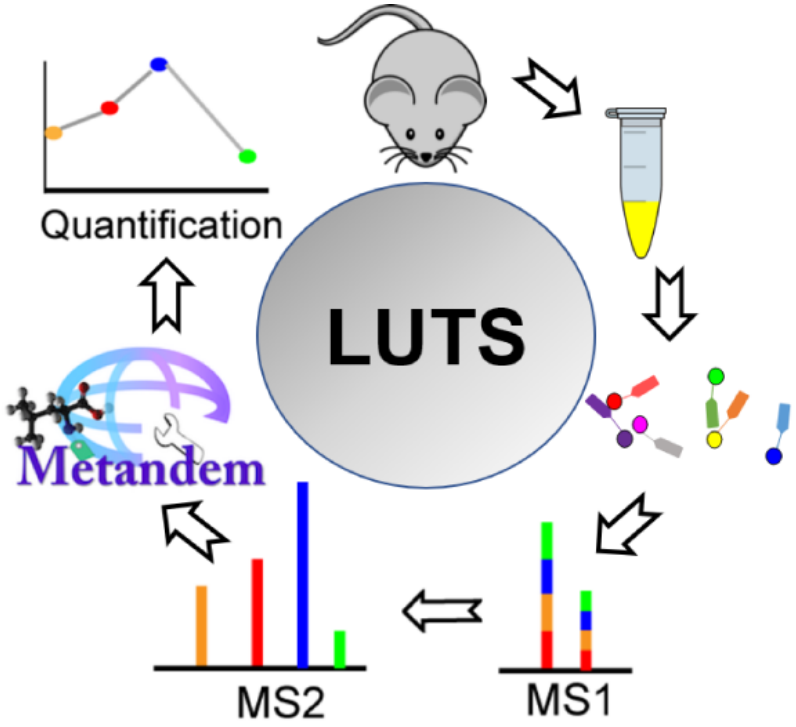
Table S3. Target Compounds produced in response to Candida infection.

Molecular Weight	RT [min]	Log2 Fold Change: (MonoCand) / (HumCand)	Log2 Fold Change: (HumUn) / (HumCand)	P-value: (MonoCand) / (HumCand)	P-value: (HumUn) / (HumCand)
201.07486	0.975	-1.09	-1.04	3.71E-05	1.39E-03
205.07728	1.701	-2.71	-1.24	1.58E-04	7.52E-03
211.09564	1.649	-1.17	-1.51	9.99E-03	3.31E-03
211.09570	2.585	-1.19	-1.29	1.31E-02	3.70E-02
243.11071	13.181	-1.31	-1.33	4.33E-03	1.57E-02
248.10457	16.250	-5.65	-1.26	2.32E-06	1.77E-03
251.10170	2.650	-2.75	-1.82	3.73E-05	1.38E-03
258.09629	1.041	-1.46	-1.20	2.71E-05	1.02E-03
259.07598	0.908	-1.95	-2.41	6.63E-03	2.47E-03
265.64015	11.918	-1.67	-1.56	2.25E-02	2.78E-02
267.09656	2.413	-4.55	-2.23	4.09E-04	4.14E-05
268.05177	9.113	-8.56	-1.71	3.27E-07	1.13E-02
271.11657	1.529	-2.65	-1.41	3.54E-04	2.84E-02
272.11193	1.546	-2.11	-1.42	7.92E-06	5.94E-03
272.13700	1.577	-2.26	-1.62	9.87E-04	6.64E-03
275.08271	13.147	-1.92	-1.62	3.32E-03	1.66E-02
275.08282	12.026	-1.96	-1.49	1.96E-03	1.53E-02
285.23016	16.114	-2.81	-1.85	9.23E-05	4.78E-02
297.19393	18.697	-2.31	-2.30	6.15E-03	2.15E-02
301.12715	1.318	-1.30	-1.43	1.23E-02	1.57E-02
301.12718	1.771	-3.33	-1.13	2.44E-03	2.94E-03
324.17987	11.899	-7.21	-2.59	6.40E-06	1.30E-03
325.18873	16.306	-4.06	-1.15	5.19E-07	4.46E-02
330.11748	1.154	-1.30	-1.15	8.03E-04	3.48E-03
331.06787	1.423	-3.83	-4.41	3.97E-03	8.79E-03
343.17429	13.260	-2.91	-1.29	2.61E-03	7.21E-03
344.07956	18.793	-1.21	-1.51	3.99E-03	1.05E-03
347.06263	1.160	-5.43	-3.59	1.98E-04	1.99E-03
361.31891	20.303	-7.44	-1.09	1.65E-05	4.79E-02
362.15090	12.765	-2.31	-1.15	3.45E-03	9.42E-03
363.05745	1.201	-2.10	-2.45	1.45E-02	5.23E-03
363.05749	1.539	-2.47	-2.55	2.43E-03	3.93E-03
374.17233	15.957	-5.01	-1.54	3.02E-05	9.40E-03
377.35012	23.394	-1.17	-1.43	3.07E-02	9.74E-03
383.26684	18.042	-2.67	-1.56	3.86E-03	4.92E-02
384.18942	1.861	-5.29	-1.48	1.83E-05	1.10E-04
400.10655	12.780	-5.28	-2.46	1.02E-05	1.32E-03
402.29149	22.011	-3.07	-1.83	3.55E-02	8.62E-03
409.20930	15.959	-5.78	-1.55	8.63E-06	2.75E-03
409.84062	0.846	-2.90	-2.29	9.37E-05	6.72E-03
414.12236	13.887	-5.12	-3.02	3.17E-05	7.86E-04
422.19370	15.731	-8.38	-1.26	1.90E-06	7.66E-03
434.13253	21.400	-1.26	-1.81	1.19E-02	4.19E-02
436.18077	1.044	-2.60	-1.91	1.27E-04	1.77E-03
439.22054	15.722	-12.37	-1.21	1.53E-07	1.12E-02
444.22551	15.979	-7.87	-1.34	3.32E-06	2.04E-03
457.25758	14.152	-9.28	-1.60	2.77E-07	2.74E-03
458.24107	16.480	-3.02	-1.66	8.24E-03	1.80E-02
459.27311	15.826	-6.91	-1.26	2.29E-06	1.73E-02
466.20760	15.980	-8.53	-1.32	2.03E-06	9.55E-04
477.82814	0.847	-3.06	-2.52	4.74E-05	5.79E-03
487.23176	18.208	-6.26	-1.48	4.29E-05	6.32E-03
487.26795	14.316	-5.60	-1.50	4.11E-06	5.49E-03
489.28332	15.982	-8.74	-1.53	3.67E-06	1.98E-03
491.26225	19.414	-3.55	-1.25	5.09E-03	3.86E-02
492.13306	14.741	-4.65	-1.88	6.35E-04	4.41E-02
514.32266	11.735	-1.53	-2.29	7.79E-03	7.65E-05
531.27977	11.920	-1.99	-1.94	2.01E-02	4.72E-02
534.19309	2.412	-4.31	-2.25	2.32E-04	4.07E-05
544.27427	1.565	-2.25	-1.62	8.25E-04	6.79E-03

568.28570	12.585	-6.25	-2.54	3.07E-06	3.09E-03
569.44962	23.286	-2.16	-1.86	4.00E-02	4.74E-03
576.28608	2.273	-4.19	-4.05	1.95E-02	1.93E-02
599.46011	23.308	-1.36	-1.61	2.33E-02	1.31E-03
613.47567	23.290	-1.65	-1.53	4.50E-03	8.56E-03
657.50174	23.285	-2.06	-1.95	1.54E-02	3.50E-03
671.39622	13.210	-1.12	-1.24	1.60E-02	2.98E-03
701.52782	23.226	-2.23	-3.03	2.77E-02	3.24E-03
729.15936	12.708	-7.37	-1.79	6.83E-07	4.80E-02
743.17508	13.423	-5.52	-2.43	3.36E-05	2.83E-02
787.41922	13.610	-1.37	-1.13	1.44E-04	5.67E-03
789.33925	8.794	-6.24	-1.53	9.69E-04	1.76E-02
852.42355	13.088	-3.38	-1.65	4.87E-04	1.89E-02
887.48635	14.818	-3.99	-2.87	6.11E-05	3.43E-03
900.49216	13.681	-2.09	-2.19	1.05E-02	1.18E-02
906.44739	12.381	-1.74	-1.13	1.04E-03	8.45E-03
910.43236	15.981	-4.70	-2.88	1.32E-03	6.83E-04
911.47117	12.548	-5.18	-3.93	3.81E-05	2.81E-02
939.50251	14.865	-1.94	-1.37	9.25E-04	1.63E-02
1028.50089	15.029	-2.79	-2.87	1.79E-02	1.49E-02
1040.59812	14.478	-3.96	-3.63	1.02E-03	1.15E-03
1173.64252	15.720	-3.15	-4.24	4.19E-02	1.40E-03
1193.34153	9.749	-10.04	-1.29	1.81E-07	1.99E-02

Chapter 8

Urinary Biomarker Investigation of Lower Urinary Tract Symptoms with Mass Spectrometry



Written in collaboration with the Wisconsin Initiative for Science Literacy to communicate this thesis research to non-specialists.

Key Words: lower urinary tract symptoms (LUST), mass spectrometry, urine, metabolite, protein, biomarker, DiLeu labeling, Hormone, Inflammation.

First of all, I would like to thank the Wisconsin Initiative for Science Literacy (WISL) at the University of Wisconsin-Madison for sponsoring and supporting the creation of this chapter. I also really appreciate the guidance and editorial support provided by the science communications program of WISL. It is a great opportunity for me and many others to explain our research and share science enthusiasm with the non-scientific audience. Hopefully, this chapter could help them to understand the biomarker discovery process and inspire the next generation scientists.

8.1 What is LUST?

Lower urinary tract symptoms (LUTS) is a disease that the patients either have trouble urinating or they are urinating too often (1). It is prevalent in men and affects 50% to 90% of men aged 50 years or older (2). Recent studies have shown that many reasons may cause LUTS, and age-related hormone change and prostatic inflammation are two plausible etiologies (3, 4). While patients may suffer a variety of etiologic mechanisms, the current subjective symptom score diagnosis method usually could not distinguish the pathogenic factors from different individuals. As a result, patients are treated empirically with either medication or surgery, which can be ineffective and bothersome. Therefore, objective biomarkers associated with specific etiologic mechanisms are greatly needed to refine the diagnostic approach and provide personalized treatment.

8.2 What is Biomarker?

Biomarkers can be specific cells, molecules, or genes, gene products, enzymes, or hormones. These large or small molecules present in biological systems and involved in the diseased development, which could be used as an indicator of diseased processes are defined as a biomarker (5). Although the term biomarker is relatively new, biomarkers have been used in pre-clinical

research and clinical diagnosis for a considerable time. For example, sweat chloride may be used as a diagnostic biomarker to confirm cystic fibrosis (6). Since age-related hormone change and prostatic inflammation are two etiologies of LUTS, it is plausible that urinary metabolite and protein biomarkers could be identified and used to diagnose hormone-induced LUTS and inflammation-induced LUTS. Therefore, our research goal is to investigate the urinary metabolite and protein biomarkers, which could be used as a non-invasive diagnose method of LUST to achieve individual treatment.

8.3 What method did I use?

8.31 Mouse Model

To investigate the LUTS biomarkers according to the etiology and to further achieve the individual treatment, the first step is to group the patients according to the etiology. However, the complex and variegated composition of the LUTS patient population makes it challenging to validate the contributions of different mechanisms. Since many reasons could lead to LUTS, both the patients and doctors could not tell how the LUTS were caused, do not even mention to group them. But a mouse model could be used to solve a problem like this. Since in the lab condition, it is much easier to control the etiology. We have successfully built two LUTS models with mice, one is primarily caused by the hormone changed, and the other one is caused by the prostatic inflammation. Both have been proved to be justified mouse model for LUTS study (7, 8).

The hormone change model is to mimic the etiology of LUTS, which is causing by the aging change. As we move through the stages of life, our hormone levels fluctuate. For example, a woman’s menopause is caused by the decrease releasing of estrogen and progesterone. In our case,

to mimic the hormone changing in men caused by the age, testosterone (25 mg) and estradiol (2.5 mg + 22.5 mg cholesterol binder) were slow releases in adult male C57/BL6 mice. The generate marked increases in bladder volume, bladder mass, and prostate mass are the same symptoms as the LUST patients (8). The mouse urine was collected before hormone treatment and 2 weeks, 4 weeks, and 8 weeks after hormone treatment. Then the mouse urine samples were store in -80 refrigerator before the analysis. The low temperature could avoid the degrading of some unstable metabolites.

For the prostatic inflammation mouse model, Ecoli, a kind of bacterial, was used to treat the Bladder of C57BL/6J mice to cause the inflammation in the prostate, which mimics the LUTS, which cause by the inflammation (7). The control group was treated with phosphate-buffered saline (PBS), which could not cause any inflammation. Urine samples were collected from both prostatic inflammation group and control group mice seven days after treatment and store in -80 until analysis.

8.32 Mass Spectrometry (MS)

After we get the mouse models and urine samples, the next step is looking for an unbiased tool to identify and quantify all the urinary metabolites and proteins to find the candidate biomarker. Mass spectrometry (MS) is a sensitive tool like this, which can detect a wide variety of metabolites and proteins present in a complex system (9). **Fig. 1A** shows the general set-up of a tandem mass spectrometer. After injecting a sample into the ion source, some molecules like metabolites and proteins will be charged and separated in MS1 according to their mass to charge ratio (m/z), which

called precursor ion. Since there may be two molecules with the same mass and charge, such as a peptide sequence with ADFG, although its molecule weight and charge state are the same with GFAD, they are different peptides. Therefore, to further confirm the structure of the molecule, the precursor ion will be broken into pieces (product ion) and separated again in MS2, then detected in the monitors. If the pieces we detected are A**DF** and **DF**G, then we could drive the structure of this molecule is ADFG instead of GFAD, just like solve a puzzle. After the separation in MS1 and MS2, the molecule could be identified confidently and the intensity of the detected report ions will be further used as the quantification.

8.33 12-plex DiLeu tags

Our 12-plex DiLeu tags are designed base on the theory of tandem MS, which has been proved to be high throughput and accurate quantification method for proteome study (10). In this thesis, it is the first time applied to the identification and quantification of the metabolite. **Fig. 1B** shows the basic structure of DiLeu tags, and it consists of three parts, amine-reactive group (A), balance group (B), and reporter group (C). Part A, which is the same in all the twelve tags, could be reacted with the amine group in the peptides or the metabolites. For part B and part C, heavy elements and light elements were both involved. Such as the two isomers of carbon ^{12}C and ^{13}C, ^{12}C is the light element, and ^{13}C is the heavy element. By adjusting the distribution of heavy and light elements, although the molecule weight of part C is different in all the twelves DiLeu tags, the total molecule weight of each DiLeu tags is the same because the complemental from the balance group. For example, tag 115 (part A=same, part B=31, part C=115) and tag 118 (part A=same, part B=28, part C=118) have the same molecule weight but different reporter group (C).

12-plex DiLeu tags were used to label twelve urine samples (Labeling map, **Fig. 3D)** and mixed the labeling samples in the same ratio. **Fig. 1C** shows an example of how DiLeu tags reacted with metabolites. Then the mixture was injected into the MS. During the first separation in MS1, the same metabolite from twelve samples would stay together since they all have the sample mass and charge state. After they came out from the MS1, their structure would be broken into pieces, and part B and part C were also broken apart. Part C, which could represent the metabolite from each sample, could be detected, and their intensity could be used for the measurement of the metabolite concentration in each sample.

8.34 Metandem

Except for the analysis instrument, the software is also import during the data analysis. To date, few metabolomics software tools can process stable isotope label-based metabolomics data, particularly when using reporter ions produced by MS/MS for quantification. Therefore, our lab developed a novel online software tool for isobaric labeling-based metabolomics, named Metandem (11), which integrates metabolite quantification, identification, and statistical analysis in the same software package and is freely available at http://metandem.com/web/. Metandem is also the first omics data analysis software to provide straightforward, online parameter optimization functionality for customization to a particular dataset (**Fig. 2**).

8.4 What have I found?

8.41 Hormone-induced mouse model

12-plex DiLeu isobaric labeling with Metandem data processing presents an accessible and efficient workflow for amine-containing metabolome study in biological specimens. Twelve-plex DiLeu isobaric labeling allowed multiplexed metabolite identification and quantification in mouse urine samples while also reducing instrumentation time demand, decreasing run-to-run variation, and also improving quantification accuracy. The same labeled metabolite from 12 urine samples showed a single peak in the MS^1 spectrum with a mass shift of 145.1280 Da (**Fig. 3A**). For this peak, twelve distinct reporter ion peaks are present in the MS^2 low *m/z* region (**Fig. 3B**). The intensity of each reporter ion in MS^2 spectra reflects the labeled metabolites' abundance in each sample and, thus, can be compared for relative quantification (**Fig. 3C**).

A total of 59 amine-containing metabolites were identified and quantified, 9 of which were changed significantly by the hormone treatment **Fig. 4**. Metabolic pathway analysis is based on the association between identified metabolites and their related biological processes. Herein, all identified metabolites were input into the MetaboAnalyte 4.0 software for metabolic pathway analysis. Three potentially perturbed metabolic pathways were identified as **Table. 1**: (1) the arginine and proline metabolism pathway; (2) the aminoacyl-tRNA biosynthesis pathway; and (3) the tryptophan metabolism pathway. Among them, the arginine and proline metabolism pathway was significantly dysregulated both in this model and in a prior analysis of LUTS patient samples (12). Proline and citrulline were significantly changed in both and present as attractive candidate biomarkers as **Fig. 8**.

8.42 Prostatic-inflammation mouse model

In the hormone-induced mouse model, we investigated the influence of hormone changed on the urinary metabolites. In the prostatic-inflammation mouse model, we went a step further. We used the 3K molecular cut-off filter (MWCO) to separate the urine samples into metabolite fraction and protein fraction and investigated both the urinary protein and metabolite-based biomarkers. Notably, the investigation of isobaric dimethylated leucine (DiLeu) labeling on metabolites allowed metabolomics and proteomics analysis on the same liquid chromatography (LC)-MS platform. In total, 143 amine-containing metabolites and 1058 urinary proteins were identified and quantified; among them, 14 metabolites (**Fig. 5**) and 168 proteins (**Fig. 6**) were significantly changed by prostatic inflammation. Five metabolic pathways and four inflammation-related biological processes (**Fig. 7**) were potentially disrupted. By comparing our findings with urinary biomarkers identified in a mouse model of genetic-induced prostate inflammation (13, 14) and with those previously found to be associated with LUTS in older men (12, 15), we identified creatine, haptoglobin, immunoglobulin kappa constant and polymeric Ig receptor as conserved biomarkers for prostatic inflammation associated with LUTS (**Fig. 8**). The combined metabolomics and proteomics analyses provided an informative method for the study of molecular disease mechanisms, which contributed to biomarker discovery and patient stratification, according to etiologies.

8.5 What is the significance of my research?

For the method significance part, twelve-plex DiLeu isobaric labeling is an attractive high throughput strategy for identification and quantification of amine-containing metabolites, and Metandem is a useful tool for large-scale stable isotope labeling-based metabolomics data analysis. Paired together, these tools offer a powerful and accessible method for relative quantification of

amine-containing metabolites in disease biomarker research. Besides, the investigation of DiLeu labeling on metabolites allowed metabolomics and proteomics analysis on the same LC-MS platform.

From the biological significance part, the arginine and proline metabolism pathway was the critical pathway disrupted during the LUTS development, and candidate biomarkers involved in this pathway were found from both hormone-induced mice model and prostatic-inflammation mouse model. Especially, different biomarkers were detected in two mouse models (**Fig. 8**), which could be further used as the LUTS patient stratification and individual treatment. Furthermore, urine samples from voluntary voiding, meaning a non-invasive method for sample collection, have been proved to be available for the future LUTS diagnosis. Future studies will focus on clinical validation of the metabolite and protein biomarkers found in this thesis with a separate large cohort of patient samples.

Reference

1. YAMANISHI T. 2004. Lower Urinary Tract Symptoms (LUTS) in Middle-Aged and Elderly Men Tomonori YAMANISHIJournal of the Japan Medical Association.

2. Hale N, Choi K, Lohri J. 2014. Primary care evaluation and treatment of men with lower urinary tract symptoms. J Am Osteopath Assoc. American Osteopathic Association.

3. Noguchi N, Chan L, Cumming RG, Blyth FM, Handelsman DJ, Seibel MJ, Waite LM, Le Couteur DG, Naganathan V. 2016. Lower Urinary Tract Symptoms and Incident Falls in

Community Dwelling Older Men: The Concord Health and Ageing in Men Project. J Urol 196:1694–1698.

4. Nickel JC, Roehrborn CG, Castro-Santamaria R, Freedland SJ, Moreira DM. 2016. Chronic Prostate Inflammation is Associated with Severity and Progression of Benign Prostatic Hyperplasia, Lower Urinary Tract Symptoms and Risk of Acute Urinary Retention. J Urol 196:1493–1498.

5. Strimbu K, Tavel JA. 2010. What are Biomarkers? NCBI.

6. Farrell PM, Rosenstein BJ, White TB, Accurso FJ, Castellani C, Cutting GR, Durie PR, LeGrys VA, Massie J, Parad RB, Rock MJ, Campbell PW. 2008. Guidelines for Diagnosis of Cystic Fibrosis in Newborns through Older Adults: Cystic Fibrosis Foundation Consensus Report. J Pediatr 153:S4.

7. Bushman WA, Jerde TJ. 2016. The role of prostate inflammation and fibrosis in lower urinary tract symptoms. Am J Physiol Physiol 311:F817–F821.

8. Nicholson TM, Ricke EA, Marker PC, Miano JM, Mayer RD, Timms BG, Vom Saal FS, Wood RW, Ricke WA. 2012. Testosterone and 17β-estradiol induce glandular prostatic growth, bladder outlet obstruction, and voiding dysfunction in male mice. Endocrinology 153:5556–5565.

8. Li L, Sweedler J V. 2008. Peptides in the Brain: Mass Spectrometry–Based Measurement Approaches and Challenges. Annu Rev Anal Chem 1:451–483.

10. Frost DC, Greer T, Li L. 2015. High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. Anal Chem 87:1646–1654.

11. Hao L, Zhu Y, Wei P, Johnson J, Buchberger A, Frost D, Kao WJ, Li L. 2018. Metandem: An online software tool for mass spectrometry-based isobaric labeling metabolomics. Anal Chim Acta 1088:99–106.

12. Hao L, Greer T, Page D, Shi Y, Vezina CM, Macoska JA, Marker PC, Bjorling DE, Bushman W, Ricke WA, Li L. 2016. In-Depth Characterization and Validation of Human Urine Metabolomes Reveal Novel Metabolic Signatures of Lower Urinary Tract Symptoms. Sci Rep 6:30868.

13. Hao L, Thomas S, Greer T, Vezina CM, Bajpai S, Ashok A, De Marzo AM, Bieberich CJ, Li L, Ricke WA. 2018. Quantitative proteomic analysis of a genetically induced prostate inflammation mouse model via custom 4-plex DiLeu isobaric labeling. Am J Physiol - Ren Physiol 316:F1236–F1243.

14. Hao L, Shi Y, Thomas S, Vezina CM, Bajpai S, Ashok A, Bieberich CJ, Ricke WA, Li L. 2018. Comprehensive urinary metabolomic characterization of a genetically induced mouse model of prostatic inflammation. Int J Mass Spectrom 434:185–192.

15. Greer T, Hao L, Nechyporenko A, Lee S, Vezina CM, Ricke WA, Marker PC, Bjorling DE, Bushman W, Li L. 2015. Custom 4-Plex DiLeu Isobaric Labels Enable Relative Quantification of Urinary Proteins in Men with Lower Urinary Tract Symptoms (LUTS). PLoS One 10:e0135415.

TABLES AND FIGURES

Table 1. Potentially disrupted metabolic pathways via MetaboAnalyte 4.0 and KEGG, metabolites with red bold are significantly changed metabolites (*p*-value is from Fisher's exact test).

Metabolic Pathway	KEGG ID	Matched Metabolites	*p*-value
Arginine and proline metabolism	Map00330	citrulline, N-acetylputrescine, proline, glutamic acid, creatine, GABA, 4-aminobutyraldehyde	4.90E-05
Aminoacyl-tRNA biosynthesis	Map00970	leucine, proline, cysteine, glycine, alanine, glutamic acid	4.88E-03
Tryptophan metabolism	Map00380	2-aminobenzoic acid, 3-hydroxyanthranilic acid, 5-hydroxyindoleacetic acid	6.81E-02

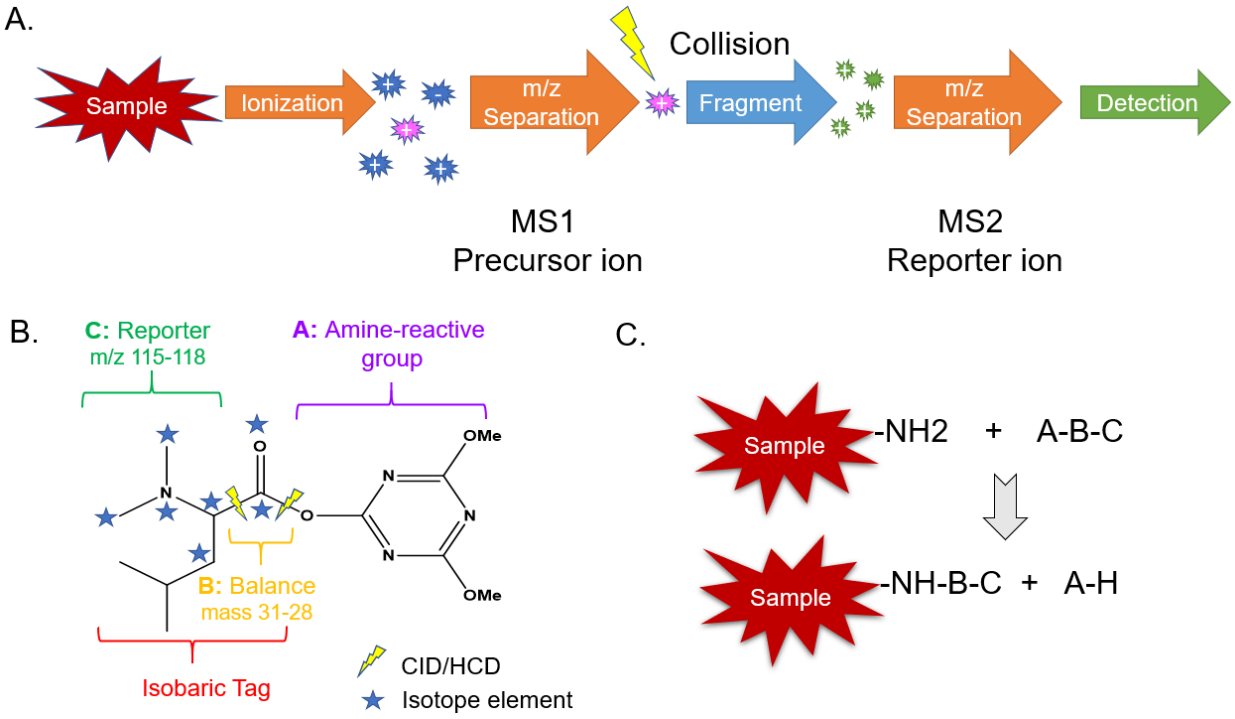
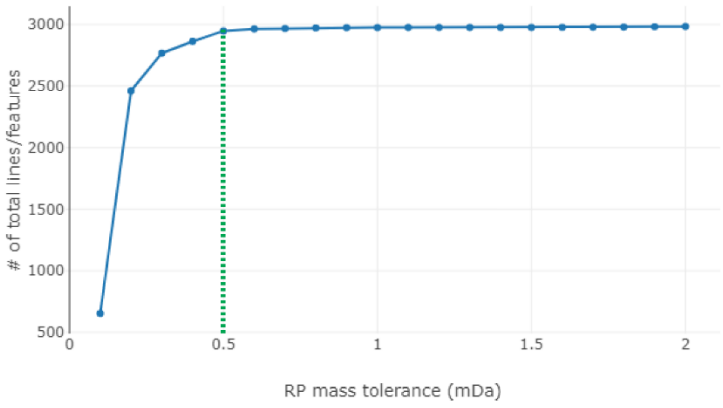
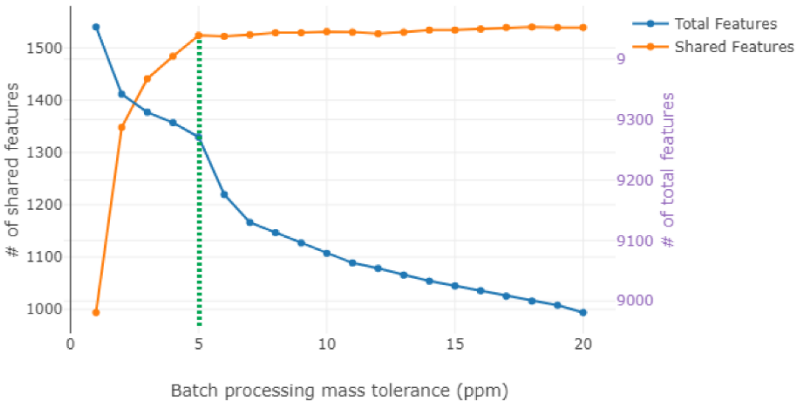


Figure 1. A: General set-up of a tandem mass spectrometer. **B:** Basic structure of DiLeu tags. **C:** How DiLeu tags reacted with metabolites

A. Optimizing reporter ion mass tolerance



B. Optimizing batch processing mass tolerance



C. Optimizing batch processing retention time tolerance

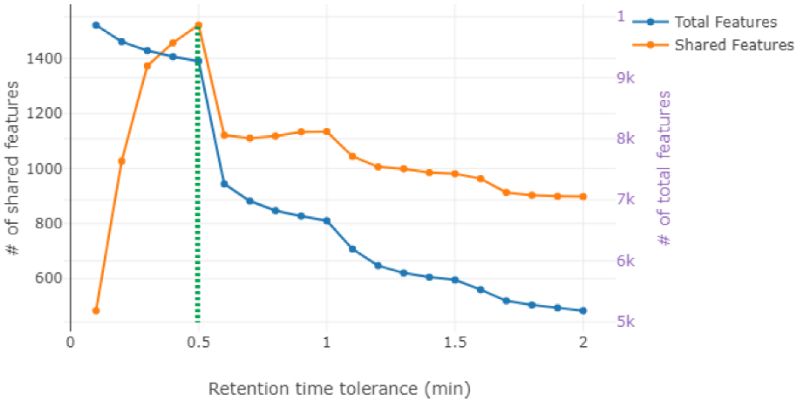


Figure 2. Metandem parameter optimization results: A: Optimizing reporter ion mass tolerance (0.5 mDa); **B:** Optimizing batch processing retention time tolerance (5 ppm); **C:** Optimizing batch processing retention time tolerance (0.5 min).

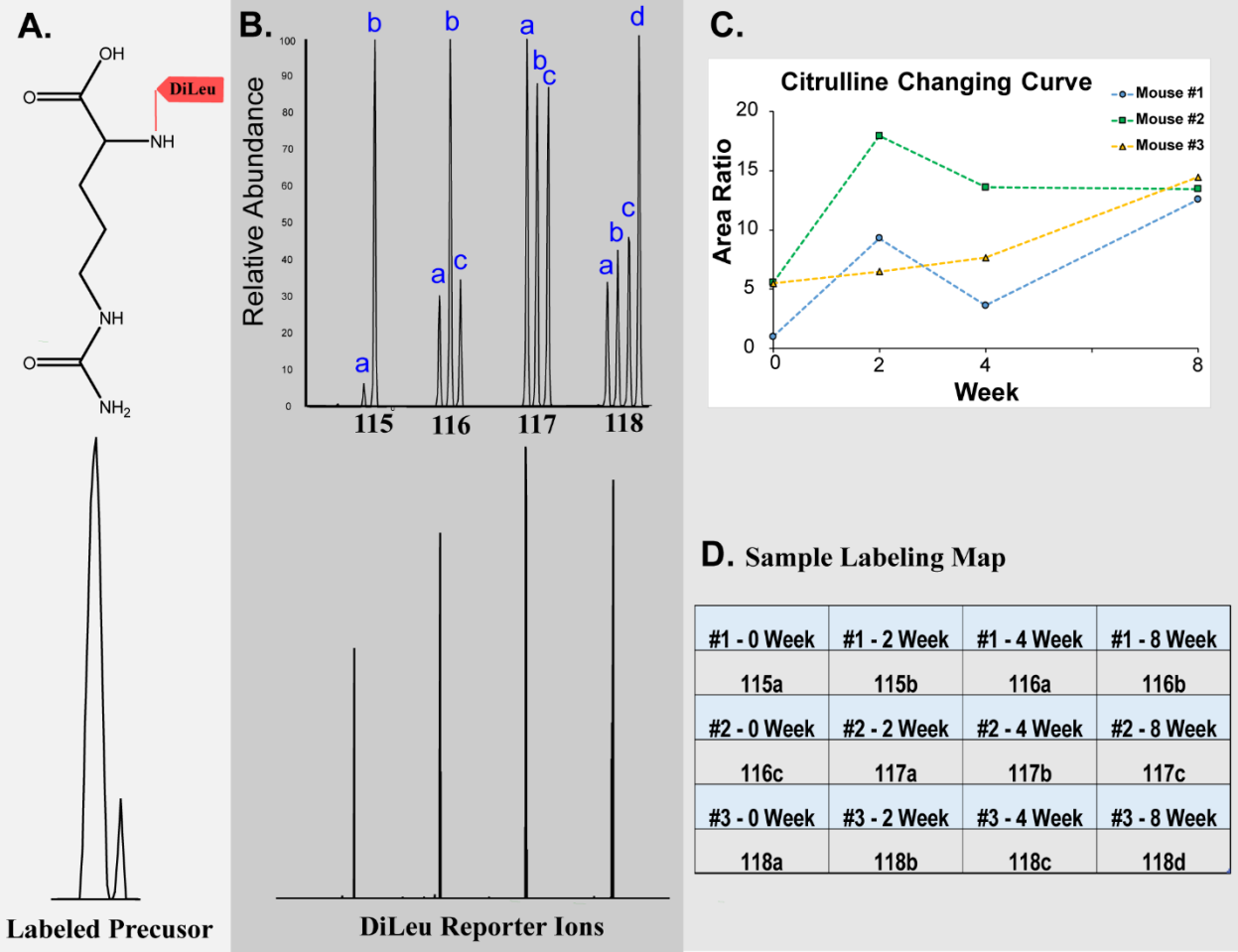


Figure 3. Details of 12-plex DiLeu labeling. A: Precursor ion of DiLeu-labeled citrulline; **B:** low *m/z* region showing DiLeu reporter channels at high resolution (top) and zoom in spectra (bottom); **C:** citrulline changing curve from different time points of three biology replicates. **D:** sample labeling map (randomized)

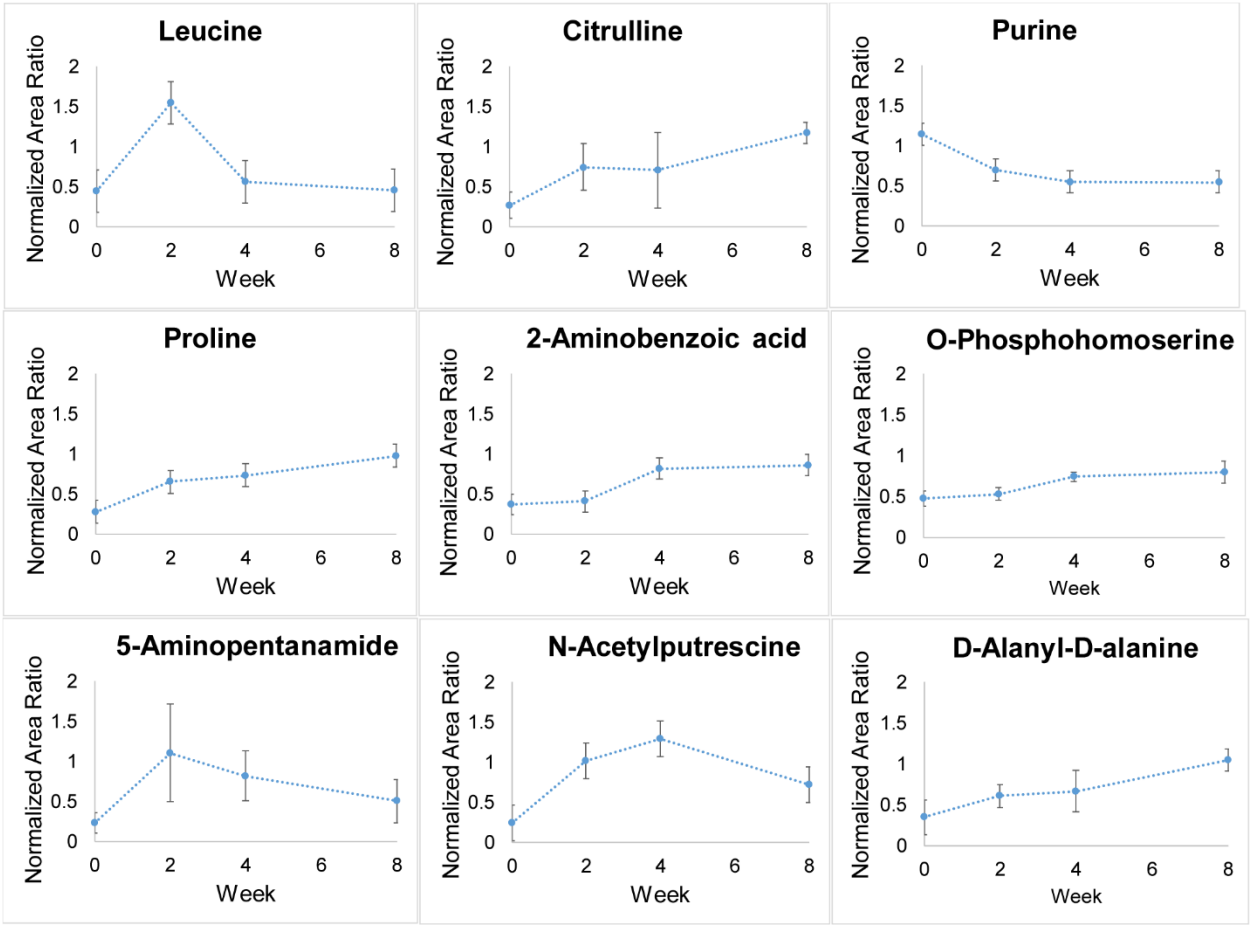


Figure 4. Quantification curves of 9 significantly changed urine metabolites (*n* = 3; RM-ANOVA, $p < 0.05$).

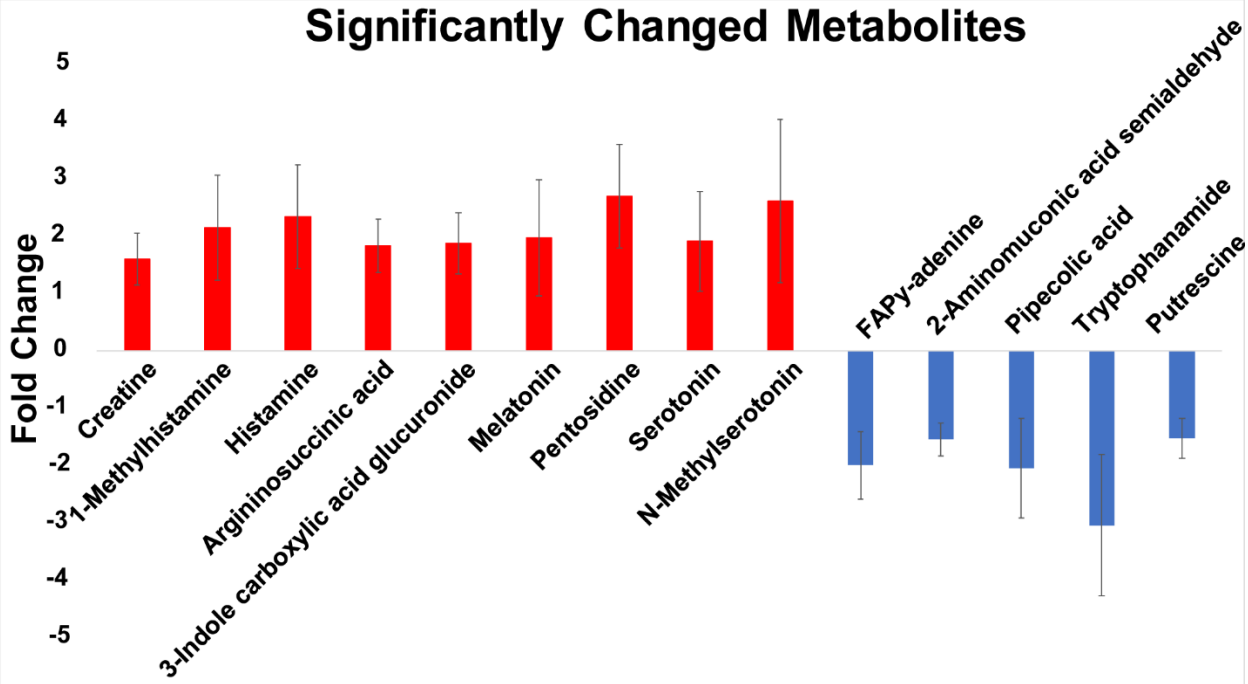


Figure 5. Significantly changed metabolites. 14 metabolites were identified as statistically significant biomarker candidates as shown here by a fold change (bacterial-induced prostatic inflammation mice with respect to control mice). 9 metabolites were up-regulated (red bars) and 5 metabolites were down-regulated (blue bars).

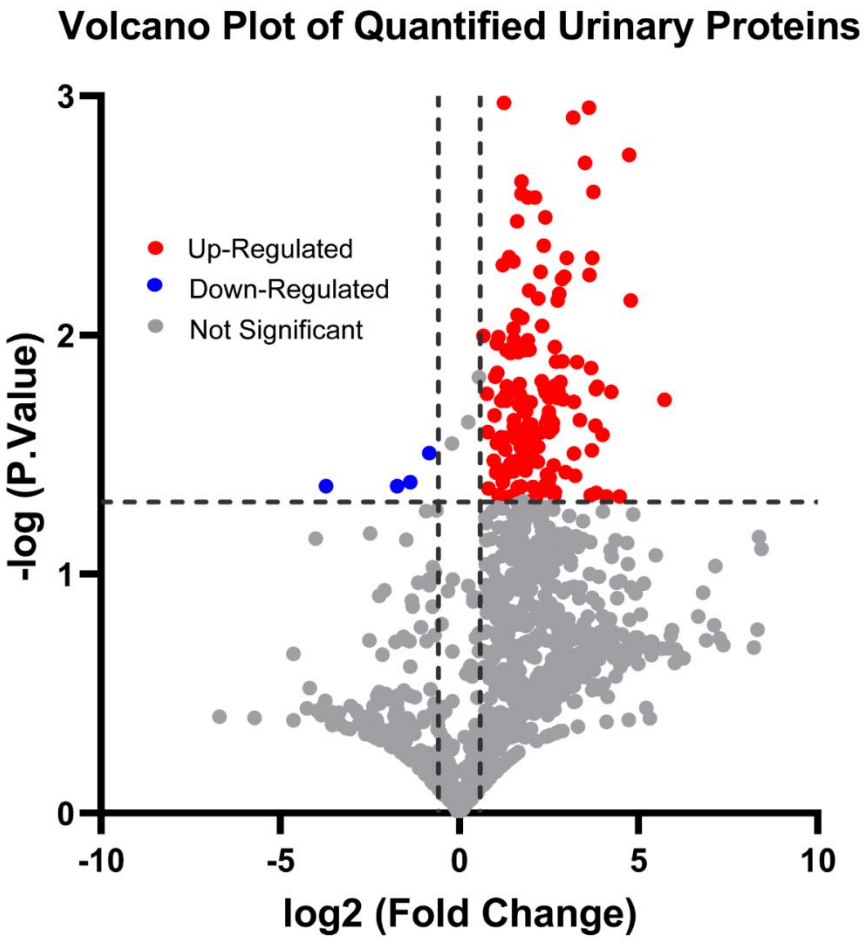


Figure 6. Volcano plot of quantified urinary proteins. The volcano plot of quantified urinary proteins reveals that 168 such proteins were significantly changed in bacterial-induced prostatic inflammation mice compared to control samples. Proteins with fold changes > 1.5 and p-values < 0.05 are shown (red spots: up-regulated; blue spots: down-regulated). Significant proteins were further characterized through literature searches and GO-term enrichment analysis.

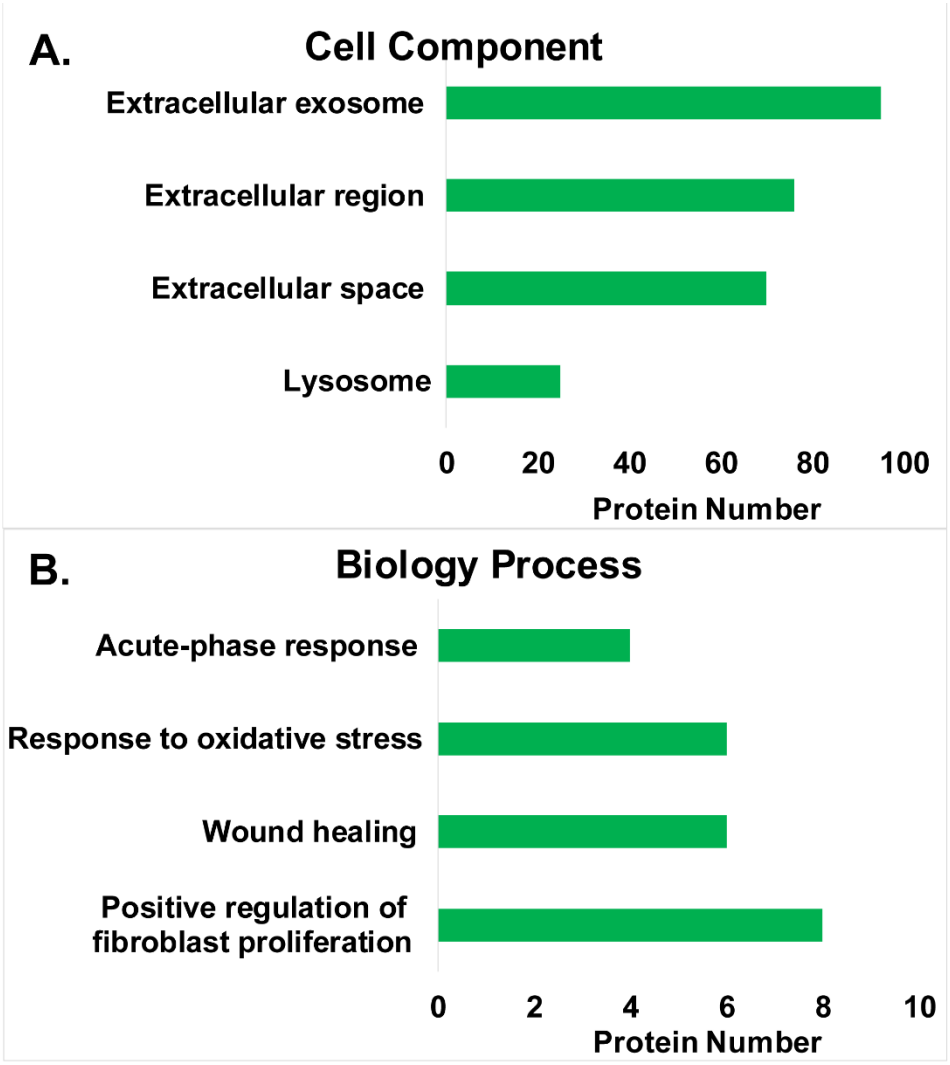


Figure 7. Gene ontology analysis of significantly changed proteins. A: Cell component (top four categories) of significantly changed proteins. **B:** Biology processes (inflammation-related) of significantly changed proteins.

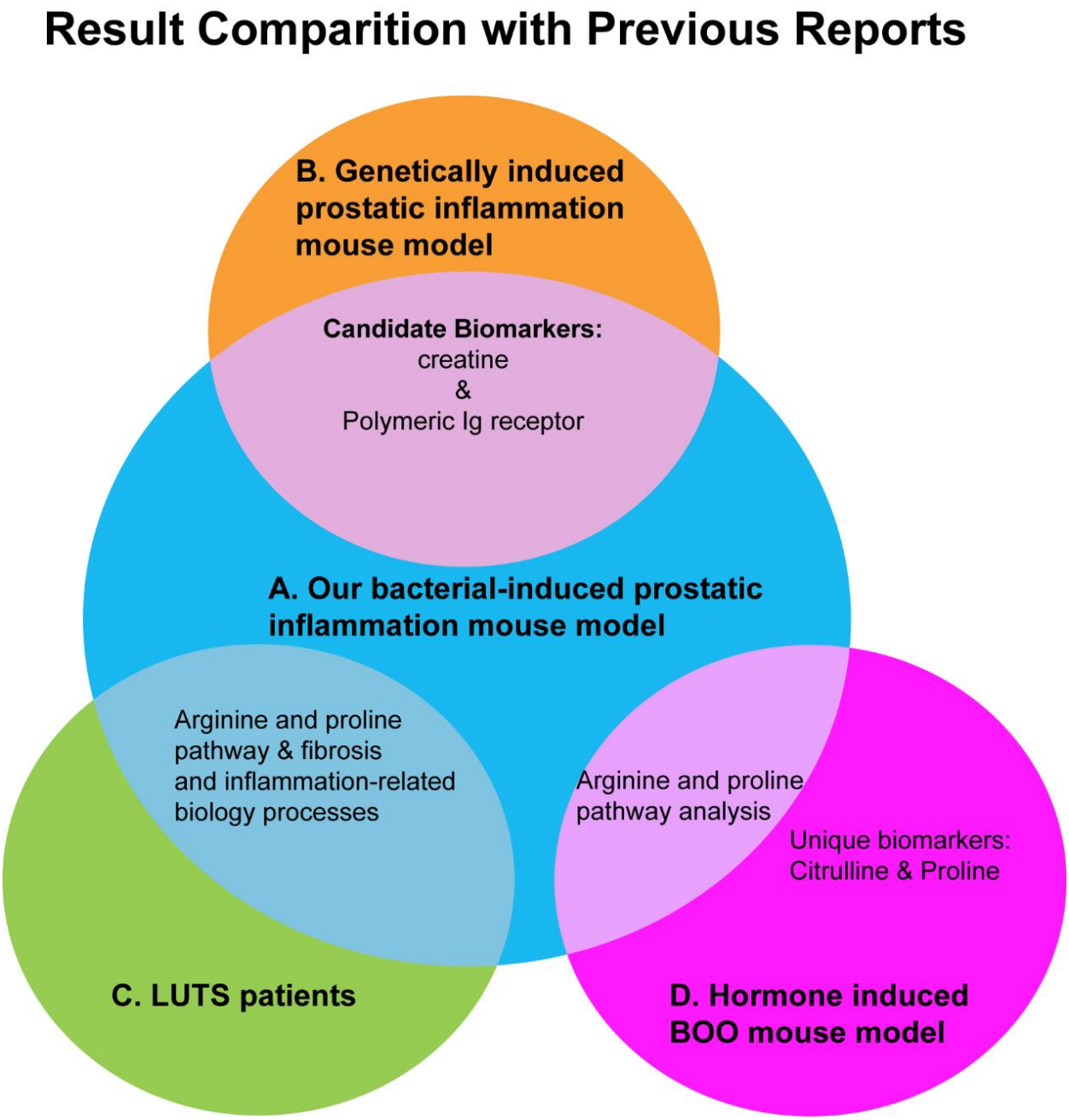
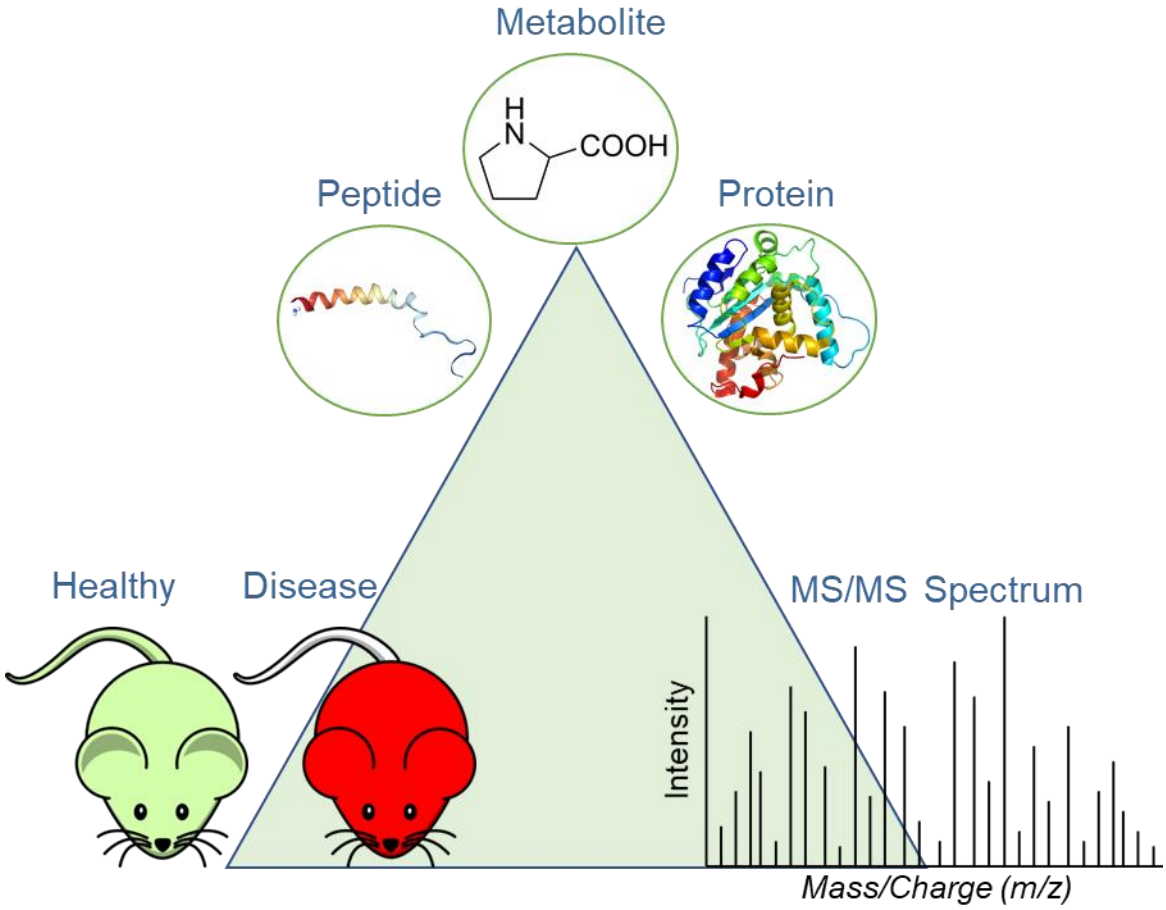


Figure 8. Result comparison with the previous study. A: Our bacterial-induced prostatic inflammation mouse model. **B:** Genetically induced prostatic inflammation mouse model. **C:** LUTS patients. **D:** Hormone induced BOO mouse model.

Chapter 9

Conclusions and Future Directions



Conclusions

This dissertation advances the field of mass spectrometry (MS)-based multi-omics through the development of stable isotope labeling method, decision tree method, and multi-omics integration method. The analytical methodologies are mainly focused on the use of nanoflow RPLC-ESI-MS platform. These methods have been successfully applied to human disease studies via clinical specimens collected from human patients and animal models.

Chapters 2 demonstrates that twelve-plex N, N-dimethyl leucine (DiLeu) isobaric labeling is an attractive high throughput strategy for the identification and quantification of amine-containing metabolites. Metandem is a useful tool for large-scale stable isotope labeling-based metabolomics data analysis. Paired together, these tools offer a powerful and accessible method for relative quantification of amine-containing metabolites in disease biomarker research. With the isobaric labeling method and Metandem software (1) for the data analysis, we were able to identify and quantify 59 amine-containing metabolites in the urine samples of hormone-induced urinary obstruction mice model and relate them to metabolic pathway analysis to explore their role in the lower urinary tract symptoms (LUTS) development. LUTS patients (2) and this mouse model shared common pathways that are dysregulated compared to control groups, for instance, the arginine and proline metabolism pathway. Proline presents an especially attractive candidate biomarker for hormone-induced LUTS, as it was significantly increased in both human LUTS and this mouse model.

Chapters 3 demonstrated that urine metabolomics and proteomics provide an informative and non-invasive method for candidate biomarker determination and etiologies classification of LUTS

induced by inflammation. Also, the investigation of DiLeu labeling on metabolites allowed metabolomics and proteomics analysis on the same LC-MS platform. In total, 143 amine-containing metabolites and 1058 urinary proteins were identified and quantified; among them, 14 metabolites and 168 proteins were significantly changed by prostatic inflammation. Five metabolic pathways and four inflammation-related biological processes were potentially disrupted. By comparing our findings with urinary biomarkers identified in a mouse model of genetic-induced prostate inflammation (3, 4) and with those previously found to be associated with LUTS in older men (5, 6), we identified creatine, haptoglobin, immunoglobulin kappa constant and polymeric Ig receptor as conserved biomarkers for prostatic inflammation associated with LUTS. These data suggest that these biomarkers could be used to identify men in which prostate inflammation is present and contributing to LUTS.

A sensitive MS method is developed for comprehensive neuropeptide characterization in **Chapter 4,** and every step in the workflow, from sampling, instrumentation analysis, to data analysis, were optimized. From the comparison of the three most commonly used extraction methods, a high acidified methanol method shows the highest extraction efficiency and lowest methylation percentage, which was further applied to the human pituitary tumor neuropeptide extraction. To improve the instrumentation analysis, a data-dependent decision tree scheme was established with Higher-energy C-trap dissociation (HCD) as the primary fragmentation method at all charge states, with Electron-transferred Higher-Energy Collision Dissociation (EThcD) complemented at charge state 2 and Electron transfer dissociation (ETD) at charge states 3-6. Furthermore, both mature neuropeptide database and preprohormone database were applied to the database searching. After applying this novel MS approach to 4 human pituitary tumor samples, in total, 119 neuropeptides

were detected, and 39 were shared among four samples. Known neuropeptides and proteins, such as Neurosecretory protein VGF (regulate cell-cell interactions), Secretogranin-1, and Secretogranin-2 (neuroendocrine secretory granule protein), ProSAAS (control of the neuroendocrine secretory pathway) were all identified. Furthermore, seven novel sulfation modification sites and four novel neuropeptides have been identified from human pituitary tumor samples. These results demonstrated the efficiency of this novel approach.

Chapters 5 focuses on the subcommissural organ (SCO) neuronal function exploration. A total of 12 neuropeptides were detected in the SCO, with nine of them exhibiting 100% sequence coverage. Most of the identified neuropeptides were secreted neuropeptides and peptide manserin, which could regulate the neuroendocrine system, was identified in mice for the first time. Also, to identify more PTM modified neuropeptides in the SCO, an advanced setting that searched raw data against 313 built-in modifications was used. Besides the common peptide PTMs, such as Acetylation (N-terminal), Amidation, and Oxidation (M), rare PTMs, such as citrullination, and Hexose (NSY) were also identified with full Ascore (PTM confidence score). Since SCO could secrete glycoproteins into cerebrospinal fluid (CSF), glyco-neuropeptides were explored by searching raw data against both 309 mammalian N-Glycan and 78 mammalian O-glycan modification lists. Eleven O-glycosylated peptides were found in thymosin beta 10, thymosin beta 4, and Acyl-CoA-binding proteins. Furthermore, the preprohormone database that included all the mouse proteins with signal peptide sequences was also used for the novel neuropeptide exploration. While comparing with the control sample, a candidate novel neuropeptide, DVGSYQEKVDVVLGPIQLQSPSKE, was only found in the SCO. It has been identified as a bioactive peptide in rat plasma after hemorrhagic shock.

The mini-review in **Chapters 6** summarizes the significant progress which has been made over the past decades in recognizing the importance of neuropeptides in the bidirectional communication between the brain and the gut and their influence on host immunity and stress (7). Neuropeptides such as CGRP are involved in host monitoring of the gut environment and have an essential function that connects nervous and immune systems during infection. Neuropeptides such as PACAP could directly act on the bacteria membrane to kill the bacteria. The effective defense and low bacterial resistance probability make the antimicrobial neuropeptides an attractive new class of antibiotics. Key findings show that stress influences the composition of the gut microbiota and that bidirectional communication between microbiota and the CNS influences stress reactivity and responses. Neuropeptide CRH acts as a common mediator during this dynamic gut-brain axis signaling process. The positive effects of PNX on anxiety and cognitive function have the potential to be impactful for the development of therapeutic strategies to treat these disorders.

In **Chapter 7**, I integrated MS-based multi-omics to investigate how the host immune system is shaped by the human microbiome during infection. The results from the metabolomics, peptidomics, and proteomics studies indicated that the 92 microbial strains in the human microbiome enhanced the host immune system and could help the host defense during infection. Also, the neuropeptide of the cathelin-related antimicrobial peptide could be a potential novel antibiotic for the treatment of *Salmonella* infection. Furthermore, the neuropeptides found in the gut facilitated further understanding of the bidirectional communication between gut microbiota, brain, and immune system.

Chapter 8 describes the two studies of LUTS biomarker discovery for the general public. Overall, this work improved techniques for both qualitative and quantitative profiling of metabolites, peptides, and proteins, which can be applied to answer some challenging questions in analytical chemistry, pharmaceutical sciences, and clinical settings.

Future Directions

To utilize the biomarkers identified in **Chapter 2** and **Chapter 3** to stratify patients with LUTS, which are induced with aging-related hormone change and prostatic inflammation, it will first be necessary to stratify patients by mechanism or by clinical features. For example, urine could be obtained from patients prior to prostate surgery, and the proteomic and metabolomic analysis performed on those patients in whom substantial inflammation is present in the removed prostate tissue. Alternatively, an analysis could be performed on patients who exhibit much greater irritative than obstructive symptoms. Once a correlation is established with a subsegment of the LUTS patients, then the biomarkers might be used as a tool for stratification and individualized treatment.

The optimized workflow developed in **Chapter 4** could be further used for neuropeptide identification and quantification. And the novel neuropeptides identified in both **Chapters 4** and **5** could be evaluated, and their neuronal function exploration could be conducted in the follow-up studies. Specifically, the novel sulfated sites found in neuropetides of Secretograini-1 and in the human pituitary tumor could be further investigated. There might be a possibility that the sulfated peptides play a vital role during pituitary tumor development.

We are just beginning to understand the meaning of gut-to-brain microbiome interactions and what role neuropeptides ultimately play for host homeostasis, including immunity and stress. Although **Chapter 6** reviewed their implications with some highlighted neuropeptides, some of the precise mechanisms and overall effects are not fully understood. Further progress in understanding the various processes involved in neuropeptide modulation of the interactions between the gut microbiome and the central and peripheral nervous systems is essential to develop effective treatments for immune and stress disorders with neurogenic components. Because of studies such as those highlighted in **Chapter 7**, our understanding of the interactions between the brain and gut through the gut-brain axis continues to expand, and novel therapies will be developed to treat gut microbiome-mediated immune and stress-related diseases. Neuropeptides, their receptors, and the proteases that degrade the same neuropeptides may become the special target of new pharmacological approaches.

Overall, this dissertation has significant impacts on both analytical method development and disease applications. The successful disease applications offer new insights and promising molecular targets for LUTS, pituitary tumor and gut-brain axis. It is anticipated that the analytical methodologies and disease applications presented in this dissertation can be potentially applied to the pharmaceutical industry and clinical settings.

References

1. Hao L, Zhu Y, Wei P, Johnson J, Buchberger A, Frost D, Kao WJ, Li L. 2019. Metandem: An online software tool for mass spectrometry-based isobaric labeling metabolomics. Anal Chim Acta 1088:99–106.

2. Hao L, Greer T, Page D, Shi Y, Vezina CM, Macoska JA, Marker PC, Bjorling DE, Bushman W, Ricke WA, Li L. 2016. In-Depth Characterization and Validation of Human Urine Metabolomes Reveal Novel Metabolic Signatures of Lower Urinary Tract Symptoms. Sci Rep 6:1–11.
3. Hao L, Shi Y, Thomas S, Vezina CM, Bajpai S, Ashok A, Bieberich CJ, Ricke WA, Li L. 2018. Comprehensive urinary metabolomic characterization of a genetically induced mouse model of prostatic inflammation. Int J Mass Spectrom 434:185–192.
4. Hao L, Thomas S, Greer T, Vezina CM, Bajpai S, Ashok A, De Marzo AM, Bieberich CJ, Li L, Ricke WA. 2019. Quantitative proteomic analysis of a genetically induced prostate inflammation mouse model via custom 4-plex DiLeu isobaric labeling. Am J Physiol - Ren Physiol 316:F1236–F1243.
5. Hao L, Greer T, Page D, Shi Y, Vezina CM, Macoska JA, Marker PC, Bjorling DE, Bushman W, Ricke WA, Li L. 2016. In-Depth Characterization and Validation of Human Urine Metabolomes Reveal Novel Metabolic Signatures of Lower Urinary Tract Symptoms. Sci Rep 6:30869.
6. Greer T, Hao L, Nechyporenko A, Lee S, Vezina CM, Ricke WA, Marker PC, Bjorling DE, Bushman W, Li L. 2015. Custom 4-Plex DiLeu Isobaric Labels Enable Relative Quantification of Urinary Proteins in Men with Lower Urinary Tract Symptoms (LUTS). PLoS One 10:e0135415.
7. Wei P, Keller C, Li L. 2020. Neuropeptides in gut-brain axis and their influence on host immunity and stress. Comput Struct Biotechnol J 18:843–851.

Appendix I

List of Publications and Presentations

Publications

1. **Pingli Wei**, Caitlin Keller, Lingjun Li. "Neuropeptides in Microbiome-Gut-Brain Axis and Their Influence on Host Immunity and Stress." Comput Struct Biotechnol J 18:843–851.

2. **Pingli Wei**, Ling Hao, Samuel Thomas, Amanda Rae Buchberger, William Ricke, Paul Marker, Lingjun Li. "Urinary Amine Metabolomics Characterization of a Hormone-Induced Urinary Obstruction Mouse Model with Custom 12-plex Isobaric DiLeu labeling." Under reviewed by JASMS.

3. **Pingli Wei**, Ling Hao, Fengfei Ma, Qing Yu, Amanda Buchberger, Sanghee Lee, Wade Bushman, Lingjun Li. "Urinary Metabolomic and Proteomic Analyses in a Mouse Model of Prostatic Inflammation." Accepted by *Urine*.

4. **Pingli Wei**, Caitlin Keller, Jennifer R. Bratburd, Rui Liu, Eugenio Vivas, Erin Gemperline, Federico E. Rey, Cameron R. Currie, Lingjun Li. "Integrated, Multi-Omics Strategy to Study the Host Immune Response Induced by Human Microbiome during *Salmonella enterica* Typhimurium and *Candida albicans* Infection." To be submitted.

5. **Pingli Wei*,** Qing Yu*, Haidan Sun*, Fengfei Ma, Vaishali P Bakshi, Wei Sun, Zhi Zheng, Chun Zeng, Lingjun Li. "Development of a Novel Mass Spectrometry Approach for Comprehensive Neuropeptidome Characterization and its Application to the Analysis of Human Pituitary Tumor." (*Co-first authors) To be submitted.

6. **Pingli Wei**, Fengfei Ma, Woo-Ping Ge, Lingjun Li. "Neuropeptidomic Study of the Mammalian Subcommissural Organ (SCO) by High Resolution Mass Spectrometry." To be submitted.

7. **Pingli Wei,** Ewa Jankowska-Gan, Diego Lema, Lynn D Haynes, Matthew Brown, William J Burlingham, Lingjun Li. "Self-target Antigen Characterization to Improve Heart and Lung Post-Transplant Survival Rate." Manuscript in preparation.

8. Ning Yang, Krishna D. B. Anapindi, Stanislav S. Rubakhin, **Pingli Wei**, Qing Yu, Lingjun Li, Paul J. Kenny, Jonathan V. Sweedler. "Neuropeptidomics of the Rat Habenular Nuclei." *Journal of Proteome Research*, 2018.

9. Ling Hao, Yuerong Zhu, **Pingli Wei**, Jillian Johnson, Amanda Buchberger, Dustin Frost, W. John Kao, Lingjun Li. "Metandem: A novel online software platform for mass spectrometry-based isobaric labeling metabolomics." *Journal of Analytica Chimica Acta*, 2019.

10. Ling Hao, **Pingli Wei**, Jingxin Wang, Tyler Greer, Ozioma Okonkwo, Lingjun Li. "Absolute Quantification of Amine Metabolites in Human Cerebrospinal Fluid via MS1-centric Isotopic DiLeu (iDiLeu) Labeling." To be submitted.

11. Caitlin Keller, **Pingli Wei**, Benjamin Wancewicz, Tzu-Wen L Cross, Frederico Rey, Lingjun Li. "Extraction Optimization for Combined Metabolomics, Peptidomics, and Proteomics Analysis of Cecum Gut Microbiota Samples." To be submitted.

12. Rui Liu, **Pingli Wei,** Caitlin Keller, Nicola Salvatore Orefice, Yatao Shi, Zihui Li, Junfeng Huang, Yusi Cui, Dustin C. Frost, Shuying Han, Tzu-Wen L. Cross, Federico E. Rey, Lingjun Li. “Integrated Label-Free and Ten-Plex DiLeu Tag Quantitative Methods for Profiling Changes in the Mouse Hypothalamic Neuropeptidome and Proteome under Different Gut Microbiota Environments.” To be submitted.

13. Fengfei Ma, **Pingli Wei,** Jingxin Wang, Vaishali Bakshi, Brian Baldo, Lingjun Li. “Profiling peptidomic changes in major brain regions in response to cocaine intoxication and withdrawal.” Manuscript in preparation.

14. Chunlan Tang, **Pingli Wei,** Min Ma, Lingjun Li. “Multi-omics profiling of rat liver reveals linoleic acid metabolic pathways associated with Alzheimer’s disease.” Manuscript in preparation.

Presentations

1. **Pingli Wei**, Ewa Jankowska-Gan, Diego Lema, Lynn D Haynes, Matthew Brown, William J Burlingham, Lingjun Li. "Self-target Antigen Characterization to Improve Heart and Lung Post-Transplant Survival Rate" (Oral-Pittcon, Chicago, IL, March 2020)

2. **Pingli Wei**, Fengfei Ma, Woo-Ping Ge, Lingjun Li. "Neuropeptidomic Study of the Mammalian Subcommissural Organ (SCO) by High Resolution Mass Spectrometry." (Poster-ASMS, Houston, TX, June 2020)

3. **Pingli Wei**, Caitlin Keller, Jennifer R. Bratburd, Rui Liu, Eugenio Vivas, Erin Gemperline, Federico E. Rey, Cameron R. Currie, Lingjun Li. "Integrated, Multi-Omics Strategy to Study the Gut Microbiota Response to Salmonella enterica Typhimurium Infection in Humanized Mice." (Poster-AbbVie Analytical Research Symposium, Chicago, IL, July 2019)

4. **Pingli Wei**, Caitlin Keller, Jennifer R. Bratburd, Rui Liu, Eugenio Vivas, Erin Gemperline, Federico E. Rey, Cameron R. Currie, Lingjun Li. "Integrated, Multi-Omics Strategy to Study the Gut Microbiota Response to Salmonella enterica Typhimurium Infection in Humanized Mice." (Poster- ASMS, Atlanta, GA, June 2019)

5. **Pingli Wei**, Qing Yu, Haidan Sun, Fengfei Ma, Vaishali P Bakshi, Wei Sun, Zhi Zheng, Chun Zeng, Lingjun Li. "Novel Mass Spectrometry Approach for Comprehensive Neuropeptidome Characterization and its Application to Human Pituitary Tumor." (Poster-Wisconsin Human Proteomics Symposium, Madison, WI, August 2018)

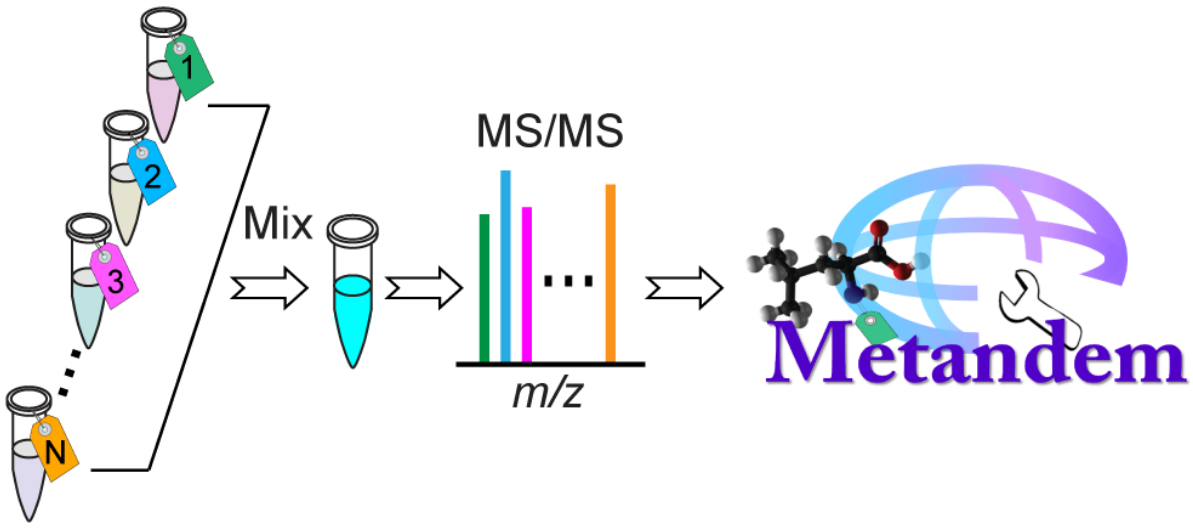
6. **Pingli Wei**, Qing Yu, Haidan Sun, Fengfei Ma, Vaishali P Bakshi, Wei Sun, Zhi Zheng, Chun Zeng, Lingjun Li. "Novel Mass Spectrometry Approach for Comprehensive Neuropeptidome Characterization and its Application to Human Pituitary Tumor." (Poster-ASMS, San Diego, CA, June 2018)

7. **Pingli Wei,** Ling Hao, Fengfei Ma, Qing Yu, Amanda Buchberger, Sanghee Lee, Wade Bushman, Lingjun Li. "Urinary Metabolomics and Proteomics Analysis of an *E.coli*-Induced Prostatic Inflammation Mouse Model via Isobaric DiLeu Tagging and Label-free Methods." (Poster-ASMS, Indianapolis, IN, June 2017)

8. **Pingli Wei**, Ling Hao, Jingxin Wang, Tyler Greer, Ozioma Okonkwo, Lingjun Li. "Absolute Quantification of Amine Metabolites in Human Cerebrospinal Fluid via MS1-centric Isotopic DiLeu (iDiLeu) Labeling." (Poster- ASMS, San Antonio, TX, June 2016)

Appendix II

Metandem: A Novel Online Software Platform for Mass Spectrometry-based Isobaric Labeling Metabolomics



Adapted from L. Hao, Y. Zhu, **P. Wei**, J. Johnson, A. Buchberger, D. Frost, W. J. Kao, L. Li. "Metandem: a Novel Online Software Platform for Mass Spectrometry-Based Isobaric Labeling Metabolomics" *Journal of Analytica Chimica Acta*, 2019. PW conducted labeling experiments and software evaluation.

Abstract

Mass spectrometry-based stable isotope labeling is a key technology for the analysis of biomolecules with the advantages of multiplexing capability and accurate quantification. Isobaric labeling has been widely used for quantitative proteomics and peptidomics and was successfully adopted to metabolomics analysis in the past decade. Despite the rapid advancement of methodology and data analysis in the metabolomics field, barely any current metabolomics software is able to process stable isotope labeling-based metabolomics data, particularly for isobaric labeling using reporter ions produced by MS/MS for quantification. The well-established proteomics software tools cannot be readily adapted for metabolite identification and quantification. Therefore, we developed Metandem, a novel online software tool for isobaric labeling-based metabolomics, integrating metabolite quantification, identification, and statistical analysis, freely available at http://metandem.com/web/. Metandem is also the first omics data analysis software that provides straightforward online parameter optimization function for custom datasets. Systematic evaluation of the Metandem tool was demonstrated by duplex, 4-plex, 10-plex, and 12-plex custom isobaric N,N-Dimethyl Leucine (DiLeu) labeling and the applications to various biological samples.

Introduction

Metabolomics is an essential component of systems biology and has embraced rapid advancements over the past decades. The development of mass spectrometry (MS)-based techniques allows metabolomics to be carried out with sophisticated methodologies, as well as wide applications to biological and clinical studies [1]. MS-based metabolomics generates multidimensional datasets where thousands of features can be measured in a single instrument run, pressing significant challenges on data processing and analysis. Particularly for large-scale metabolomics studies, state-of-the-art analytical techniques must be paired up with proper bioinformatics software for automated and efficient data analysis [2].

Quantitative and qualitative metabolite analyses are typically performed by label-free or stable isotope labeling approaches. Stable isotope labeling, in particular isobaric labeling, has gained substantial popularity in proteomic and peptidomic studies and has also been successfully adopted to quantify small molecules in recent years [3-7]. Biomolecules derivatized by multiplexed isobaric labels have nearly identical mass shift of precursor ions, which can be fragmented into a panel of MS/MS reporter ions for quantification, using commercial isobaric tags TMT [8,9] and iTRAQ [10,11], or custom synthesized reagents like DiLeu [12-14], DiAla [15], DiVal [15], and DiART [16]. Besides amine-reactive isobaric labels, aminoxy TMT [17] was also developed to target molecules with carbonyl group such as carbohydrates and steroids. Stable isotope labeling provides advantages of multiplexing capability and accurate quantification, but often requires tailored bioinformatics tools for data analysis. A wide range of proteomics software have been well established for processing stable-isotope labeling-based datasets. Yet proteomics software platforms are specialized in protein/peptide analysis and cannot be easily adapted for small molecule analysis because of their distinct isotope distribution and identification algorithms. In

regards of metabolomics software, numerous packages have been developed for label-free metabolomics studies, such as XCMS [18,19], MZmine [20], MAVEN [21], SIEVE (Thermo), and Compound Discoverer (Thermo). However, barely any current metabolomics software packages are able to process stable isotope labeling-based metabolomics data, particularly for MS/MS generated reporter ion-based isobaric labeling. Scientists have to write their own program or script in order to process these datasets [9,11,16], which severely hampers the progress and applicability of isobaric labeling-based metabolomics.

To address this limitation and critical technological gap, we developed Metandem, a novel online software platform for the data analysis of isobaric labeling-based metabolomics. Metandem is freely available at http://metandem.com/web/ and very easy to use through graphic interface design. It provides a comprehensive data analysis pipeline integrating feature extraction, metabolite quantification, metabolite identification, batch processing of multiple data files, parameter optimization, median normalization, and statistical analysis (Figure 1). All graphs are interactive, which can be visualized/edited online or downloaded for offline analysis. The highlighted function of parameter optimization in Metandem allows straightforward selection of the best set of parameters based on different custom dataset. We evaluated the Metandem software tool using duplex, 4-plex, 10-plex, and 12-plex isobaric DiLeu labeling on different LC-MS and MALDI-MS platforms, and applied the complete isobaric labeling metabolomics workflow to various biological samples.

Experimental

Metabolite sample preparation

Metabolite standard mixtures and metabolite fractions from various biological samples were prepared in this study. Commercially-available pancreatic cancer cells (PANC1) and breast cancer cells (MCF7) were routinely cultured in the lab, and cells were quenched with methanol. Cellular metabolites were extracted with a methanol/chloroform/water extraction method. Metabolite fractions from the urine samples were obtained using 3 kDa molecular weight cut-off ultracentrifugation filters [22]. Metabolite standard mixtures were prepared by mixing individual stock solutions of 12 representative metabolites including histidine, valine, tyrosine, leucine, lysine, phenylalanine, tryptophan, alanine, serotonin, dopamine, γ-aminobutyric acid, and norepinephrine.

Multiplexed isobaric DiLeu reagents were custom synthesized following the procedure described previously with steps of reductive dimethylation of leucine and ^{18}O exchange [13]. Dried DiLeu reagents were stored in a desiccator at 4 °C and activated to the triazine ester form right before the labeling reaction with amine groups. A 20-fold molar excess of activated DiLeu was reacted with metabolite samples at a 70% of organic: aqueous ratio. Each plex of labeled sample was combined, dried, and purified by SCX Ziptip as described previously [14].

Mass spectrometry analysis

Multiple MS instrument platforms were used in this study to demonstrate the applicability of the Metandem tool, including a Dionex UltiMate 3000 nanoLC system coupled with a Fusion™ Lumos™ Orbitrap MS, a Dionex nanoLC coupled with a Q Exactive™ HF Orbitrap MS, a Dionex UHPLC coupled with a Q Exactive™ Orbitrap MS, and a MALDI LTQ Orbitrap MS. LC-MS analyses were conducted with a C18 column under a 40 min LC gradient separation and top 20 data-dependent acquisition. Mobile phase A was 0.1% formic acid in H_2O, and mobile phase B was 0.1% formic acid in ACN. Flow rate was 0.3 μL/min. Full MS scans

were acquired from *m/z* 100 to 1000 at a resolution of 60 K, automatic gain control at 5×10^5, and maximum injection time of 100 ms. MS/MS scans were acquired at a resolution of 60 K, an isolation window of 1 m/z, and a lower mass limit of 110 *m/z*. Normalized collision energy for MS/MS fragmentation was 30% with higher-energy collisional dissociation (HCD).

Data analysis using Metandem

Data analysis was achieved using the Metandem software tool, which is freely available at http://metandem.com/web/. The web interface allows users to select the number of reporter ions, input accurate mass and purity of each reporter ion, and upload data sets. Thermo .raw data files were converted to .txt format using COMPASS [23] software suite. The complete reporter ion information of different isobaric DiLeu reagents used in this study is listed in Supplemental Table S1. Data analysis parameters for the demonstration data set were optimized using the parameter optimization graphs (POGs) and optimal settings were set as the following: reporter ion mass tolerance of 0.4 mDa, batch processing mass tolerance of 6 ppm, and batch processing retention tolerance of 0.5 min. Reporter ion intensities were extracted from the dataset for metabolite quantification and statistical analysis. Molecular weight of the detected compound can be calculated based on the charge and mass shift caused by labeling, which was then searched against Human Metabolome Database [24] for metabolite identification.

Results and discussion

Metandem functionality and web interface

Metandem has four main functions including quantification, metabolite identification, parameter optimization, and statistical analysis (Figure 1). Metabolite quantification is achieved

at the MS/MS level by extracting reporter ion intensities from the isobaric labeling experiment. Metabolite identification is conducted through accurate mass matching against the Human Metabolome Database (HMDB) [24]. Since chemical derivatization can alter the fragmentation pattern of target molecules, MS/MS matching for ID confirmation needs to be performed manually instead of searching MS/MS databases. The data analysis pipeline of Metandem for isobaric labeling-based metabolomics can be summarized into five steps: 1) upload data; 2) define reporter ion information; 3) optimize parameters (optional); 4) submit job; and 5) generate output and interpret results. Metandem can process both individual data file and multiple data files as batch processing. The graphic interface of Metandem is shown in Figure 2. Metandem currently accepts .txt format for data upload. Available data conversion software tools include COMPASS [23] or ProteoWizard [25]. After uploading the data files, reporter ion information must be defined based on the isobaric labeling experiment. Currently up to 20 multiplexing reporter ion channels can be selected for quantification. Reporter ion information of commercial isobaric tags are summarized in Supplemental Table S2, including TMT, aminoxyTMT, and iTRAQ. Metandem also provides the function of purity correction to account for the isotopic interference and impurities from synthetic isotope reagents [14,26]. The purity of each label (default =1) can be input along with the accurate masses of reporter ions. In most cases, we recommend to select the option of only output data containing all reporter ions to ensure the best data quality (Figure 2). However, if some target molecules are known to be not present in some samples or with an extremely high fold change compared to other samples, this box can be unselected.

For most of the data analysis tools, software performance and quality of results are highly dependent on the set of parameters selected for data processing. Different datasets generated from diverse mass spectrometers may require unique parameter settings to ensure the best

performance. However, customization of parameters is often difficult and time-consuming. Some software packages provide recommended parameter settings or tailored settings for several types of instrument platforms, but rarely for specific custom datasets. To our knowledge, Metandem is the first omics data analysis software that provides online parameter optimization function. The program runs automatically with a range of parameter values to calculate the corresponding number of total/shared features to generate the POG curve. By simply checking the box of Output Parameter Optimization Graphs and submitting the job, users can view the POGs and select the best parameters for their specific custom datasets.

There are three POGs in total to optimize the reporter ion mass tolerance, the batch processing mass tolerance, and the batch processing retention time tolerance. Figure 3 A, B and C illustrate the POGs using the demonstration dataset. Reporter ion mass tolerance (0.1 to 2 mDa) is used to find specific reporter ions based on the input reporter ion masses and extract their intensities from the dataset for quantification. As illustrated in Figure 3A, the total number of features containing report ions increases with a mass tolerance-dependent manner and then reaches a plateau, indicating an optimal reporter ion mass tolerance of 0.4 mDa for the demo dataset. When multiple data files are uploaded for analysis, quantitative information is generated for each file and merged among multiple files on the feature level. With the increase of batch processing mass tolerance (from 0.1 to 20 ppm), the total number of features dropped as more features can be merged together. The shared features among multiple input files rises and then reaches a plateau at a mass tolerance of 6 ppm for batch processing of demo dataset (Figure 3B). For batch processing retention time tolerance (0.1-2 min), if the tolerance is too high, distinctive features are incorrectly merged together, causing the decreased number of both total features and shared features. As shown in Figure 3C for the demo dataset, the optimal retention time tolerance

is 0.5 min. Depending on the size of the dataset, the parameter optimization step only takes from several seconds up to 10 min to be finished. Test run function is also available to provide a quick evaluation of a subset of data.

Results of data analysis are summarized into tables (.csv) and statistical graphs to be either downloaded or visualized/edited online by simply clicking on the output link through Plotly web platform. The output result tables include all individual file tables, merged table, and metabolite identification table. If median normalization is selected, another merged table after ratio normalization will be generated. Reporter ion ratios are calculated from intensities with tag #1 as the denominator. The screenshots of these results tables were provided in Supplemental Figure S1. Each individual result table consists precursor ion information and reporter ion intensities and ratios of all quantified features. For the merged table, average reporter ion ratios and relative standard deviations (RSD) of ratios are calculated across all input data files.

Interactive statistical graphs are generated with only the shared features among all merged files to ensure the best data quality. Output graphs include histogram distribution of precursor mass (Figure 3D), histogram distribution of retention time (Figure 3E), and box plots of reporter ion ratios before and after median normalization (Figure 3F).

Evaluation and Applicability

Metandem tool was systematically evaluated with isobaric labeling experiments on both LC-MS/MS and MALDI-MS platforms. Duplex, 4-plex, 10-plex, and 12-plex isobaric DiLeu labeling was applied to standard, biofluid, and cancer cell samples for metabolite quantification and identification.

Accuracy and dynamic range of quantification are evaluated by labeling the same samples with multiplexed reagents which were then mixed at known ratios and analyzed by LC-MS/MS. Breast cancer cellular metabolites were labeled with 12-plex isobaric DiLeu, generating a total of 3510 merged features with three technical replicates. Pancreatic cancer cellular metabolites were labeled with duplex isobaric DiLeu with a total of 5456 merged features. Experimental average ratios can be plotted against theoretical ratios, where slope =1 represents perfect accuracy and R^2=1 represents perfect consistency and precision. As shown in Figure 4, box plots of reporter ion ratios demonstrate within 12% of accuracies for both 12-plex and duplex DiLeu labeling. Slopes for 12-plex and duplex labeling are 0.9691 and 1.122, respectively; R^2 for 12-plex and duplex labeling are 0.999 and 0.998, respectively, indicating excellent quantification accuracy and precision of the results generated by Metandem.

In order to validate feature extraction and metabolite identification functions of Metandem, we first use a simple dataset of 4-plex DiLeu labeled twelve metabolite standards mixture analyzed by UPLC-MS/MS platform. All 12 targets were successfully identified and quantified using Metandem. For complex samples like mouse urine, MCF7 breast cancer cells and PANC1 pancreatic cancer cells, over 2000 features were quantified, and over 500 features can be identified as metabolites (Table 1). A major pitfall of isobaric labeling-based metabolomics is that only subsets of the metabolome can be derivatized depending on the structure of chemical tags. Therefore, the results of metabolite identification need to be examined to exclude unreasonable matches from the database, which on the other hand, enhances the confidence of identification by the presence of reporter ions. For instance, TMT, iTRAQ and DiLeu labeled metabolites are amine-containing small molecules, and aminoxyTMT targets

carbonyl-containing molecules such as carbohydrates and steroids. These additional functional group requirements further increase the confidence of metabolite identification.

Conclusions

MS-based quantitative analysis has fostered the development of state-of-the-art methodologies in both proteomics and metabolomics studies. There is a pressing need for associated software and computational tools to process complex data automatically and efficiently. Metandem fills a critical gap in the metabolomics field, enabling isobaric labeling-based dataset to be processed with simple graphic interface design. Besides basic functions of metabolite quantification and identification, Metandem also provides unique feature of online parameter optimization to ensure the best performance of such large-scale metabolomic experiments. It is fast, accurate, easy to use, and freely available at http://metandem.com/web/.

Acknowledgements

The authors would like to thank Professor Wei Xu and Dr. Fabao Liu at the University of Wisconsin for providing breast cancer cells and Professor Paul Marker for providing mouse urine samples. We acknowledge the members of the Li group for their use, evaluation, and critical feedback on this newly developed software tool.

References

(1) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nature Reviews Molecular Cell Biology* **2012**, *13*, 263-269.

(2) Sarpe, V.; Schriemer, D. C. *Curr Opin Biotech* **2017**, *43*, 110-117.

(3) Rauniyar, N.; Yates, J. R., 3rd. *J Proteome Res* **2014**, *13*, 5293-5309.

(4) Chokkathukalam, A.; Kim, D. H.; Barrett, M. P.; Breitling, R.; Creek, D. J. *Bioanalysis* **2014**, *6*, 511-524.

(5) Guo, K.; Li, L. *Anal Chem* **2009**, *81*, 3919-3932.

(6) Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L. J. *Anal Chem* **2017**, *89*, 1138-1146.

(7) Greer, T.; Hao, L.; Nechyporenko, A.; Lee, S.; Vezina, C. M.; Ricke, W. A.; Marker, P. C.; Bjorling, D. E.; Bushman, W.; Li, L. J. *Plos One* **2015**, *10*.

(8) Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. *Anal Chem* **2003**, *75*, 4942-4942.

(9) Murphy, J. P.; Everley, R. A.; Coloff, J. L.; Gygi, S. P. *Anal Chem* **2014**.

(10) Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Molecular & Cellular Proteomics* **2004**, *3*, 1154-1169.

(11) Kaspar, H.; Dettmer, K.; Chan, Q.; Daniels, S.; Nimkar, S.; Daviglus, M. L.; Stamler, J.; Elliott, P.; Oefner, P. J. *J Chromatogr B* **2009**, *877*, 1838-1846.

(12) Xiang, F.; Ye, H.; Chen, R. B.; Fu, Q.; Li, L. J. *Anal Chem* **2010**, *82*, 2817-2825.

(13) Frost, D. C.; Greer, T.; Li, L. *Anal Chem* **2015**, *87*, 1646-1654.

(14) Hao, L.; Zhong, X. F.; Greer, T.; Ye, H.; Li, L. J. *Analyst* **2015**, *140*, 467-475.

(15) Yu, Q.; Shi, X. D.; Greer, T.; Lietz, C. B.; Kent, K. C.; Li, L. J. *Journal of Proteome Research* **2016**, *15*, 3420-3431.

(16) Yuan, W.; Zhang, J. X.; Li, S. W.; Edwards, J. L. *Journal Of Proteome Research* **2011**, *10*, 5242-5250.

(17) Zhong, X. F.; Chen, Z. W.; Snovida, S.; Liu, Y.; Rogers, J. C.; Li, L. J. *Anal Chem* **2015**, *87*, 6527-6534.

(18) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal Chem* **2006**, *78*, 779-787.

(19) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal Chem* **2012**, *84*, 5035-5039.

(20) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *Bmc Bioinformatics* **2010**, *11*.

(21) Melamud, E.; Vastag, L.; Rabinowitz, J. D. *Anal Chem* **2010**, *82*, 9818-9826.

(22) Hao, L.; Greer, T.; Page, D.; Shi, Y. T.; Vezina, C. M.; Macoska, J. A.; Marker, P. C.; Bjorling, D. E.; Bushman, W.; Ricke, W. A.; Li, L. J. *Sci Rep-Uk* **2016**, *6*.

(23) Wenger, C. D.; Phanstiel, D. H.; Lee, M. V.; Bailey, D. J.; Coon, J. J. *Proteomics* **2011**, *11*, 1064-1074.

(24) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. *Nucleic Acids Res* **2013**, *41*, D801-D807.

(25) Holman, J. D.; Tabb, D. L.; Mallick, P. *Curr Protoc Bioinformatics* **2014**, *46*, 13 24 11-19.

(26) Shadforth, I. P.; Dunkley, T. P. J.; Lilley, K. S.; Bessant, C. *Bmc Genomics* **2005**, *6*.

Figures and Tables

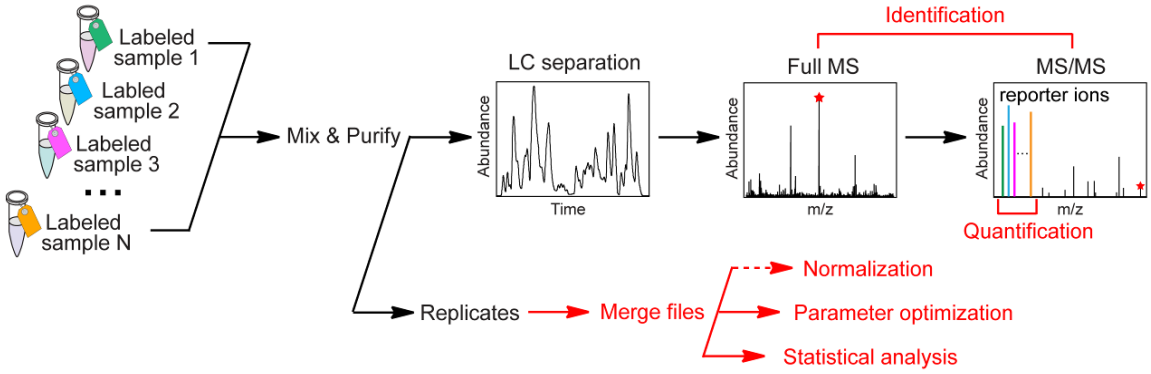


Figure 1. Pipeline of mass spectrometry-based isobaric labeling metabolomics using Metandem tool. Data analysis capability of Metandem is illustrated and highlighted in red color, including metabolite quantification, identification, parameter optimization, and statistical visualization.

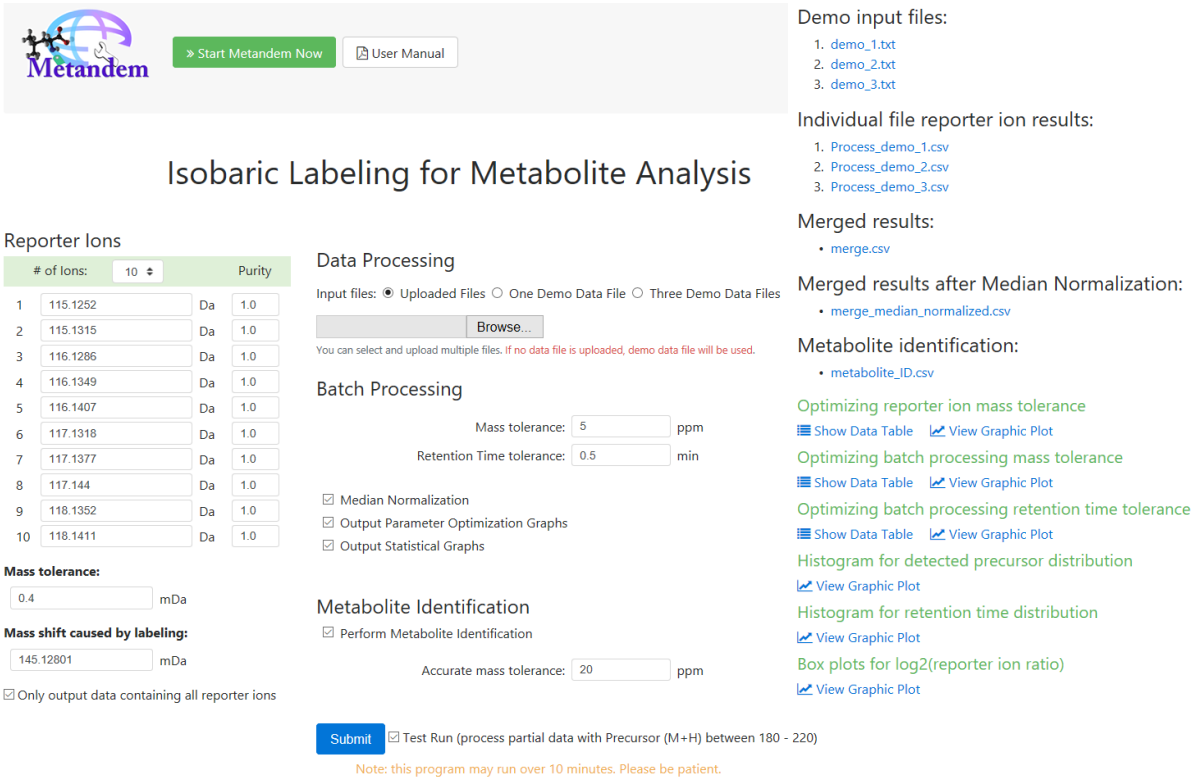


Figure 2. Screenshot of the Metandem web interface. Output results of three demo data file are illustrated on the right which can be downloaded as .csv files. Parameter optimization and statistical plots can also be edited/visualized online.

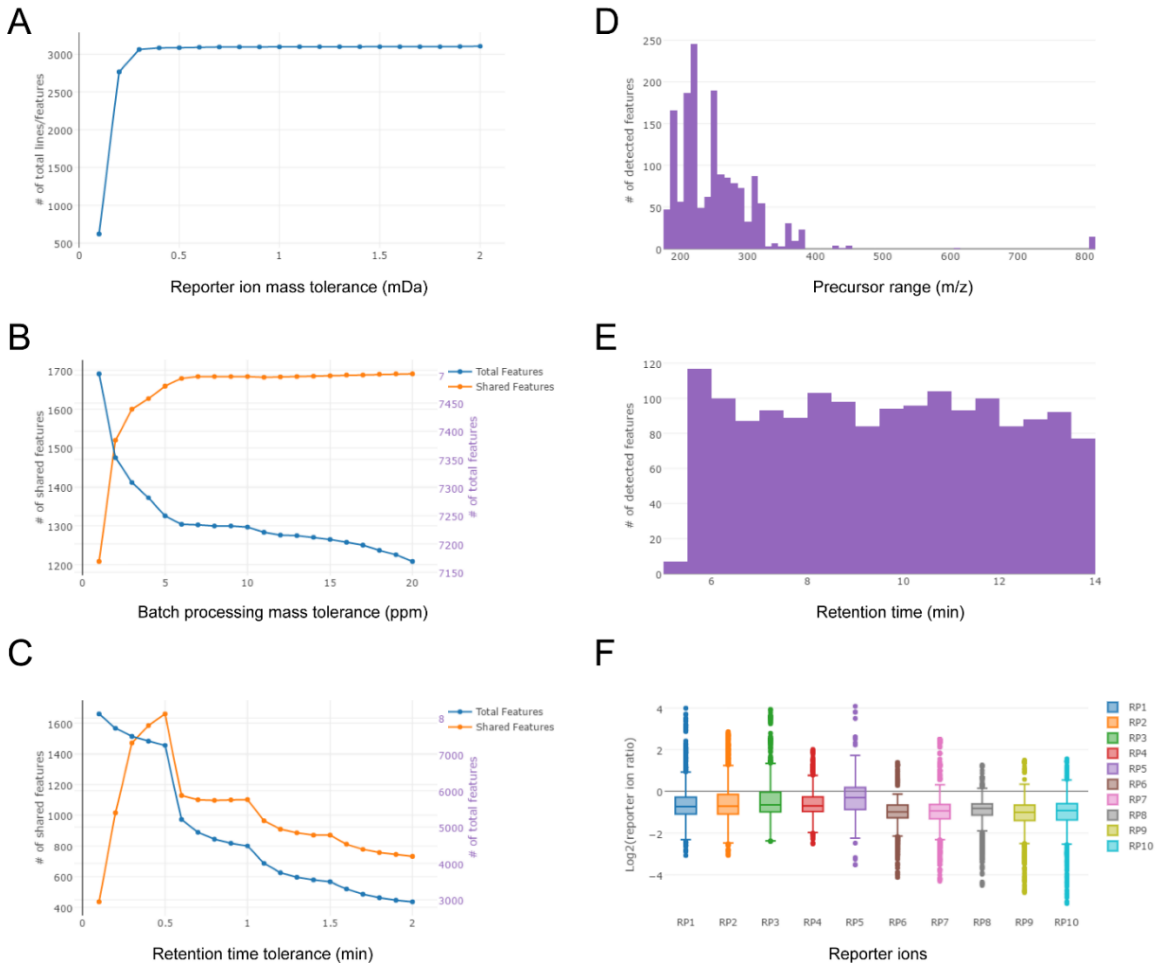


Figure 3. Parameter optimization graphs (POGs) and statistical analysis of Metandem using three demo files as input files. Parameter optimization graphs for reporter ion mass tolerance (A), batch processing mass tolerance (B), and batch processing retention time tolerance (C). Statistical graphs were generated using the shared features among all merged input files, including histogram of precursor mass range (D), histogram of retention time (E), and boxplots of reporter ion ratios (F). Box denotes 25th and 75th percentiles; line inside the box denotes the median; whisker denotes standard deviation.

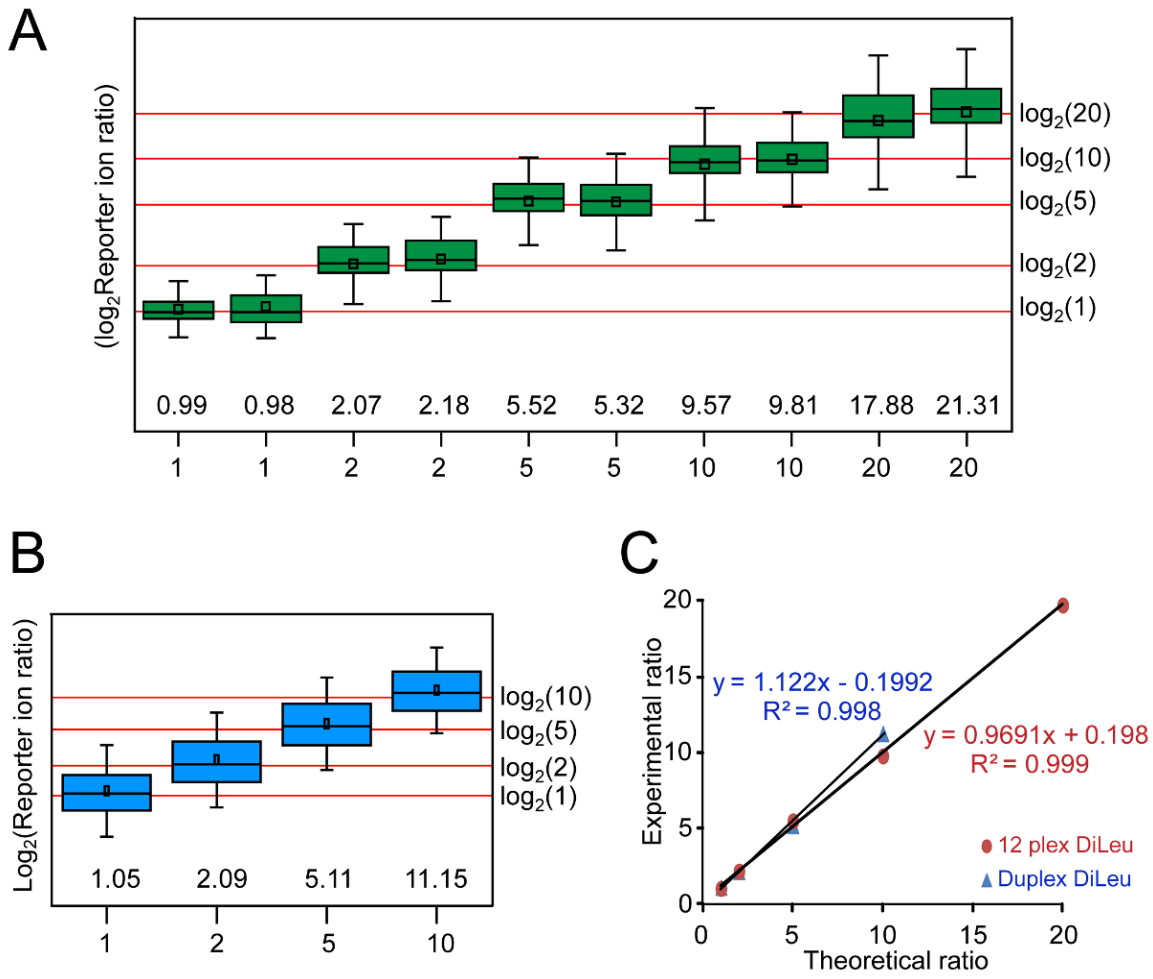


Figure 4. Quantification accuracy of Metandem tool demonstrated by isobaric DiLeu labeling of cellular metabolites. Boxplots of reporter ion ratios illustrate isobaric 12-plex DiLeu labeling of MCF7 cellular metabolites (A) and isobaric duplex DiLeu labeling of PANC1 cellular metabolites (B). Box denotes 25th and 75th percentiles; line inside the box denotes the median; whisker denotes standard deviation. Excellent linearity and dynamic range of relative quantification was shown in (C) with slope and R^2 close to 1.

Table 1. Results of isobaric labeling-based various biological samples using Metandem.

Biological sample	MS platform	Isobaric labels	Quantified feature[a]	Identified feature
Standards mix	UPLC-MS/MS	4-plex DiLeu	20	12
Mouse urine	MALDI MS/MS	4-plex DiLeu	55	9
Mouse urine	nanoLC-MS/MS	10-plex DiLeu	2108	510
MCF7 cells	nanoLC-MS/MS	12-plex DiLeu	3510	849
PANC1 cells	nanoLC-MS/MS	Duplex DiLeu	5456	861

[a] Quantified features often contain labeling artifacts that cannot be identified as metabolites.

Supplemental Information

A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Precursor (M+H)	Charge	MW	Time	RP1	RP2	RP3	RP4	RP5	RP6	RP7	RP8	RP9	RP10	Ratio RP1	Ratio RP2	Ratio RP3	Ratio RP4	Ratio RP5	Ratio RP6	Ratio RP7	Ratio RP8	Ratio RP9	Ratio RP10
2	221.1734	1	75.0376	5.40	1.E+07	1.E+07	1.E+07	1.E+07	8.E+06	3.E+07	3.E+07	3.E+07	8.E+07	8.E+07	1	0.93	0.78	0.91	0.54	2.19	1.97	2.12	5.80	5.55
3	251.1782	1	105.0423	5.41	2.E+06	2.E+06	2.E+06	2.E+06	2.E+06	4.E+06	3.E+06	4.E+06	1.E+07	1.E+07	1	0.96	0.84	0.79	0.75	1.80	1.59	1.84	4.65	4.67
4	251.1845	1	105.0487	5.41	2.E+06	2.E+06	2.E+06	2.E+06	2.E+06	4.E+06	3.E+06	4.E+06	1.E+07	1.E+07	1	0.82	0.79	0.70	0.72	1.87	1.56	1.89	4.61	4.76
5	613.1592	2	321.8954	5.50	3.E+04	2.E+04	2.E+04	3.E+04	4.E+04	6.E+04	5.E+04	6.E+04	2.E+05	1.E+05	1	0.73	0.79	0.81	1.15	1.77	1.68	1.76	4.73	3.72
6	219.1555	1	73.0196	5.50	2.E+06	2.E+06	1.E+06	2.E+06	1.E+06	9.E+04	5.E+04	7.E+04	1.E+05	9.E+04	1	0.85	0.71	0.96	0.60	0.05	0.03	0.04	0.08	0.05
7	280.1973	1	134.0614	5.51	1.E+05	1.E+05	1.E+05	1.E+05	1.E+05	3.E+05	2.E+05	3.E+05	7.E+05	6.E+05	1	0.95	1.03	1.04	1.07	2.33	1.96	2.65	6.50	5.66
8	809.2445	2	517.9806	5.51	3.E+04	4.E+04	3.E+04	3.E+04	4.E+04	6.E+04	7.E+04	1.E+05	2.E+05	2.E+05	1	1.19	0.97	1.06	1.36	2.04	2.39	3.27	7.16	7.34

B

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Number	Average Precursor (M+H)	Charge	Average MW	Average Time	Average _Ratio_ RP1	Average _Ratio_ RP2	Average _Ratio_ RP3	Average _Ratio_ RP4	Average _Ratio_ RP5	Average _Ratio_ RP6	Average _Ratio_ RP7	Average _Ratio_ RP8	Average _Ratio_ RP9	Average _Ratio_ RP10	RSD_ Ratio _RP1	RSD_ Ratio _RP2	RSD_ Ratio _RP3	RSD_ Ratio _RP4	RSD_ Ratio _RP5	RSD_ Ratio _RP6	RSD_ Ratio _RP7	RSD_ Ratio _RP8	RSD_ Ratio _RP9	R R _ 0
2	1	221.1733	1	75.0374	5.48	1	0.95	0.72	0.90	0.52	2.13	1.95	2.06	5.59	5.48	0	0.04	0.09	0.06	0.02	0.02	0.03	0.07	0.03
3	2	251.1781	1	105.0423	5.57	1	0.86	0.79	0.69	0.73	1.68	1.51	1.76	4.26	4.40	0	0.09	0.08	0.11	0.04	0.07	0.07	0.07	0.09
4	3	251.1844	1	105.0486	5.41	1	0.86	0.75	0.69	0.71	1.75	1.56	1.77	4.40	4.60	0	0.03	0.04	0.04	0.02	0.06	0.01	0.05	0.03
5	4	193.1983	1	47.0625	5.47	1	0.14	5.27	43.52	46.35	15.43	13.27	112.67	15.75	23.11	0	0.18	0.05	0.03	0.03	0.05	0.03	0.02	0.07
6	5	231.1706	1	85.0347	5.60	1	1.33	0.81	0.84	1.18	1.47	1.64	1.72	4.70	4.25	0	0.15	0.14	0.08	0.27	0.05	0.13	0.20	0.09
7	6	337.0866	1	190.9508	5.43	1	2.56	2.23	2.67	3.86	5.13	4.29	6.02	10.47	11.75	0	0.77	0.87	0.82	0.91	0.77	0.86	0.81	0.69
8	7	295.1153	1	148.9795	5.63	1	1.09	1.30	1.41	0.79	3.51	3.53	4.00	7.33	10.30	0	0.07	0.07	0.05	0.07	0.01	0.01	0.05	0.04
9	8	190.1782	1	44.0423	5.69	1	0.39	1.26	0.20	0.20	0.24	0.19	0.22	0.32	0.33	0	0.08	0.02	0.01	0.02	0.01	0.05	0.04	0.03
10	9	216.1598	1	70.0239	5.49	1	3.78	2.02	1.79	2.86	2.76	2.49	2.82	6.23	5.62	0	0.36	0.41	0.28	0.32	0.19	0.27	0.22	0.29

C

	A	B	C	D	E	F	G	H	I	J	K	L
1	Input Number	Average Precursor (M+H)	Charge	Average MW	Average Time	dppm	name	chemical_ formula	monisotopic _MW	kegg_id	pubchem_ compound _id	hmdb_id
2	20	301.2075	1	155.0717	0.35	14.03	L-Histidine	C6H9N3O2	155.0695	C00135	6274	HMDB00177
3	20	301.2075	1	155.0717	0.35	11.91	4-Phenylpyridine	C11H9N	155.0735	C11310	13651	HMDB33123
4	32	235.1831	1	89.0473	0.56	4.29	L-Alanine	C3H7NO2	89.0477	C00041	5950	HMDB00161
5	32	235.1831	1	89.0473	0.56	4.29	Sarcosine	C3H7NO2	89.0477	C00213	1088	HMDB00271
6	32	235.1831	1	89.0473	0.56	4.29	Beta-Alanine	C3H7NO2	89.0477	C00099	239	HMDB00056
7	32	235.1831	1	89.0473	0.56	4.29	D-Alanine	C3H7NO2	89.0477	C00133	71080	HMDB01310

Supplemental Figure S1. Screenshots of the example result tables. A is the quantification result of individual file consisting of precursor ion information and reporter ion intensities and ratios. B is the merge table. For each merged feature, average reporter ion ratios and relative standard deviations (RSD) of ratios are calculated across all input data files, followed by the original reporter ion intensities and ratios from each data file. C is the metabolite identification table with ID numbers from KEGG, PubChem and HMDB databases.

Supplemental Table S1. Reporter ion information of the custom isobaric DiLeu reagents.

DiLeu reagent	Mass shift	Reporter ion accurate mass[a]	Purity
Duplex	145.1302	115.1252	0.9569
		118.1532	0.9631
4-plex	145.1302	115.1252	0.9236
		116.1407	0.9125
		117.1377	0.9786
		118.1532	0.9607
12-plex[b]	145.1280	115.1252	0.9405
		115.1315	0.9750
		116.1286	1.0990
		116.1349	0.9983
		116.1407	0.9390
		117.1318	0.9178
		117.1377	1.0019
		117.144	1.0270
		118.1352	1.0156
		118.1411	1.0575
		118.1473	1.0462
		118.1532	0.9823

[a] Reporter ion accurate masses are the experimental *m/z* values not theoretical *m/z* values.

[b] 10-plex DiLeu experiments used the first 10 reagents of the 12-plex DiLeu.

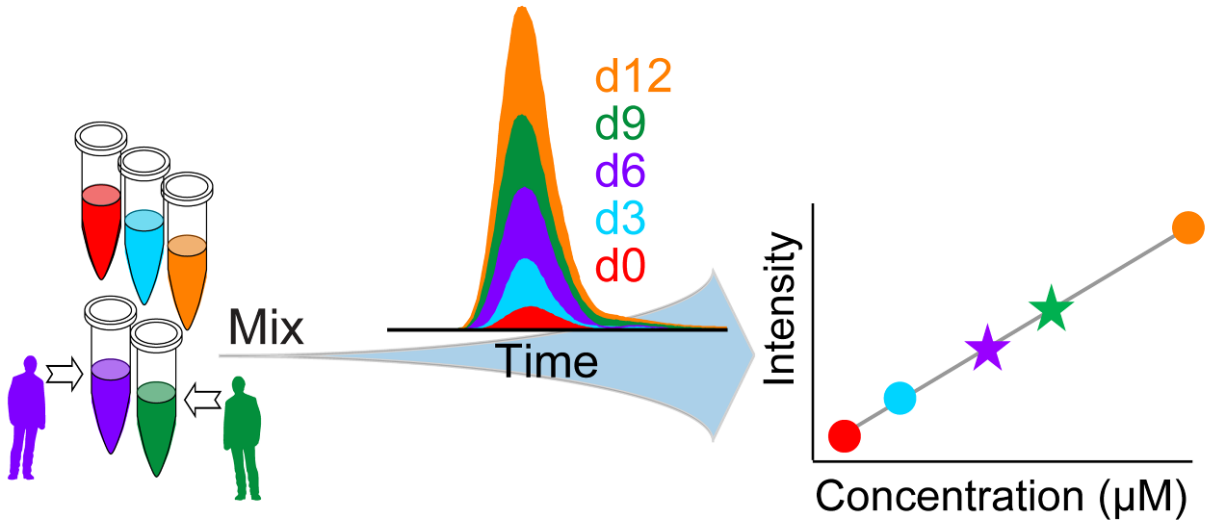
Supplemental Table S2. Reporter ion information of commercial isobaric labels.

Commercial reagent	Mass shift[a]	Reporter ions (HCD)	Reporter ions (ETD)	Commercial reagent	Mass shift[a]	Reporter ions (HCD)
TMT 10-plex	229.1629	126.1277	114.1277	iTRAQ 4-plex	144.1024	114.1112
		127.1248	115.1248			115.1083
		127.1311	114.1277			116.1116
		128.1281	115.1248			117.1150
		128.1344	116.1344	iTRAQ 8-plex	304.2022	113.1078
		129.1315	117.1315			114.1112
		129.1378	116.1344			115.1082
		130.1348	117.1315			116.1116
		130.1411	118.1411			117.1149
		131.1382	119.1382			118.1120
AminoxyTMT 6-plex	301.2317	126.1277	114.1277			119.1153
		127.1248	115.1248			121.1220
		128.1344	116.1344			
		129.1315	117.1315			
		130.1411	118.1411			
		131.1382	119.1382			

[a] Mass shift is the average accurate mass shift of the complete set of tags.

Appendix III

Absolute Quantification of Amine Metabolites in Human Cerebrospinal Fluid via MS1-centric Isotopic DiLeu (iDiLeu) Labeling



Adapted from L. Hao**, P. Wei**, J. Wang, T. Greer, O. Okonkwo, L. Li. "Absolute Quantification of Amine Metabolites in Human Cerebrospinal Fluid via MS1-centric Isotopic N,N-Dimethyl-Leucine (iDiLeu) Labeling" *To be submitted*. PW conducted labeling experiments and data analysis.

Abstract

Quantitative measurement of metabolites is essential to understanding biological and physiological processes. Absolute quantification by spiking isotopic internal standards and analyzing on mass spectrometry (MS) platform is a key method to determine the concentration of metabolites in biological samples. However, MS-based absolute quantification is often limited by the commercial availability and high costs of isotopic internal standards. Here, we evaluated the use of custom 5-plex *N*,*N*-dimethyl-leucine (iDiLeu) labeling technique for absolute quantification of amine metabolites. Metabolites labeled by iDiLeu demonstrated improved detection sensitivity and chromatographic separation compared to the routine label-free approach on a standard flow reverse phase LC-MS platform. Accurate and high-throughput quantification was achieved on the MS1 level with the ability to construct calibration curves in a single LC-MS run. The application to human CSF samples resulted in the identification of 87 amine metabolites which provided 40 additional amine metabolites to expand the current human CSF metabolome database. The concentrations of ten amine metabolites were determined in 17 human CSF samples ranging from 5.16 to 109.09 μM.

Introduction

In biological systems, small molecule metabolites are downstream end products of transcriptional and translational processes and can directly reflect enzymatic activities at the time of sampling [1]. Mass spectrometry (MS) has become a central technology for metabolite analysis with advantages that offer sensitive, accurate, and reproducible measurements in a wide dynamic range, yielding important insights into biological processes [2]. MS-based metabolite quantification can be categorized into relative and absolute quantifications. Relative quantification through label-free or chemical derivatization compares the relative abundances of metabolites in different biological states, and has been successfully applied to the discovery of candidate metabolic biomarkers of various human diseases [3-6]. Absolute quantification determines the absolute concentration of metabolites, typically achieved with an isotopic analogue of the analyte (^{2}H- or ^{13}C- analogues) spiked into samples as an internal standard to construct a calibration curve [7,8]. Absolute quantification is usually employed in targeted assays, which can be applied to verify candidate biomarkers generated from discovery studies [4,9,10].

However, absolute quantification of metabolites is always limited by the commercial availability and high costs of isotopic internal standards. Alternative approaches have been developed in recent years via chemical derivatization, particularly stable isotope labeling [11,12]. Stable heavy isotopes can be differentially incorporated into analytes to allow simultaneous comparison of multiple samples and the construction of calibration curves without the need of expensive isotopic internal standards. Quantification can be achieved by MS2-centric isobaric labeling or mass difference labeling on the MS1 level [13]. MS1-centric mass difference labeling is often more accurate since it is virtually free from precursor isolation interferences as with isobaric labeling [14,15]. Yet isobaric labeling provides greater multiplexing capability than mass

difference labeling, which is an important feature in terms of constructing multi-point calibration curves.

In order to develop an MS1-centric mass difference labeling technique with greater multiplexing capability, our research group has designed and synthesized the 5-plex isotopic *N*,*N*-dimethyl leucine (iDiLeu) reagents for the absolute quantification of peptides/proteins [16]. Each iDiLeu reagent was synthesized with simple synthetic route, high yield, and greatly reduced cost compared to commercial mTRAQ labels [17]. The amine reactive group in iDiLeu label can target the N-terminus and lysine side chains of peptides which should also be applicable to amine-containing small molecules. Therefore, we aim to expand the 5-plex iDiLeu labeling technique to the analysis of amine-containing metabolites in the present study (Scheme 1). Proof-of-principle experiments were conducted with metabolite standard mixture on a standard flow LC-MS platform which is routinely used for label-free metabolomics studies. We also applied iDiLeu labeling method to the absolute quantification of amine metabolites in human cerebrospinal fluid samples, where two iDiLeu channels can be used to label CSF samples and the other three channels to label metabolite standards. Thus, three-point calibration curve and multiplexed quantification can be efficiently achieved on a single LC-MS run.

Experimental

Human cerebrospinal fluid and standard sample preparation

Human CSF samples were collected from consenting human subjects in the Wisconsin Alzheimer's Disease Research Center (ADRC), approved by the University of Wisconsin Institutional Review Board. For each human subject, lumbar puncture was performed at L3/4 or

L4/5 following local anesthesia after 12-hour fast, and CSF sample was collected into a tube, gently mixed, and centrifuged at 2000 g for 10 min. Supernatant CSF sample was collected, stored at -80 °C, and thawed on ice before sample preparation. Metabolite fraction of CSF sample was obtained using 3 kDa molecular weight cut-off (MWCO) ultracentrifugation (Millipore Amicon Ultra, MA). Forty microliters of each CSF metabolite fraction were freeze dried and stored at -80 °C until the labeling experiment.

Metabolite standard mixture was prepared by mixing 14 metabolite standard stock solutions, including glycine, alanine, GABA, creatinine, valine, isoleucine, leucine, lysine, methionine, dopamine, phenylalanine, serotonin, tyrosine, and tryptophan. Serial dilutions of metabolite standard mixture were conducted for the construction of label-free standard curves. Several aliquots of metabolite standard mixture were dried down. All metabolite standard samples were stored at -20 °C prior to the labeling experiment.

iDiLeu synthesis and labeling reactions

The synthesis procedure of five-plex iDiLeu reagents was described previously [16]. One milligram of each iDiLeu reagent was activated to triazine ester form before labeling reaction [16]. Dried CSF metabolite or metabolite standard sample was re-dissolved in 0.5 M triethylammonium bicarbonate (TEAB) solution and labeled with 20 molar-fold excess of activated iDiLeu. Anhydrous DMF was added to reach 70% of organic: aqueous ratio and the labeling reaction was maintained for 2 h with vortexing at room temperature. The labeling reaction was quenched with 0.25% hydroxylamine (v/v), and each labeled sample was dried in *vacuo* separately and combined afterwards.

LC-ESI-MS analysis

Ultra-performance LC-MS analysis was conducted using a Dionex UltiMate 3000 LC system coupled with a Q-Exactive™ Orbitrap mass spectrometer (San Jose, CA). A Phenomenex C18 column (2.1 × 100 mm, 1.7 μm, 100 Å) was used for metabolite separation at a column temperature of 30 °C and a flow rate of 0.3 ml/min. Mobile phase A was 0.1% formic acid in optima water and mobile phase B was 0.1% formic acid in optima acetonitrile. For label-free metabolite samples, a 15 min gradient was set as the following: 0-5 min, 0-3% solvent B; 5-10 min, 3-20% B; 10-12 min, 20-80% B; 12-13min, 80% B; 13-15 min, 0% B. Full MS was acquired in selected ion mode (SIM) with an inclusion list of fourteen standards at a resolution of 70 K, an automatic gain control (AGC) of 1×10^6 and a maximum injection time (IT) of 100 ms. For iDiLeu labeled metabolite samples, a 23 min gradient was as the following: 0-5 min, 3% solvent B; 5-16 min, 3-30% B; 16-17 min, 30-80% B; 17-19min, 80% B; 19-23 min, 3% B. The first 5 min of LC flow was diverted to waste to remove impurities. Full MS scanned from *m/z* 200-1000 at a resolution of 70 K, a AGC of 1×10^6 and a maxIT of 100 ms.

Data analysis

Raw data files were acquired by Thermo Scientific Xcalibur software and then uploaded to SIEVE™ software for data processing. Peak alignment and framing algorithm was selected with a frame time width of 2 min and *m/z* width of 5 ppm. The maximum retention time shift was 0.2 min for peak alignment. ICIS algorithm was used for peak detection. A database lookup file (.csv) was created containing the accurate masses and names of 14 metabolite standards. Targeted quantification and identification of metabolites were achieved in SIEVE using the database lookup file with a mass tolerance of 5 ppm. For human CSF samples, the complete component table was downloaded, and the molecular weight of each compounds was calculated based on the charge and mass shift caused by iDiLeu labeling in an Excel sheet. MetaboSearch [18]

software was then used for metabolite identification (mass error < 5 ppm) by searching against multiple online databases: Human Metabolome Database (HMDB), Madison Metabolomics Consortium Database (MMCD), Metlin, and LIPID MAPS. Five channels of iDiLeu labeled metabolite samples were separately analyzed by LC-MS to calculate the purity of each iDiLeu label, and purity correction was performed in an Excel sheet following a previously described method [19].

Results and discussion

Characterization of iDiLeu labeling

Five-plex iDiLeu labeling was originally designed for protein/peptide quantification as MS1-centric mass difference reagents [16]. For the first time, we expanded this technique to the analysis of amine-containing metabolites. The general structures of activated iDiLeu reagents are shown in Scheme 1. After the labeling reaction, mass shifts of 141.1154, 144.1313, 147.1409, 150.1631, and 153.1644 were introduced into metabolites by the d0, d3, d6, d9, and d12 labels. A mixture of 14 metabolite standards was used to characterize the performance of iDiLeu labeling. The accurate masses before and after labeling were listed in Table 1. Peak areas of extracted ion chromatograms (EIC) of the labeled metabolites were used for quantification. As discussed previously for DiLeu labeled metabolites, hydrophobicity of the molecule was modified by labeling which can improve the separation of polar metabolites and enable the detection of polar metabolites in nanoLC-MS [19,20]. Since metabolomics analysis is typically performed on a standard flow LC system, here, we aimed to evaluate the performance of iDiLeu

labeled metabolites on a standard flow RPLC-MS platform and its comparison with the label-free approach.

Both label-free and d0-labeled metabolite standard mixture were analyzed on a reveser phase (RP)LC-MS for comparison. The chromatographic separations were illustrated in Figure 1. Free metabolites eluted within the first few minutes, particularly small polar metabolites like glycine, alanine, GABA, lysine, and creatinine were barely retained on the column and eluted before 1 min. Free isoleucine and leucine were not completely separated in this label-free platform (Figure 1A). Optimization of LC gradient was not helpful to separate these early eluting metabolites. However, after iDiLeu labeling, polar metabolites can be better retained on the RPLC column and were well separated from each other with modified hydrophobicity and increased molecular size (Figure 1B). The isomers (leucine and isoleucine) were also completely separated with retention time of 9.7 min and 10.7 min, respectively.

In order to compare the detection sensitivity for iDiLeu labeling vs. label-free approach, metabolite samples before and after iDiLeu labeling were serially diluted to construct calibration curves and determine their limits of detection. The results are listed in Table 2. LODs and LOQs of most metabolite standards were greatly improved after iDiLeu (d0) labeling. Although the improvements of metabolite detection sensitivity were not as significant as using nanoLC-MS or CE-MS system [19,20], the present results demonstrated the utility of iDiLeu labeling on the routine standard flow RPLC-MS platform with the advantages of better metabolite separation and creating calibration curves in a single run.

Quantitative performance of iDiLeu labeling

Five-plex iDiLeu labeled metabolites created five distinct peaks, differing from one another by 3, 6, 9, and 12 Da. An example iDiLeu labeled serotonin standard was illustrated in Figure 2. Because of the subtle mass difference of 3 Da, iDiLeu labeling does not suffer from isotope interference for metabolite analysis. Deuterium effect was minimized by grouping deuterium atoms around polar amine group of leucine [16,21], so that five labeled peaks can elute at roughly the same time. EIC peak areas of five isotopic forms of the same labeled metabolite can be compared for quantification. Quantitative accuracy of iDiLeu labeling was demonstrated by labeling 14 metabolite standards and mixing five channels at theoretical 1:1:1:1:1 and 1:2:5:8:10 ratios. As shown in Figure 3, satisfactory accuracies were achieved with an average accuracy < 12% error. Median ratios measured among 14 metabolites (three replicates each) were 1:1.05:0.98:1.02:1.10 and 1:2.06:4.49:7.11:9.64, respectively.

Absolute quantification of a molecule is typically achieved by constructing a calibration curve where quantification accuracy is highly dependent on the curve's linearity (also called correlation coefficient). The linearity of iDiLeu-based quantification was characterized by two approaches. Firstly, iDiLeu labeled metabolite standards were serially diluted and calibration curves were constructed across multiple LC-MS run; Alternatively, five channels of iDiLeu labeled metabolite standards were mixed at different ratios across two order of magnitude, and calibration curves were established in a single LC-MS run. We compared these two methods with the label-free approach and the correlation coefficients of 14 metabolites standards are provided in Table 2. The linearity using serially diluted labeled metabolites was comparable with free metabolite standards (R> 0.99) except for creatinine, while the single-run calibration curves demonstrated the best correlation coefficient (R>0.999) because of the ability to generate calibration curves in a single LC-MS injection to avoid run-to-run variations.

The Analysis of human CSF

CSF is a valuable biofluid sample circulating within the brain ventricular system. It protects the brain and maintains the metabolic homeostasis of the central nervous system [22,23]. Amino acids in CSF samples serve as the basic building blocks and key regulators in CSF metabolism, which has been found to be involved in various neurological diseases, such as Alzheimer's disease and Parkinson's disease [5,24,25]. The established absolute quantification method using iDiLeu labeling can be applied to determine the concentrations of amine-containing metabolites in CSF samples. By using three channels of iDiLeu to label standards and the remaining channels to label two human CSF samples, we can construct a three-point calibration curve and determine the metabolite concentration in two samples on a single MS scan. Because of the good chromatographic separation of iDiLeu labeled metabolites, absolute quantifications of many amine metabolites can be achieved in a single LC-MS run. To minimize the influence of variations from five iDiLeu tags, we carefully designed the experiment to rotate different iDiLeu channels for the labeling of standards and CSF samples (Supplemental Table S1). Eight labeling experiments were conducted for the measurement of 16 human CSF samples.

Metabolite identification was achieved in Metabosearch software [18] by searching against multiple online databases with 5 ppm mass tolerance. The resulted list of metabolites was then manually examined to only include primary and secondary amine-containing molecules. A total of 87 amine metabolites were identified from human CSF samples (Supplemental Table S2). We then matched the metabolites with the human CSF metabolome database [22] and found that 47 of them are documented in the database and 40 additional amine metabolites were identified in our study. Since iDiLeu tags can only target primary and secondary amine groups, the confidence of metabolite identification was also enhanced by the presence of five isotopic peaks. With the

excellent linearity and quantification accuracy of iDiLeu labeling, we determined the concentrations of 10 amine metabolites in 16 human CSF samples, including lysine, tryptophan, valine, phenylalanine, tyrosine, methionine, creatinine, alanine, isoleucine, and leucine. As shown in Figure 4, concentrations of the 10 measured molecules ranged from 5.16 to 109.09 μM. Creatinine is the most abundant compounds with an average concentration of 109.09 ± 30.77 μM. Example calibration curves were shown in Supplemental Figure S1.

Conclusions

An accurate and high throughput absolute quantification strategy was developed for amine metabolites using the custom 5-plex iDiLeu labeling. The labeling technique was characterized on a standard flow RPLC-MS platform for the first time and achieved enhanced metabolite separation and detection sensitivity compared with routine label-free approach. Accurate absolute quantification was achieved on the MS1 level with the ability to construct calibration curves in a single LC-MS run. We then successfully applied this method to the identification and quantification of amine metabolites in human CSF samples. We believe that iDiLeu labeling based absolute quantification will be a useful tool to measure the concentration of amine metabolites in biological samples, serving as a promising validation platform for disease biomarker studies.

Acknowledgements

We would like to thank Wisconsin Alzheimer's Disease Research Center for providing the CSF samples and related study data. This work was supported in part by National Institutes

of Health through grant 1P50AG033514. The Q-Exactive Orbitrap instrument was purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531). We also wish to thank Molly Pellitteri-Hahn and Cameron O. Scarlett in the Analytical Instrumentation Center in School of Pharmacy for instrument access and helpful suggestions.

References

(1) Nicholson, J. K.; Lindon, J. C. *Nature* **2008**, *455*, 1054-1056.

(2) Villas-Boas, S. G.; Mas, S.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spectrometry Reviews* **2005**, *24*, 613-646.

(3) Ganti, S.; Weiss, R. H. *Urologic Oncology-Seminars And Original Investigations* **2011**, *29*, 551-557.

(4) Hao, L.; Greer, T.; Page, D.; Shi, Y. T.; Vezina, C. M.; Macoska, J. A.; Marker, P. C.; Bjorling, D. E.; Bushman, W.; Ricke, W. A.; Li, L. J. *Sci Rep-Uk* **2016**, *6*.

(5) Trushina, E.; Dutta, T.; Persson, X. M. T.; Mielke, M. M.; Petersen, R. C. *Plos One* **2013**, *8*.

(6) Peng, J.; Chen, Y. T.; Chen, C. L.; Li, L. *Anal Chem* **2014**, *86*, 6540-6547.

(7) Carter, M. D.; Calcutt, M. W.; Malow, B. A.; Rose, K. L.; Hachey, D. L. *Journal Of Mass Spectrometry* **2012**, *47*, 277-285.

(8) Pauwels, S.; Antonio, L.; Jans, I.; Lintermans, A.; Neven, P.; Claessens, F.; Decallonne, B.; Billen, J.; Vanderschueren, D.; Vermeersch, P. *Analytical And Bioanalytical Chemistry* **2013**, *405*, 8569-8577.

(9) Zheng, P.; Wang, Y.; Chen, L.; Yang, D. Y.; Meng, H. Q.; Zhou, D. Z.; Zhong, J. J.; Lei, Y.; Melgiri, N. D.; Xie, P. *Molecular & Cellular Proteomics* **2013**, *12*, 207-214.

(10) Wettersten, H. I.; Weiss, R. H. *Organogenesis* **2013**, *9*, 11-18.

(11) Guo, K.; Li, L. *Anal Chem* **2009**, *81*, 3919-3932.

(12) Yuan, W.; Anderson, K. W.; Li, S. W.; Edwards, J. L. *Anal Chem* **2012**, *84*, 2892-2899.

(13) Chokkathukalam, A.; Kim, D. H.; Barrett, M. P.; Breitling, R.; Creek, D. J. *Bioanalysis* **2014**, *6*, 511-524.

(14) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C. *J Proteome Res* **2009**, *8*, 5347-5355.

(15) Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J. *Nat Methods* **2013**, *10*, 332-334.

(16) Greer, T.; Lietz, C.; Xiang, F.; Li, L. *Journal of The American Society for Mass Spectrometry* **2015**, *26*, 107-119.

(17) DeSouza, L. V.; Taylor, A. M.; Li, W.; Minkoff, M. S.; Romaschin, A. D.; Colgan, T. J.; Siu, K. W. *J Proteome Res* **2008**, *7*, 3525-3534.

(18) Zhou, B.; Wang, J. L.; Ressom, H. W. *Plos One* **2012**, *7*.

(19) Hao, L.; Zhong, X. F.; Greer, T.; Ye, H.; Li, L. J. *Analyst* **2015**, *140*, 467-475.

(20) Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L. J. *Anal Chem* **2017**, *89*, 1138-1146.

(21) Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. *Anal Chem* **2002**, *74*, 3662-3669.

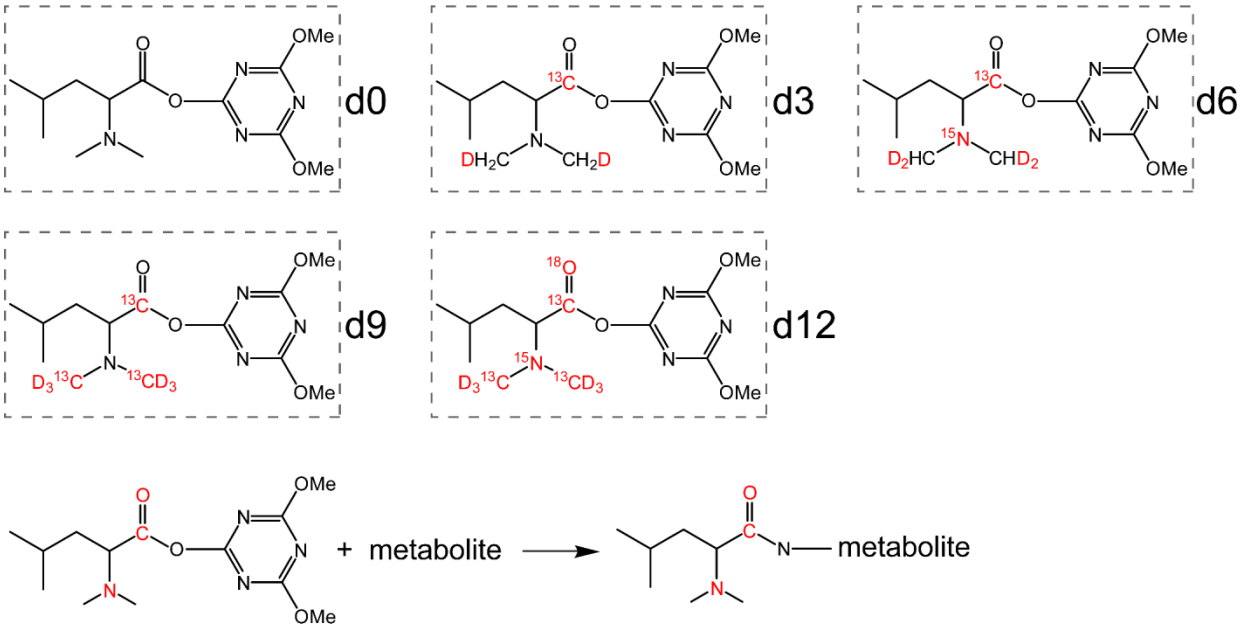
(22) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y. P.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. *J Chromatogr B* **2008**, *871*, 164-173.

(23) Mandal, R.; Guo, A. C.; Chaudhary, K. K.; Liu, P.; Yallou, F. S.; Dong, E.; Aziat, F.; Wishart, D. S. *Genome Med* **2012**, *4*.

(24) Oresic, M.; Hyotylainen, T.; Herukka, S. K.; Sysi-Aho, M.; Mattila, I.; Seppanan-Laakso, T.; Julkunen, V.; Gopalacharyulu, P. V.; Hallikainen, M.; Koikkalainen, J.; Kivipelto, M.; Helisalmi, S.; Lotjonen, J.; Soininen, H. *Translational Psychiatry* **2011**, *1*.

(25) Zhang, J.; Sokal, I.; Peskind, E. R.; Quinn, J. F.; Jankovic, J.; Kenney, C.; Chung, K. A.; Millard, S. P.; Nutt, J. G.; Montine, T. J. *American Journal Of Clinical Pathology* **2008**, *129*, 526-529.

Figures and Tables



Scheme 1. General structure of iDiLeu reagents (d0, d3, d6, d9, and d12) in their activated triazine ester forms, and the reaction to label amine-containing metabolites.

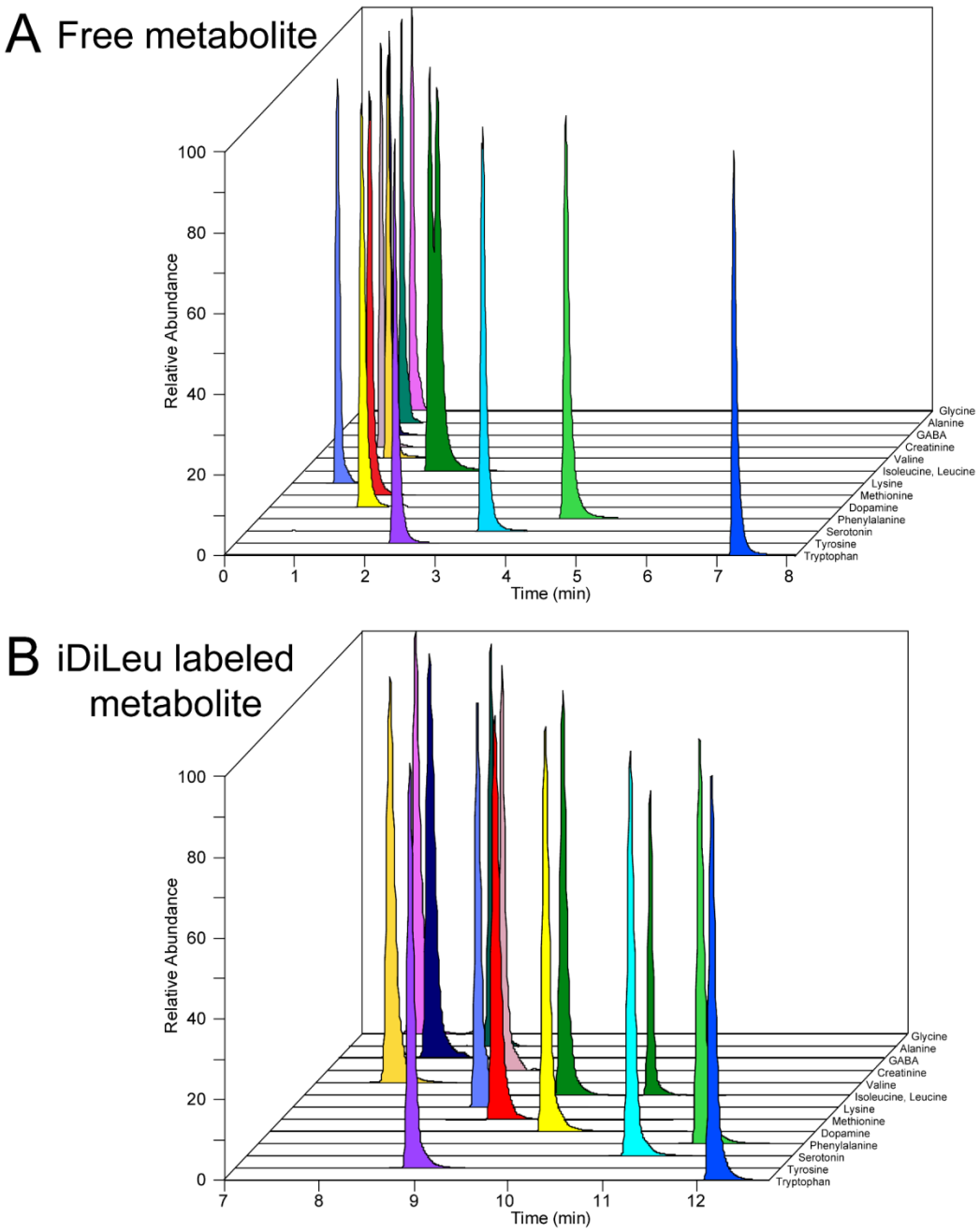


Figure 1. Chromatographic comparison of label-free metabolites (A) and iDiLeu (d0) labeled metabolite standards (B) via standard flow UPLC-MS platform.

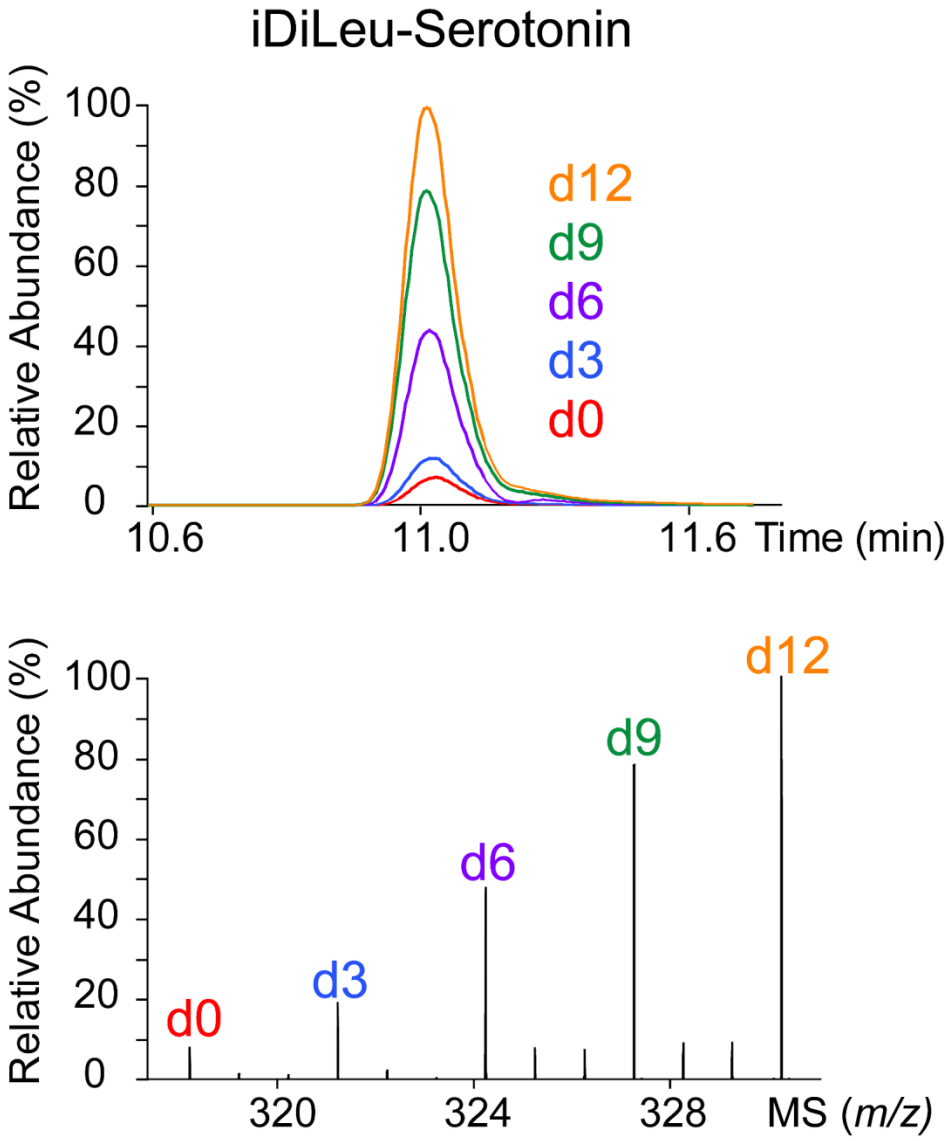


Figure 2. Example chromatography and MS spectrum of 5-plex iDiLeu labeled serotonin standard. Five-plex iDiLeu labeled serotonin samples were mixed at 1:2:5:8:10 ratio.

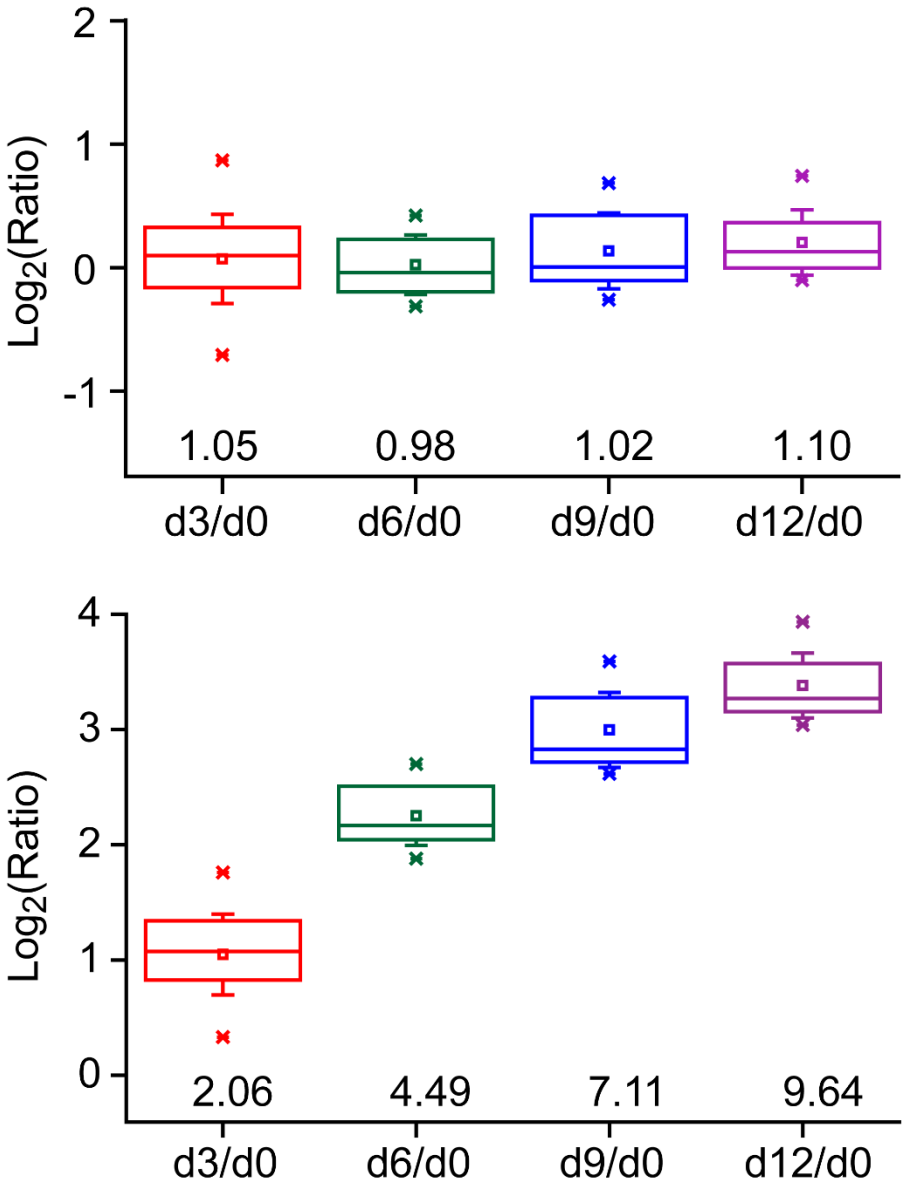


Figure 3. Quantification accuracy of iDiLeu labeling. Each box contains results from 14 metabolite standards labeled by iDiLeu. Theoretical mixing ratios were 1:1:1:1:1 and 1:2:5:8:10. Box denotes 25th and 75th percentiles. Line within the box denotes 50th percentile. Whisker denotes standard deviation.

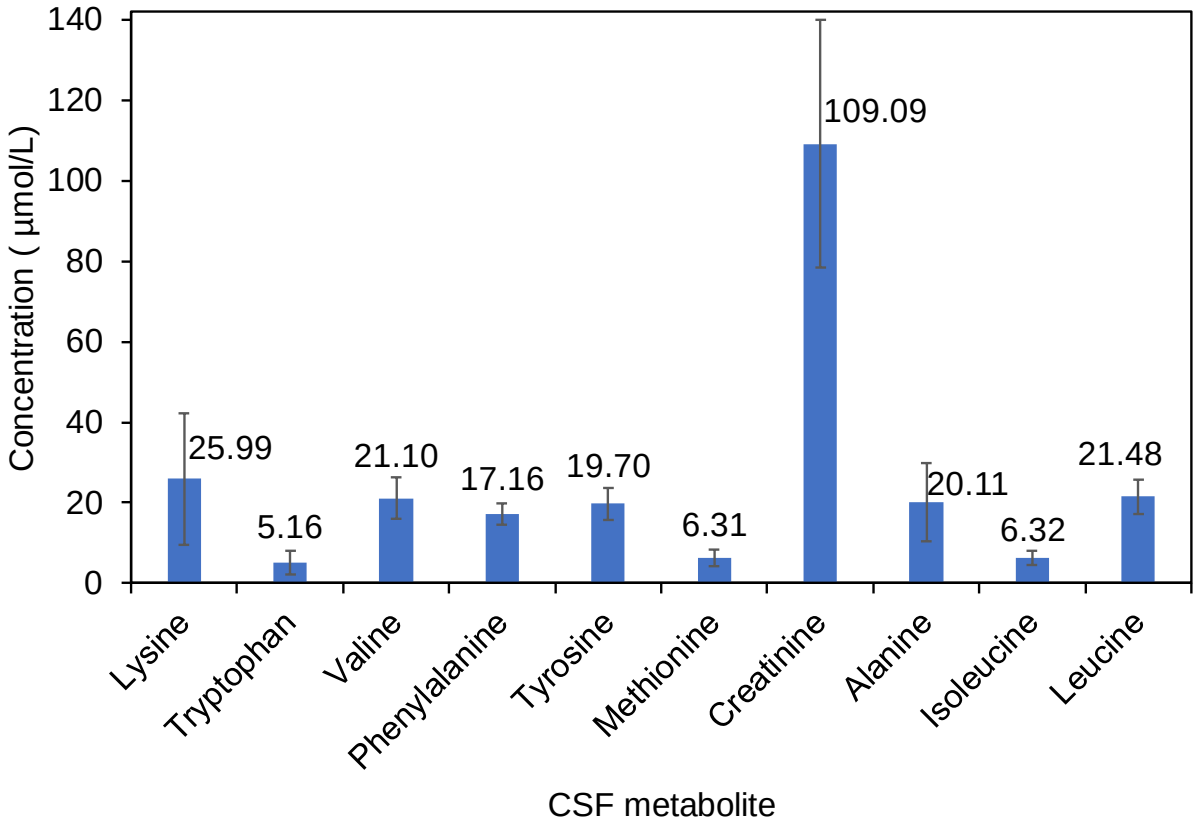


Figure 4. Metabolite concentrations in human CSF samples, measured by iDiLeu labeling technique.

Table 1. Accurate masses of 14 metabolite standards before and after iDiLeu labeling.

Metabolite	M+H	d0 labeled	d3 labeled	d6 labeled	d9 labeled	d12 labeled
Glycine	76.03983	217.1552	220.1711	223.1807	226.2029	229.2042
Alanine	90.05553	231.1709	234.1868	237.1964	240.2186	243.2199
GABA	104.0711	245.1865	248.2024	251.212	254.2342	257.2355
Creatinine	114.0667	255.1821	258.198	261.2076	264.2298	267.2311
Valine	118.0868	259.2022	262.2181	265.2277	268.2499	271.2512
Isoleucine	132.1024	273.2178	276.2337	279.2433	282.2655	285.2668
Leucine	132.1024	273.2178	276.2337	279.2433	282.2655	285.2668
Lysine	147.1133	429.3441	435.3759	441.3951	447.4395	453.4421
Methionine	150.0588	291.1742	294.1901	297.1997	300.2219	303.2232
Dopamine	154.0868	295.2022	298.2181	301.2277	304.2499	307.2512
Phenylalanine	166.0868	307.2022	310.2181	313.2277	316.2499	319.2512
Serotonin	177.1028	318.2182	321.2341	324.2437	327.2659	330.2672
Tyrosine	182.0817	323.1971	326.213	329.2226	332.2448	335.2461
Tryptophan	205.0977	346.2131	349.229	352.2386	355.2608	358.2621

Table 2. Retention time, sensitivity, and linearity (calibration curves) of metabolite analysis with iDiLeu labeling and label-free approaches.

	iDiLeu labeling					Label-free			
Metabolite	Time	LOD[a]	LOQ[b]	R (d0)[c]	R (5-plex)[d]	Time	LOD[a]	LOQ[b]	R
Glycine	7.5	4.20E-04	7.00E-04	0.9948	0.9996	0.72	9.53E-03	1.59E-02	0.9994
Alanine	8.5	2.17E-03	3.62E-03	0.9945	0.9990	0.7	3.45E-03	5.75E-03	0.9988
GABA	7.9	5.06E-05	8.43E-05	0.9991	0.9993	0.71	1.27E-03	2.12E-03	0.9997
Creatinine	8.8	1.13E-04	1.88E-04	0.9757	0.9998	0.74	1.20E-04	2.00E-04	0.9966
Valine	7.7	5.23E-05	8.72E-05	0.9968	0.9999	1.02	1.09E-01	1.82E-01	0.9998
Isoleucine	9.7	1.15E-04	1.92E-04	0.9807	0.9999	1.79	9.63E-04	1.61E-03	0.9995
Leucine	10.7	6.93E-05	1.16E-04	0.9956	0.9999	1.82	6.04E-04	1.01E-03	0.9997
Lysine	8.9	5.06E-05	8.43E-05	0.9995	0.9998	0.65	4.49E-04	7.48E-04	0.9998
Methionine	9.2	4.74E-05	7.90E-05	0.9969	0.9998	1.26	1.77E-04	2.95E-04	0.9999
Dopamine	9.9	1.20E-05	2.00E-05	0.9913	0.9991	1.34	4.62E-05	7.70E-05	0.9999
Phenylalanine	11.6	1.43E-03	2.38E-03	0.9985	0.9999	4.31	7.74E-05	1.29E-04	0.9998
Serotonin	11	4.16E-05	6.93E-05	0.9985	0.9991	3.24	9.51E-04	1.59E-03	0.9996
Tyrosine	8.8	4.98E-05	8.30E-05	0.9968	0.9999	2.26	3.33E-04	5.55E-04	0.9999
Tryptophan	12.1	1.18E-04	1.97E-04	0.9887	0.9999	7.34	4.29E-04	7.15E-04	0.9995

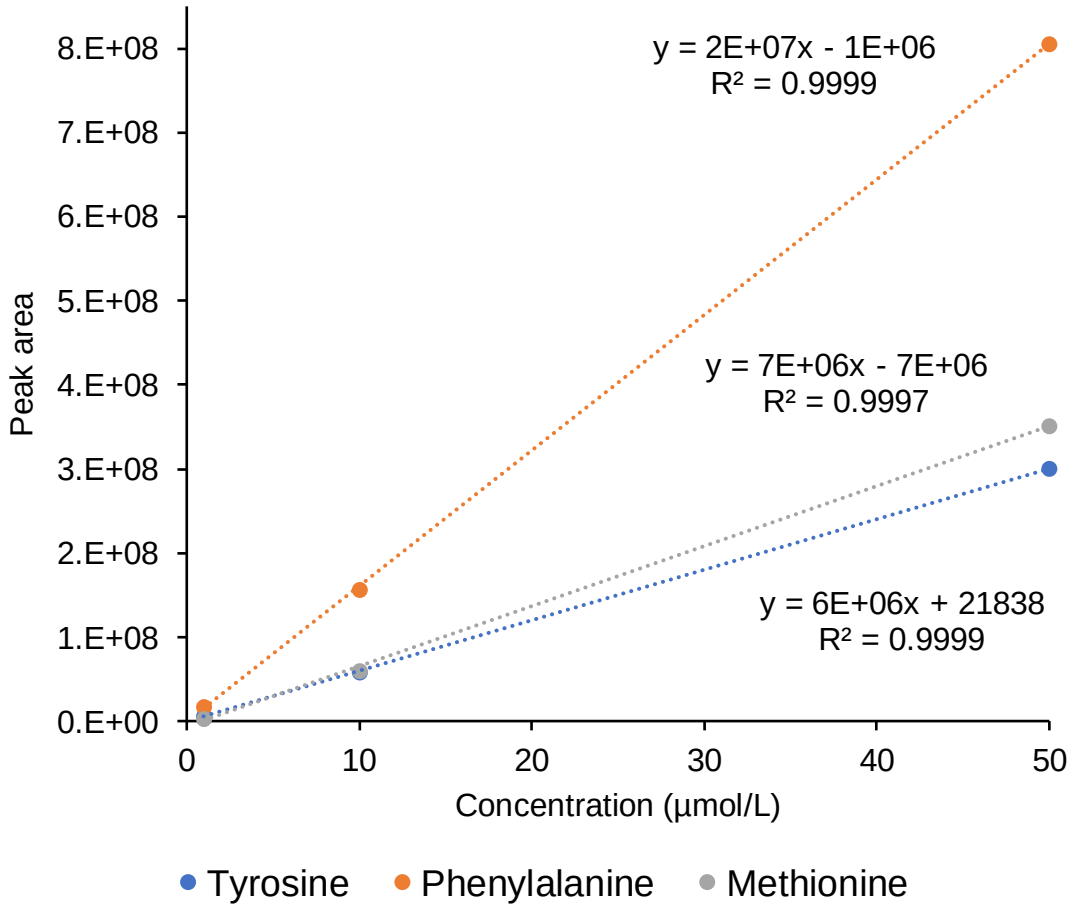
[a]Limit of detection, µmol/L;

[b]Limit of quantification, µmol/L;

[c]Linearity of calibration curve generated by serial dilution of iDiLeu d0-labeled metabolite;

[d]Linearity of calibration curve generated by five plex iDiLeu in a single LC-MS run.

Supplemental Information



Supplemental Figure S1. Example calibration curves for the absolute quantification of CSF metabolites using iDiLeu labeling. Three iDiLeu channels were used to label standard to construct the calibration curve, while the other two channels were used to label CSF samples.

Supplemental Table S1. Experimental design for the absolute quantification of amine metabolites in 16 human CSF samples using iDiLeu labeling.

	#1	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**
d0	a	49	80	c	b	c	159	169
d3	b	a	44	109	c	a	c	138
d6	c	b	a	68	114	b	a	c
d9	15	c	b	a	98	117	b	a
d12	30	27	c	b	a	144	125	b

d0, d3, d6, d9, and d12 represent five iDiLeu channels;

a, b, and c represent metabolite standard solutions with three different concentrations;

#1 to #8 represent eight labeling experiments

15, 30, 49, 27…..and other numbers inside the table represent the code of human subjects.

Supplemental Table S2. The list of detected amine metabolites in human CSF samples.

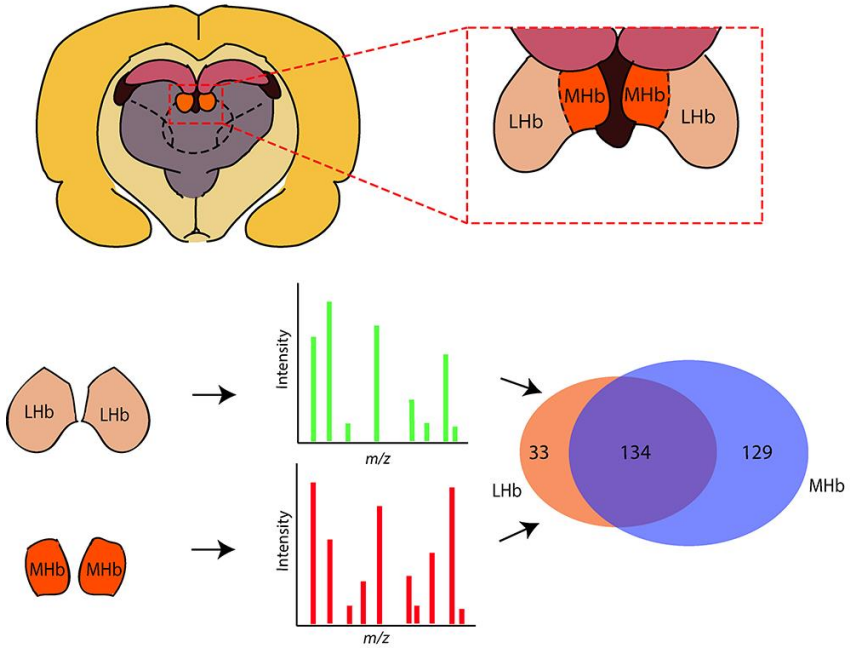
Name	MW	Detected *m/z*	Charge	Time	Δppm	HMDB ID	KEGG ID
Creatine	131.0695	207.6576	2	10.18	1.93	HMDB00064	C00300
Ornithine	132.0899	208.1678	2	8.56	0.97	HMDB00214	C01602
Glutamine	146.0691	215.1575	2	8.49	0.16	HMDB00641	C00303
Lysine	146.1055	215.1758	2	8.94	0.93	HMDB00182	C16440
Glycine	75.0320	217.1550	1	7.54	1.17	HMDB00123	C00037
Methyl-lysine	160.1212	222.1835	2	11.50	0.91	HMDB02038	C02728
Arginine	174.1117	229.1790	2	8.01	0.44	HMDB00517	C02385
Alanine	89.0477	231.1706	1	8.50	1.31	HMDB00161	C01401
GABA	103.0633	245.1861	1	7.59	1.85	HMDB00112	C00334
Serine	105.0426	247.1655	1	8.70	1.29	HMDB00187	C00716
Creatinine	113.0589	255.1817	1	8.79	1.71	HMDB00562	C00791
Valine	117.0790	259.2018	1	7.66	1.56	HMDB00883	C16436
Cystine	240.0238	262.1347	2	9.61	0.24	HMDB00192	C01420
Leucine	131.0946	273.2175	1	10.66	1.29	HMDB00687	C16439
Isoleucine	131.0946	273.2176	1	9.71	0.93	HMDB00172	C16434
Methionine	149.0510	291.1738	1	9.13	1.63	HMDB00696	C00073
Histidine	155.0695	297.1923	1	8.41	1.35	HMDB00177	C00768
Phenylalanine	165.0790	307.2020	1	11.64	0.66	HMDB00159	C02057
Methylhistidine	169.0851	311.2081	1	8.88	0.81	HMDB00479	C01152
Tyrosine	181.0739	323.1969	1	8.77	0.68	HMDB00158	C00082
Tryptophan	204.0899	346.2128	1	12.10	0.94	HMDB00929	C00078
Asparagine	132.0535	415.2917	1	13.50	1.00	HMDB00168	C16438
Kynurenine	208.0848	491.3232	1	11.57	0.44	HMDB00684	C01718
Ethanolamine	61.0528	203.1758	1	8.27	0.93	HMDB00149	C00189
Aspartic acid	133.0375	275.1602	1	7.69	1.94	HMDB06483	C16433
Homocysteine	135.0354	277.1584	1	8.05	0.81	HMDB00742	C05330
Aminobenzoic acid	137.0477	279.1706	1	13.48	1.09	HMDB01392	D02456
Guanidinobutanoic acid	145.0851	287.2078	1	9.14	1.92	HMDB03464	C01035
Glutamate	147.0532	289.1763	1	7.37	0.29	HMDB03339	C00025
Dopamine	153.0790	295.2022	1	9.77	0.01	HMDB00073	C03758

Allantoin	158.0440	300.1668	1	9.66	1.38	HMDB00462	C01551
methoxytyramine	167.0946	309.2175	1	12.39	1.14	HMDB00022	C05587
Norepinephrine	169.0739	311.1968	1	11.50	1.02	HMDB00216	C00547
Glycylproline	172.0848	314.2081	1	8.81	-0.26	HMDB00721	
Acetylaspartate	175.0481	317.1712	1	10.62	0.31	HMDB00812	
Trimethyl-lysine	188.1525	330.2751	1	10.05	1.83	HMDB01325	C03793
3-Methoxytyrosine	211.0845	353.2076	1	9.43	0.23	HMDB01434	
Sapropterin	241.1175	383.2408	1	17.50	-0.22	HMDB00787	
Cytidine	243.0855	385.2085	1	9.16	0.64	HMDB00089	C00475
Guanidoacetic acid	117.0538	400.2921	1	8.69	0.88	HMDB00128	C00581
Ureidoisobutyrate	146.0691	429.3073	1	10.72	1.09	HMDB02031	C05100
Acetylneuraminic acid	309.1060	451.2292	1	7.74	0.01	HMDB00230	
Cystathionine	222.0674	505.3059	1	13.31	0.30	HMDB00099	C00542
Succinoadenosine	383.1077	525.2312	1	17.57	2.15	HMDB00912	
Glutamyl-Lysine	275.1481	558.3863	1	9.38	0.80	HMDB28824	
Guanidinosuccinic acid	175.0593	599.4130	1	6.18	0.55	HMDB03157	C03139
Citrulline	175.0957	599.4493	1	12.04	0.69	HMDB00904	C00327
Acetyllysine	188.1161	236.1810	2	9.64	1.10	HMDB00446	
N-Acetylserotonin	218.1055	360.2286	1	12.63	0.42	HMDB01238	C00978
Aminoacetone	73.0528	215.1757	1	12.93	1.34	HMDB02134	C01888
Aminobutyraldehyde	87.0684	229.1913	1	7.62	1.48	HMDB01080	C00555
Sarcosine	89.0477	231.1705	1	6.58	1.75	HMDB00271	C00213
Cytosine	111.0433	253.1661	1	8.44	1.53	HMDB00630	C00380
Acetamidopropanal	115.0633	257.1861	1	9.96	1.76	HMDB12880	
2-Amino-3-oxobutanoate/ Aspartate-semialdehyde	117.0426	259.1653	1	6.85	2.00	HMDB06454/ HMDB12249	C03508/ C00441
Allothreonine	119.0582	261.1811	1	7.53	1.41	HMDB04041	C05519
Methylcytosine	125.0589	267.1817	1	10.33	1.64	HMDB02894	C02376
Acryloylglycine	129.0426	271.1653	1	6.58	1.91	HMDB00783	
Acetylaminobutanal	129.0790	271.2018	1	12.05	1.49	HMDB04226	C05936
Propionylglycine	131.0582	273.1810	1	10.40	1.71	HMDB00783	
5-Amino-2-oxopentanoic acid	131.0582	273.1811	1	8.90	1.35	HMDB06272	C01110
N-Methylnicotinamide	136.0637	278.1861	1	14.28	2.83	HMDB03152	
Vinylacetylglycine	143.0582	285.1810	1	9.04	1.64	HMDB00894	

Oxoglutaramate	145.0375	287.1604	1	13.67	1.16	HMDB01552	C00940
Isobutyrylglycine/ 2-Keto-6-aminocaproate	145.0739	287.1963	1	10.66	2.85	HMDB00730/ HMDB12151	
5-Amino-3-oxohexanoate	145.0739	287.1969	1	9.60	0.76	HMDB12131	
Methyl-aspartic acid	147.0532	289.1760	1	9.28	1.32	HMDB02393	C12269
Tiglylglycine/ Methylcrotonylglycine	157.0739	299.1966	1	11.10	1.73	HMDB00959/ HMDB00459	
2-Aminooctanoic acid	159.1259	301.2488	1	12.83	1.17	HMDB00991	
Aminoadipic acid	161.0688	303.1917	1	10.14	1.10	HMDB00510	
Hydroxydopamine	169.0739	311.1967	1	11.78	1.34	HMDB01537	
Hexanoylglycine	173.1052	315.2282	1	15.72	0.69	HMDB00701	
8-Hydroxy-7-methylguanine	181.0600	323.1830	1	12.96	0.62	HMDB06037	
2-Hydroxy-phenylalanine	181.0739	323.1968	1	10.21	0.98	HMDB06050	
N-Heptanoylglycine	187.1208	329.2439	1	16.64	0.51	HMDB13010	
Hydroxyphenyl-acetylglycine	209.0688	351.1919	1	11.85	0.38	HMDB00735	
Nonanoylglycine	215.1521	357.2751	1	14.48	0.75	HMDB13279	
Tyramine-O-sulfate	217.0409	359.1637	1	12.04	1.12	HMDB06409	
5-Hydroxyl-tryptophan	220.0848	362.2077	1	12.18	0.88	HMDB00472	C01017
Acetyltyrosine	223.0845	365.2072	1	13.78	1.32	HMDB00866	C01657
Aspartylhydroxyproline	246.0852	388.2082	1	7.41	0.54	HMDB28754	
Acetylvanilalanine	253.0950	395.2182	1	11.73	0.12	HMDB11716	
Tetrahydroneopterin	257.1124	399.2353	1	10.72	0.82	HMDB00942	
S-(2-Methylpropionyl)-dihydrolipoamide-E	277.1170	419.2405	1	13.56	-0.61	HMDB06868	C15977
1-Methylguanosine	297.1073	439.2304	1	10.73	0.33	HMDB01563	
Acetylasparagine	174.0641	457.3026	1	8.72	0.18	HMDB06028	
Aminohippuric acid	194.0691	477.3075	1	12.42	0.56	HMDB01867	D06890

Appendix IV

Neuropeptidomics of the Rat Habenular Nuclei



Adapted from Ning Yang, Krishna D. B. Anapindi, Stanislav S. Rubakhin, **Pingli Wei**, Qing Yu, Lingjun Li, Paul J. Kenny, Jonathan V. Sweedler. "Neuropeptidomics of the Rat Habenular Nuclei." Journal of Proteome Research, 2018. **Pingli Wei** provided Sulfation peptide analysis.

ABSTRACT

Conserved across vertebrates, the habenular nuclei are a pair of small symmetrical structures in the epithalamus. The nuclei functionally link the forebrain and midbrain by receiving input from and projecting to several brain regions. The habenular nuclei comprise two major asymmetrical subnuclei, the medial and lateral habenula; each are associated with different physiological processes and disorders, such as depression, nicotine addiction, and encoding aversive stimuli or omitting expected rewarding stimuli. Elucidating the functions of the habenula nuclei at the molecular level requires knowledge of their neuropeptide complement. In this work, three mass spectrometry (MS) techniques – liquid chromatography (LC) coupled to Orbitrap tandem MS (MS/MS), LC coupled to Fourier transform (FT)-ion cyclotron resonance (ICR) MS/MS, and matrix-assisted laser desorption/ionization (MALDI) FT-ICR MS – were used to uncover the neuropeptide profiles of the medial and lateral rodent habenula. With the assistance of tissue stabilization and bioinformatics, a total of 263 and 167 neuropeptides produced from 27 and 20 prohormones were detected and identified from the medial and lateral habenula regions, respectively. Among these neuropeptides, 129 were exclusively found in the medial habenula, and 33 were exclusively expressed in the lateral habenula. The results demonstrate that these two small brain nuclei have a rich and differentiated peptide repertoire, with this information enabling a range of follow-up studies.

KEYWORDS: medial habenula, lateral habenula, neuropeptidomics, mass spectrometry, ESI MS, MALDI MS, FT-ICR MS

INTRODUCTION

The habenula nuclei are small structures symmetrically located above the posterior end of the thalamus, functionally linking the forebrain and the midbrain. They are phylogenetically conserved in vertebrates and part of the epithalamus.[1] The habenula nucleus has two substructures, the medial and lateral subnuclei, which interconnect with different brain regions.[2] The habenula had not attracted significant attention until the discovery of its involvement in the reward system.[3-6] To date, the habenula has been found to play important roles in a wide range of physiological processes, including sleep regulation, reward-based decision-making, avoidance and mood behaviors, and drug addiction, particularly tobacco dependence.[7-10] An important step in understanding the contribution of the habenula to these biological events and behaviors is to elucidate the cell-to-cell signaling molecules, including neuropeptides, contained in this brain region. In addition, there are orphan G-protein coupled receptors (GPCRs) densely or exclusively expressed in the habenula;[11,12] characterizing the neuropeptide complement in the habenula may help to uncover the endogenous ligands for these GPCRs.

Neuropeptides are important cell-to-cell signaling molecules involved in the coordination of behavioral, cognitive, and homeostatic events, as well as various physiological functions, such as food intake, pain sensation, circadian rhythms, and tissue regeneration, including pathophysiological processes such as drug addiction.[13-17] These signaling peptides are produced by post-translational enzymatic processing of large precursor proteins known as prohormones at conventional cleavage sites, including mono/di-basic amino acid motifs, and cleavages at less canonical amino acid sequences. To become biologically functional peptides, many neuropeptides also undergo post-translational modifications (PTMs),[18,19] resulting in chemically unique peptides from the same sequence of amino acids.

The development of modern mass spectrometry (MS) techniques, aided by improvements in sample preparation, have changed the way peptide detection, identification, and quantitation are performed, allowing characterization of a fairly complete set of neuropeptides present in a biological structure.[20-24] In the past decade, hundreds of neuropeptides have been discovered and identified from a range of animals

using matrix-assisted laser desorption/ionization (MALDI) MS and electrospray ionization (ESI) MS.[15,17,25-28]

A number of *in situ* mRNA hybridization and immunocytochemistry studies have suggested the presence of several neuropeptide receptors in the habenula region, such as the orexin-2 receptor in the lateral habenula, and the neuropeptide Y (NPY) Y1 receptor in both the medial and lateral habenula.[29-31] However, there has yet to be a systematic investigation of the neuropeptide content in this small but functionally important brain region. Here, we combined sample stabilization, high molecular mass accuracy and resolution MS, and bioinformatics to characterize the endogenous peptides within the rat habenula. We identified 263 and 167 peptides from 27 and 20 prohormones in the medial and lateral habenula, respectively. Of these, 38 and 36 peptides in the medial and lateral habenula, respectively, possess PTMs, including amidation, pyroglutamylation, acetylation, disulfide bonds, and phosphorylation.

In addition to characterizing these commonly found PTMs, a set of targeted experiments were aimed exclusively at characterizing sulfated peptides. Tyrosine (Tyr)-sulfation is a well-known PTM that is involved specifically in modulating protein-protein and ligand-receptor interactions.[32-37] Though sulfation as a modification of proteins was identified in the early 1950s,[38] determination of the exact site of Tyr-sulfation using MS is still a challenging task for several reasons. First, sulfation increases the mass of the modified peptide by almost the same amount (+79.9568 Da) as phosphorylation (+79.9663 Da). Moreover, both PTMs can occur on Tyr. Secondly, sulfation is a labile PTM where the SO_3 group falls off, even with soft ionization techniques such as ESI. Finally, even if sulfation remains after ESI, it does not survive collision-induced dissociation (CID), making it difficult to determine the site of the modification. To overcome these difficulties, we used a custom approach with beam-type CID (high-energy C-trap dissociation, or HCD) and electron transfer HCD (EThcD) dissociation techniques, followed by manual deconvolution of neutral loss peaks of the parent ion corresponding to a mass difference of 80 Da. Using this approach, we confidently localized four novel sites of modification for sulfation on the secretogranin I (SCG1) prohormone in rat.

Together, these data represent the first comprehensive neuropeptidomics study of the vertebrate habenula. Information on the neuropeptide content of the habenular nuclei,

and modulation of these peptides by PTMs, can be used as a basis for better understanding of the roles of cell-to-cell signaling molecules in different biological processes.

EXPERIMENTAL SECTION

Materials

The solvents used were of LC–MS grade and obtained from Thermo Fisher Scientific (Waltham, MA), except for triethylamine, acetone, and hydrochloric acid, which were obtained from Sigma-Aldrich (St. Louis, MA).

Habenula Isolation and Stabilization

Male Sprague Dawley rats, 8–12 weeks old, were euthanized by decapitation in compliance with animal use protocols approved by the University of Illinois Institutional Animal Care and Use Committee and in accordance with all state and federal regulations. The brain was quickly removed after fast decapitation and cooled in ice cold mGBSS buffer. Targeted brain areas were identified according to the Paxinos and Watson rat brain atlas,[39] surgically isolated, and stabilized as outlined below; samples from the habenula regions were pooled together. For LC–FT-ICR MS analysis, the medial and lateral habenula-containing regions from 10 rats were collected and stabilized in boiling water for 10 min. For LC–Orbitrap MS analysis, 6 rats were euthanized. The isolated medial and lateral habenula regions were stabilized with a Denator Stabilizor T1 (ST1) (Denator, Sweden) and pooled, respectively.[40] For LC–MALDI MS analysis, entire habenula regions from 3 rats were stabilized with the ST1 and pooled. For characterization of the sulfated peptides via LC–Orbitrap MS, habenula regions from 3 rats were isolated and stabilized by the ST1. The entire habenular nucleus (medial + lateral) was used in this case.

Extraction of Endogenous Peptides

The isolated habenula tissues were stabilized by the ST1 or hot water, followed by a three-stage peptide extraction process, as described in our previous work.[28] Briefly, habenula tissues were homogenized in 200 μL of ice cold water and incubated on ice for 1 h. Supernatants were saved in a new microcentrifuge tube after centrifugation at 14000 × *g* for 10 min. The tissue pellets were subjected to two more stages of peptide extraction with acidified acetone (acetone:H_2O:HCl 40:6:1) and 0.25% acetic acid solution, respectively, performed in a similar manner. The supernatants collected during these three extraction steps were combined.

The volume of combined supernatant sample was reduced to about 20 μL using a vacuum pre-concentrator device (GeneVac, UK) for HCl and acetone removal. Concentrated samples were desalted with a C18 spin column (Thermo Fisher Scientific). For LC–Orbitrap MS and LC–MALDI MS analysis, peptides were pre-purified on and eluted from the C18 spin column using an acetic acid organic solution buffer (pH 4.0). For LC–FT-ICR MS analysis, peptides retained on the C18 spin column were sequentially eluted into different vials using a set of buffers containing 70/30, 50/50, and 20/80 H_2O/acetonitrile (ACN) solutions, with the pH adjusted to 11.0. Each eluent fraction was collected, dried using a vacuum pre-concentrator, and stored in a –20 °C freezer until LC–MS analysis.

LC–Orbitrap MS Analysis

The peptides extracted from the rat habenular nuclei regions were analyzed using a nanoLC system (Dionex UltiMate 3000, Thermo Fisher Scientific) coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). The samples were reconstituted in loading solvent (95%/5% H_2O/ACN with 0.1% formic acid (FA)) and loaded onto a trap column. H_2O with 0.1% FA, and ACN with 0.1% FA, were used as solvents A and B, respectively. The loaded peptides were separated on a 15 cm Pepmap Acclaim analytical column (Thermo Fisher Scientific) packed with 2-μm particle sizes (100 Å pore size) at a 300 nL/min flow rate. The gradient for the peptide elution range was as follows: 3–6 min, 1–10% B; 6–90 min, 10–70% B. Data acquisition was accomplished in top speed data-dependent mode. Other parameters included: precursor scan automatic gain control (AGC),

2e+5; MS/MS scan AGC, 5e+4; isolation window, 1.6 Da; normalized collision energy, 35%; activation Q, 0.25.

LC–FT-ICR MS Analysis

The three dried eluent fractions were reconstituted in 10 μL of 95%/5% H_2O/ACN solution and analyzed individually by Eksigent nanoLC 1D Plus system (SCIEX, Framingham, MA, USA) coupled to a 11 Tesla FT mass spectrometer (LTQ-FT Ultra, Thermo Fisher Scientific) mass spectrometer, as described in a prior peptidomics work.[27] The experimental parameters were the same as described in the previous study,[27] including the columns, solvents, and flowrate. Briefly, the peptides were first loaded onto a peptide trap column and then separated on the analytical column. The sample was eluted over a gradient as follows: 0–10 min, 0–20% solvent B; 10–65 min, 20–55% solvent B; 65–75 min, 55–85% solvent B; 75–80 min, 85–85% solvent B; 80–85 min, 85–0% solvent B; 85–90 min, 0–0% B. The MS acquisition parameters included mass scans at *m/z* 300–2000. Data-dependent precursor selection was restricted to the top five most intense ions with the following parameters: precursor isolation window, *m/z* 10; activation Q, 0.25; activation time, 50 ms; collision energy 35 eV. Dynamic exclusion was enabled with a repeat count of 2 and exclusion duration of 180 s.

LC–Orbitrap MS for Characterization of Sulfation PTM

The peptide extracts were analyzed using the Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to the Dionex UltiMate 3000 UPLC system (Thermo Fisher Scientific) located at UW. A binary solvent system composed of H_2O containing 0.1% FA (A) and ACN containing 0.1% FA (B) was used for all analyses. Peptides were loaded and separated on a 75 μm × 15 cm self-fabricated column packed with 1.7 μm, 150 Å, BEH C18 material (part no. 186004661, Waters, Milford, MA). Samples were loaded with 3% solvent B, and solvent B was linearly ramped to 30% in 102 min, and ramped to 75% in another 20 min at a 300 nL/min flow rate. Data were acquired in data-dependent mode using the Orbitrap mass spectrometer with EThcD and high-energy C-trap dissociation. The precursor ions in the range of 300–1500 *m/z* were scanned with an AGC target set to 2e+5 and an Orbitrap resolution of 60,000. The fragment ion

analysis for the precursors was divided into five different scan events based on the charge state of the precursor ion. Scan event 1: intensity>1e+4; scan event 2: intensity>1e+4 and charge state 2; scan event 3: intensity>1e+4 and charge states 3–4; scan event 4: intensity>1e+4 and charge state 5; scan event 5: intensity>1e+5 and charge states 6–8. Scan event 1 analyzed all ions, whereas the other events only looked at specific charge states. The electron transfer dissociation reaction times were chosen appropriately for scan events 2–5, depending on the charge state. All of the scan events were performed with an Orbitrap resolution of 15,000 and AGC target value of 5e+4.

Peptide Identification: Untargeted Database Search

The raw files acquired on the LC–Orbitrap Fusion Lumos system were loaded into PEAKS software (Version 7.5, Canada). The loaded data were first processed with the *de novo* function in the software and then searched against a rat proteome database from UniProt[41] (36,078 entries). The mass error tolerance was set to 20 ppm for precursor ion and 0.1 Da for fragment ions. No enzyme digestion was selected and up to 3 variable PTMs were allowed for each peptide from a list of 7 modifications, including acetylation (K, N-term, and protein N-term), amidation, oxidation (M), pyro-Glu from E, pyro-Glu from Q, phosphorylation (STY), and half-disulfide bond (C). Identified peptides in the database search results were filtered by -10logP value cut-offs, corresponding to a 1% false discovery rate (FDR) or lower.

The raw data files acquired on the LC–FT-ICR MS system (LTQ-FT Ultra) were processed with the cRAWler algorithm in ProsightPC (Version 2.0, Thermo Fisher Scientific), which outputs .puf files containing the deconvoluted monoisotopic masses of precursor and fragment ions. The .puf files were then loaded into ProsightPC and searched against a lab-built rat proteome database based on information downloaded from UniProt. The database search was conducted in Biomarker search mode with the precursor and fragment mass error tolerances set at 81 Da and 10 ppm, respectively. In biomarker search mode, the detected precursor mass was compared with the calculated masses of all of the possible peptide subsequences present in the database. For subsequences with masses within 81 Da tolerance of a precursor mass, detected fragment masses were matched with

theoretical fragment masses from these candidate peptide subsequences. Identified peptides with p values lower than 1e-4 were retained.

Peptide Identification: Targeted Database Search

The .raw data files from the Orbitrap MS analysis were also searched against a lab-built rat neuropeptide prohormone library that contains prohormone sequences in FASTA format using PEAKS software. The same precursor and fragment ion mass error tolerance value were used as described above. Up to 7 variable PTMs, including acetylation (K, N-term, and protein N-term), amidation, oxidation (M), pyro-glu from E, pyro-glu from Q, phosphorylation (STY), and half-disulfide bond (C) were allowed for each peptide, so that the neuropeptides with multiple PTMs not detected in the initial round of the untargeted database search described above could be discovered and identified in the second round of targeted search. The FDR value was set at a stricter level of 0.1%.

Sulfated Peptide Identification

The. raw files from the HCD + EThcD method for the sulfated peptide analysis were performed with Byonic software (version 2.0, Protein Metrics, San Carlos, CA). Spectra were searched against a rat neuropeptide database (http://isyslab.info/NeuroPep/). A precursor tolerance of 10 ppm and a fragment mass tolerance of 20 ppm were allowed. Acetylation (N-term), oxidation (M), pyro-glu from E, and pyro-glu from Q were set as rare dynamic modifications. Common dynamic modifications consisted of phosphorylation (STY) and amidation (C-term). The mass spectra of the peptides identified to be sulfated were manually analyzed for the presence of a neutral loss of SO_3 that corresponds to a mass difference of 80 Da.[42,43]

LC–MALDI MS Analysis

Peptide extracts were injected into a Dionex capLC system equipped with a capillary C18 column (300 μm ID × 150 mm, 100 Å, Thermo Fisher Scientific). Solvents A and B were H_2O with 0.1% FA and ACN with 0.1% FA, respectively. Flow rate was set at 4 μL/min. Peptides were separated over 45 min with gradient solvent B rising from 1% to 50%. Fractions were collected every 3 min. Collected fractions were dried with a vacuum pre-

concentrator and reconstituted in 5 µL of H_2O. 1 µL of sample from each fraction was spotted on a MALDI sample plate, mixed with 50 mg/mL 2, 5-dihydroxybenzoic acid matrix, and dried at room temperature (~24°C). The dried samples were analyzed with a MALDI FT-ICR mass spectrometer (solariX XR, Bruker Daltonics, Billerica, MA) over a mass range of 200–4000 *m/z*. The putative identities of peaks in the MALDI mass spectra were assigned by matching the *m/z* of detected peaks to known neuropeptides contained in the SwePep rat database (http://www.swepep.org/), with a 0.1 Da allowable mass error.

RESULTS AND DISCUSSION

The habenular nuclei encompass two small regions symmetrically located in the brain hemispheres; these nuclei have important roles in a range of behavioral phenomena, including drug addiction and depression.[1,9,44] Neuropeptides are important cell-to-cell signaling molecules involved in these phenomena.[45-47] While immunohistochemical methods and *in situ* hybridization have been used to study the distribution of selected neuropeptides or their mRNAs within the habenular nuclei,[48-51] it appears surprising that an inventory of the neuropeptides found in the structure is not available. Here, we employed a combination of high resolution MS platforms to explore the neuropeptidome of the lateral and medial habenula in rat.

Evaluation of the morphology, connectivity to other brain regions, and biochemistry of the habenula nuclei suggests that the habenula is comprised of distinct lateral and medial subnuclei, each of which can be further divided into several smaller subnuclear regions. Difficulties in the observation, isolation, and LC–MS analysis of these smaller subnuclei restricted our investigation into the lateral and medial habenula nuclei.

Figure 1 shows the workflow of the multifaceted mass spectrometric approach used in this study. Acquired data from analysis 1 using the LC–Orbitrap system were searched against a rat proteome database for peptide identification. A targeted database search was also performed on the raw Orbitrap MS data using a prohormone library instead of a proteome database. For the targeted search, up to seven PTMs on individual peptides were allowed, facilitating the discovery of neuropeptides possessing multiple PTMs. For example, orexin-A has pyroglutamination at its N-terminus, two disulfide bonds, and amidation at its C-termini. In our analysis, six neuropeptides from the lateral habenula

region, which were missed in the initial untargeted search, were found by the targeted database search. Using LC–FT-ICR MS/MS and LC–Orbitrap MS/MS, 263 known and putative neuropeptides derived from 27 prohormones were identified from the medial habenula (Table S1); in the lateral habenula, 177 known and putative neuropeptides derived from 20 prohormones were found (Table S2). These small brain nuclei display a high molecular heterogeneity, which is in agreement with the complex spatial expression of genes and high heterogeneity of cell types.[52] Among the identified neuropeptides, 134 were detected and identified in both the medial and lateral habenula (Figure 2). Table 1 summarizes the prohormones detected in the medial and lateral habenula. Peptides from seven prohormones were exclusively detected in the medial habenula, including pro-opiomelanocortin, gastrin-releasing peptide, insulin-like growth factor-binding protein 5, neuropeptide S, neuroendocrine protein 7B2, corticoliberin, and tachykinin-3. These observations demonstrate that the medial and lateral habenula contain distinct sets of cell-cell signaling peptides. Mass spectra illustrating the results of the MS/MS characterization of selected physiologically important neuropeptides are shown in Figure S1.

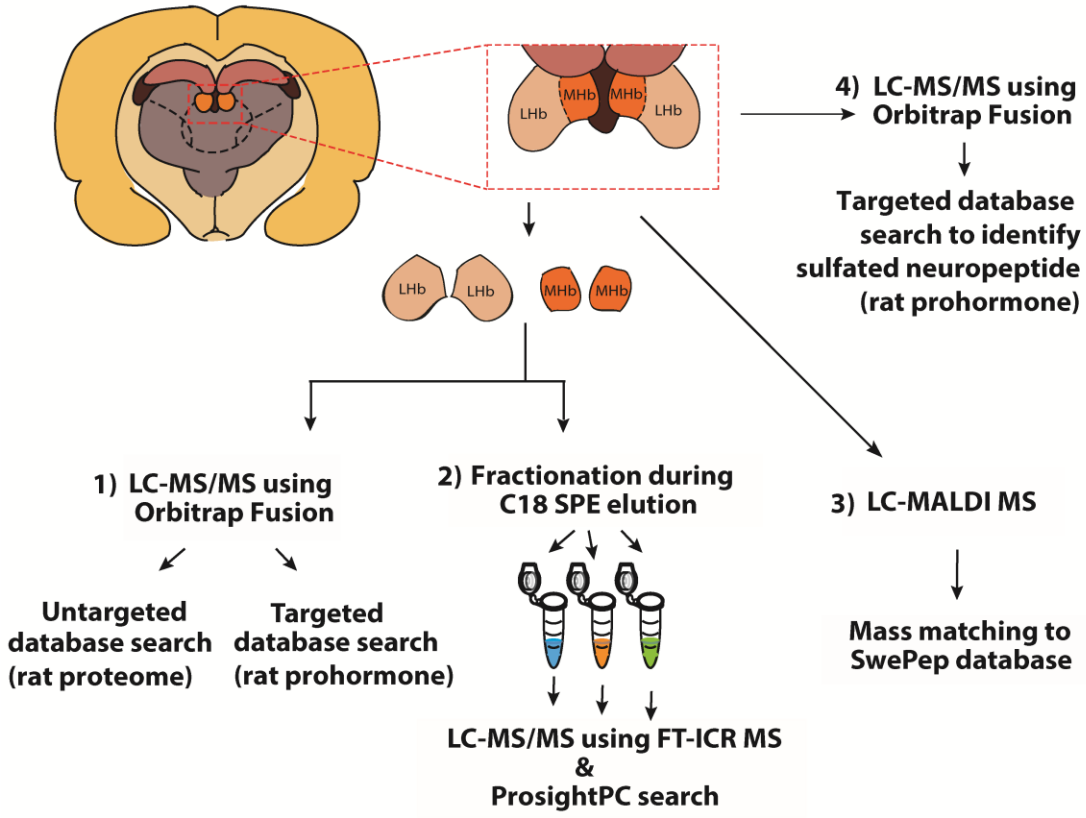


Figure 1. Workflow of the MS-based neuropeptidomic analyses of the rat habenular nuclei. Symmetrical habenular nuclei were isolated from the left and right brain hemispheres and characterized in three ways: (1) The medial and lateral habenula were analyzed with LC–ESI-Orbitrap MS and a PEAKS search against a rat proteome database (untargeted analysis), and a rat prohormone database (targeted analysis). (2) The medial and lateral habenula analyte extracts were first fractionated during C18 solid phase extraction, then analyzed with ESI-FT-ICR MS. (3) The entire habenula was fractionated with capLC and analyzed by MALDI-FT-ICR MS. (4) The entire habenula extracts were analyzed using specific settings for characterization of sulfated PTMs on the LC–ESI-Orbitrap mass spectrometer. The resultant raw spectra were searched against a rat prohormone database using Byonic 2.0.

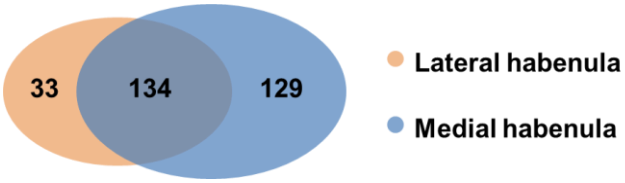


Figure 2. Venn diagram summarizing the results of peptidomic analyses of the medial and lateral habenula. A total of 263 and 177 neuropeptides were detected and identified in the medial and lateral habenula, respectively, with 134 neuropeptides detected in both regions.

Table 1. Prohormones expressed in the medial and lateral habenula and detected using MS-based peptidomic analyses.

Prohormone	Medial habenula	Lateral habenula
Cerebellin-1	X	X
Cholecystokinin	X	X
Chromogranin A	X	X
Cocaine- and amphetamine-regulated transcript protein	X	X
Corticoliberin	X	
Galanin peptides	X	X
Gastrin-releasing peptide	X	
Insulin-like growth factor-binding protein 5	X	
Neuroendocrine protein 7B2	X	
Neuropeptide S	X	
Neurosecretory protein VGF	X	X
Neurotensin	X	X
Orexin	X	X
Pituitary adenylate cyclase-activating polypeptide	X	X
Prepronociceptin	X	X
Proenkephalin-A	X	X
Proenkephalin-B	X	X
Pro-MCH	X	X
Pro-neuropeptide Y	X	X
Pro-opiomelanocortin	X	
ProSAAS	X	X
Protachykinin-1	X	X
Protachykinin-3	X	
Prothyroliberin	X	X
Secretogranin-1	X	X
Secretogranin-2	X	X
Secretogranin-3	X	X
Somatostatin	X	X

We also investigated the habenula using MALDI-FT-ICR MS. Most of the observed signals, even highly intense and well-resolved peaks, did not match to known neuropeptides compiled in the SwePep database. Part of the unassigned signals may correspond to neuropeptides possessing endogenous cleavage sites at both termini but not included in the database. Other unassigned signals could represent peptides originated from processing of non-prohormone proteins. Nevertheless, 24 signals detected by MALDI-FT-ICR MS were mass matched to known neuropeptides (Table S3) by mass matching detected peaks to known neuropeptides in the SwePep database, as described previously.[53] Of these, 22 peptides were also identified in the ESI MS data sets, improving the confidence of our assignments. Two signals were detected only with MALDI-FT-ICR MS,

and likely represent osteostatin and pro-FMRFamide-related neuropeptides. However, confirmation of the expression of these peptides needs further validation since no MS/MS spectra were acquired in the MALDI MS analysis. Nevertheless, the Allen Brain Atlas[54] *in situ* hybridization data demonstrate that mRNA of the parathyroid hormone-related protein containing the osteostatin sequence is present in the medial habenula region (Figure S2). Also, pro-FMRFamide-related neuropeptide FF receptors have been reported in the habenula region previously,[55] although the presence of parathyroid hormone-related protein mRNA and pro-FMRFamide-related neuropeptide FF receptors do not confirm their protein expression.

Although a comprehensive study of the neuropeptide complement of this physiologically important region has not been performed previously, a large body of information on the identity and function of several neuropeptides found in the habenular nuclei does exist. Our results align well with the results from previous studies, some of which are discussed below.

The prohormone protachykinin-1, which belongs to the tachykinin protein family, has been reported in the medial habenula region.[56] We identified multiple neuropeptides cleaved at mono/di-basic sites from tachykinin-1, including substance P, neurokinin A, short neuropeptide K, C-terminal flanking peptide, and protachykinin [111-118]. Prior immunocytochemistry and *in situ* hybridization investigations demonstrated the presence of cells expressing substance P and neurokinin B in the medial habenula, with their axons projecting to the interpeduncular nucleus and ventral tegmental area.[57] The neurokinin signaling pathway involving substance P, neurokinin B, and their receptors facilitates enhancement of nicotine-induced excitability in medial habenula neurons.[58] However, the biological functions of other protachykinin-1-derived peptides identified in the habenula are not clear. Further research on these neurokinin peptides can lead to a better understanding of the mechanism of nicotine addiction and perhaps neurokinin signaling-based therapies for smoking cessation and other disorders for which the actions of this neuropeptide system in the medial habenula may be relevant. Work done by Pertizziello et al.[59] on the effect of chronic nicotine treatment in the regulation of opioid and non-opioid peptides has revealed a significant difference in peptides levels derived from prohormones such as cerebellin-1, cholecystokinin, neurotensin, proenkephalin-A/B, proNPY, and

proSAAS in the dorsal striatum of model rats compared to the controls. In the current study, several endogenously cleaved full-length peptides from these prohormones were also found, further reinforcing the involvement of the habenular nuclei in nicotine addiction.

NPY [30-65] is a 36-amino acid cell-to-cell signaling molecule found in the nervous system that is involved in regulating different physiological and pathological events, such as food intake, stress, and depression.[60,61] NPY-immunoreactive fibers were detected in both lateral and medial habenula regions.[62] We detected and identified the full-length form of NPY and its C-terminal flanking peptide. A novel shortened NPY [30-47] peptide (YPSKPDNPGEDAPAEDMA.R) was also characterized with high confidence (34.0 for -10logP score). This shortened NPY has an arginine at its C-terminal, a common cleavage site in prohormone processing, supporting its endogenous nature, in contrast to the possibility that it was generated by degradation in post-mortem tissue. McBride and coworkers[63] have reported that mRNA transcripts encoding NPY are significantly reduced in the medial habenula of alcohol-preferring rats, suggesting a role for NPY in ethanol intake control.[63] This raises the interesting possibility that NPY, and novel modified forms of the peptide, expressed in the habenula may contribute to alcohol dependence vulnerability. While NPY is an appetite stimulator, cholecystokinin (CCK)-derived peptides have been associated with satiety and shown to decrease food intake.[64] Even in invertebrates, sulfakinins, the peptides with high sequence similarity to vertebrate CCK peptides, were shown to reduce food intake.[65] In this work, we detected a full-length mature peptide of the CCK prohormone (CCK[21-45]) with high confidence in both lateral and medial habenula regions, suggesting a role for the habenula in the regulation of food intake. Additionally, a recent study by Ye et al.,[66] has shown a direct correlation between food intake and brain levels of the proSAAS-derived peptides, big LEN (proSAAS[245-260]), and PEN (proSAAS[221-242]). Both of these peptides have been identified with high confidence scores in the current study.

The habenular nuclei are anatomically connected to the pineal gland, which together comprise the epithalamus.[8] The pineal gland is known to regulate sleep. Also, the epithalamus receives extrinsic circadian signals sent from the main circadian pacemaker, the suprachiasmatic nucleus (SCN).[67] The SCN is a well-investigated brain structure containing hundreds of peptides.[15,68] For example, SCN-originated orexin-A and -B

modulate sleep by regulating melatonin synthesis and secretion through the orexinergic neuron-pineal gland circuit in zebrafish and rat, as well as influence neuronal activities in the lateral habenula.[69,70] We detected and identified full-length orexin-B and other pro-orexin-derived peptides in both the lateral and medial habenula. Though the sequence identity of orexin-A was not confirmed, low intensity peaks of *m/z* corresponding to +4, +5, and +6 charged orexin-A molecules were detected (Figure S3). The relatively low levels of this molecule present in the studied structures, and/or biological matrix interferences, may have led to the low quality MS/MS spectra. In addition, the PTMs of orexin-A[71] (N-terminal pyroglutamyl cyclization, C-terminal amidation and two disulfide bonds between residues 38↔44 and 39↔46) complicate confident identification of the molecules with a targeted search. The identification of orexin-B in the habenular nuclei suggests that pro-orexin likely undergoes local translation and processing in the habenula, which may in turn contribute to sleep regulation by the epithalamus.

We also identified two new pituitary adenylate cyclase-activating polypeptide-derived neuropeptides, PACAP [111-128] and PACAP [130-143], in the habenular nuclei. Classic cleavage sites flank the N- and C-termini of both peptides. Immunohistochemistry and *in situ* hybridization studies have shown that these peptides and their receptors are present in the medial and lateral habenula.[72,73] Habenular nuclei neurons are activated by stress-inducing stimuli, resulting in a significant decrease in c-fos expression.[74] PACAP is associated with stress and anxiety-like behaviors.[8,44]

Another cell-to-cell signaling peptide observed in this study is somatostatin, which affects the stress response in both experience-dependent and independent ways in the zebrafish habenula.[75] Multiple pain-related peptides were also detected in this study, including those derived from prohormones such as nociceptin, pro-enkephalin-A, and pro-dynorphin-related, which are known to produce endogenous opiate peptides.[76-78] These peptides bind to various classes of opioid and nociception receptors, and are involved in mechanisms of pain signaling.[79] Our finding of pain-related peptides in the habenular nuclei is consistent with results from a preclinical study demonstrating an important role for the habenula in pain and analgesia.[80]

An additional experiment to identify the sites of sulfation revealed the location of five Tyr residues with O-sulfation. All of the five locations correspond to the prohormone

SCG1. One of them (SCG1 [339]) has been previously reported in rats,[81] whereas the other four are novel PTM sites of sulfation (SCG1 [330], [466], [469] and [501]). SCG1 [469] is predicted to be a sulfotyrosine by homology with a previously reported sulfation site in a bovine model.[82] While secretogranins are a class of secretory proteins that have been previously shown to be sulfated,[83] only one site of sulfation had been previously confirmed in the rat SCG1 prohormone.[81] Moreover, all of the identified peptides (Table S4) are either full-length forms or smaller fragments of two parent peptides (SCG1 [324-343] and SCG1 [454-508]), both with mono/di-basic amino acid cleavage sites. Sulfation is associated with the secretory pathway and the modification is known to occur in the *trans*-Golgi network. Modified Tyr residues have been shown to modulate ligand-receptor interaction in several previous studies.[35-37,84] The important role of sulfotyrosine in the regulation of ligand-receptor interactions, combined with the mono/di-basic cleavage sites of the corresponding peptides containing this modification, make them interesting candidates for further evaluation of biological activity. Secondly, although the sulfation experiments were performed on an MS platform with a different collision energy, ion detection, and search engine parameters in a separate laboratory, we did not observe significant differences in the total number of precursor proteins identified (data available on request). This consistency in results reflects the efficient sample stabilization and peptide extraction protocols used in this study.

Finally, we also detected and identified several novel neuropeptides, which have not been reported previously or have only corresponding mRNA transcripts detected in the habenular nuclei. These include peptides derived from CART, neurosecretory protein VGF, and several secretogranins.[85] However, their role in the habenula is unknown and further investigation will be required to understand their functions.

CONCLUSIONS

Habenular nuclei are morphologically, chemically, and functionally complex brain structures. Our peptidomic analyses utilized high-resolution MS-based and bioinformatics approaches to detect and identify 263 and 167 neuropeptides in the rat medial and lateral habenula regions, respectively. We determined the spatial specificity of the prohormone expressions and revealed the endogenous forms of neuropeptides in these habenular

structures. The habenular nuclei receive inputs from different brain regions and act as a bridge between the forebrain and midbrain. It is therefore not surprising that many normal and pathological conditions, as well as behaviors, depend on the activity of cells in this brain region. The current compendium of habenular-expressed peptides reveals the striking chemical complexity of this structure and points to some unique features of habenular signaling that involve neuropeptides. We expect these results to be useful in identifying endogenous ligands for the habenular nuclei orphan GPCRs[11,12] that may be targets for the development of novel therapeutics for neuropsychiatric disorders in which the habenula has been implicated.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Figure S1 showing MS/MS spectra of select neuropeptides detected in the habenula; Figure S2 showing parathyroid hormone-related protein mRNA expression; Tables S1–S4 listing neuropeptides detected by the approaches used in this study (PDF).

ACKNOWLEDGEMENTS

The project described was supported by Award No. P30 DA018310 from the National Institute on Drug Abuse, and Award No. R01 NS031609 from the National Institute of Neurological Disease and Stroke (J.V.S.); and Award No. R01 DK071801 from the National Institute of Digestive and Kidney Disease (L.L.). The Orbitrap Fusion Lumos Tribrid used for the sulfation experiments was purchased through the support of an NIH shared instrument grant, NIH-NCRR S10RR029531 (L.L.) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

REFERENCES

1. Wang, R.; Aghajanian, G., Physiological evidence for habenula as major link between forebrain and midbrain raphe. *Science* **1977,** 197, 89-91.

2. Kim, U.; Chang, S.-Y., Dendritic morphology, local circuitry, and intrinsic electrophysiology of neurons in the rat medial and lateral habenular nuclei of the epithalamus. *J. Comp. Neurol.* **2005,** 483, 236-250.

3. Christoph, G.; Leonzio, R.; Wilcox, K., Stimulation of the lateral habenula inhibits dopamine-containing neurons in the substantia nigra and ventral tegmental area of the rat. *J. Neurosci.* **1986,** 6, 613-619.

4. Schultz, W., Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **1998,** 80, 1-27.

5. Matsumoto, M.; Hikosaka, O., Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* **2007,** 447, 1111-1115.

6. Ji, H.; Shepard, P. D., Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABAA receptor-mediated mechanism. *J. Neurosci.* **2007,** 27, 6923-6930.

7. Medina, A. B.; Lechuga, D. A.; Escandón, O. S.; Moctezuma, J. V., Update of sleep alterations in depression. *Sleep Sci.* **2014,** 7, 165-169.

8. Hikosaka, O., The habenula: from stress evasion to value-based decision-making. *Nat. Rev. Neurosci.* **2010,** 11, 503-513.

9. Proulx, C. D.; Hikosaka, O.; Malinow, R., Reward processing by the lateral habenula in normal and depressive behaviors. *Nat. Neurosci.* **2014,** 17, 1146-1152.

10. Baldwin, P. R.; Alanis, R.; Salas, R., The role of the habenula in nicotine addiction. *J. Addict. Res. Ther.* **2011,** S1.

11. Broms, J.; Antolin-Fontes, B.; Tingström, A.; Ibañez-Tallon, I., Conserved expression of the GPR151 receptor in habenular axonal projections of vertebrates. *J. Comp. Neurol.* **2015,** 523, 359-380.

12. Ignatov, A.; Hermans-Borgmeyer, I.; Schaller, H. C., Cloning and characterization of a novel G-protein-coupled receptor with homology to galanin receptors. *Neuropharmacology* **2004,** 46, 1114-1120.

13. Hillebrand, J. J.; de Wied, D.; Adan, R. A., Neuropeptides, food intake and body weight regulation: a hypothalamic focus. *Peptides* **2002,** 23, 2283-2306.

14. Xu, P.; Hall, A. K., The role of activin in neuropeptide induction and pain sensation. *Dev. Biol.* **2006,** 299, 303-309.

15. Lee, J. E.; Atkins, N., Jr.; Hatcher, N. G.; Zamdborg, L.; Gillette, M. U.; Sweedler, J. V.; Kelleher, N. L., Endogenous peptide discovery of the rat circadian clock: a focused study of the suprachiasmatic nucleus by ultrahigh performance tandem mass spectrometry. *Mol. Cell. Proteomics* **2010,** 9, 285-297.

16. DiLeone, R. J.; Georgescu, D.; Nestler, E. J., Lateral hypothalamic neuropeptides in reward and drug addiction. *Life Sci.* **2003,** 73, 759-768.

17. Ong, T.-H.; Romanova, E. V.; Roberts-Galbraith, R.; Yang, N.; Zimmerman, T. A.; Collins, J. J.; Lee, J. E.; Kelleher, N. L.; Newmark, P. A.; Sweedler, J. V., Mass spectrometry imaging and identification of peptides associated with cephalic ganglia regeneration in Schmidtea mediterranea. *J. Biol. Chem.* **2016**, 8109-8120.

18. Fricker, L. D., Neuropeptide-processing enzymes: applications for drug discovery. *AAPS J.* **2005,** 7, E449-455.

19. Hoyer, D.; Bartfai, T., Neuropeptides and neuropeptide receptors: drug targets, and peptide and non-peptide ligands: a tribute to Prof. Dieter Seebach. *Chem. Biodivers.* **2012,** 9, 2367-2387.

20. Buchberger, A.; Yu, Q.; Li, L., Advances in mass spectrometric tools for probing neuropeptides. *Ann. Rev. Anal. Chem.* **2015,** 8, 485-509.

21. Romanova, E. V.; Sweedler, J. V., Peptidomics for the discovery and characterization of neuropeptides and hormones. *Trends Pharmacol. Sci.* **2015,** 36, 579-586.

22. Ong, T.-H.; Tillmaand, E. G.; Makurath, M.; Rubakhin, S. S.; Sweedler, J. V., Mass spectrometry-based characterization of endogenous peptides and metabolites in small volume samples. *Biochim. Biophys. Acta* **2015,** 1854, 732-740.

23. Sapio, M. R.; Fricker, L. D., Carboxypeptidases in disease: insights from peptidomic studies. *Proteomics Clin. Appl.* **2014,** 8, 327-337.

24. Hook, V.; Hwang, S.-R.; Wegrzyn, J.; Bark, S.; Begley, T. P., Neuropeptides: Chemical Activity Profiling and Proteomic Approaches. In *Wiley Encyclopedia of Chemical Biology*, John Wiley & Sons, Inc.: 2007; pp 1-13.

25. Romanova, E. V.; Lee, J. E.; Kelleher, N. L.; Sweedler, J. V.; Gulley, J. M., Comparative peptidomics analysis of neural adaptations in rats repeatedly exposed to amphetamine. *J. Neurochem.* **2012,** 123, 276-287.

26. Salisbury, J. P.; Boggio, K. J.; Hsu, Y. W.; Quijada, J.; Sivachenko, A.; Gloeckner, G.; Kowalski, P. J.; Easterling, M. L.; Rosbash, M.; Agar, J. N., A rapid MALDI-TOF mass spectrometry workflow for Drosophila melanogaster differential neuropeptidomics. *Mol. Brain* **2013,** 6, 60.

27. Tillmaand, E. G.; Yang, N.; Kindt, C. A.; Romanova, E. V.; Rubakhin, S. S.; Sweedler, J. V., Peptidomics and secretomics of the mammalian peripheral sensory-motor system. *J. Am. Soc. Mass Spectrom.* **2015,** 26, 2051-2061.

28. Bora, A.; Annangudi, S. P.; Millet, L. J.; Rubakhin, S. S.; Forbes, A. J.; Kelleher, N. L.; Gillette, M. U.; Sweedler, J. V., Neuropeptidomics of the supraoptic rat nucleus. *J. Proteome Res.* **2008,** 7, 4992-5003.

29. Marcus, J. N.; Aschkenasi, C. J.; Lee, C. E.; Chemelli, R. M.; Saper, C. B.; Yanagisawa, M.; Elmquist, J. K., Differential expression of orexin receptors 1 and 2 in the rat brain. *J. Comp. Neurol.* **2001,** 435, 6-25.

30. Lynch, D. R.; Walker, M. W.; Miller, R. J.; Snyder, S. H., Neuropeptide Y receptor binding sites in rat brain: differential autoradiographic localizations with 125I-peptide YY and 125I-neuropeptide Y imply receptor heterogeneity. *J. Neurosci.* **1989,** 9, 2607-2619.

31. Kopp, J.; Xu, Z. Q.; Zhang, X.; Pedrazzini, T.; Herzog, H.; Kresse, A.; Wong, H.; Walsh, J. H.; Hökfelt, T., Expression of the neuropeptide Y Y1 receptor in the CNS of rat and of wild-type and Y1 receptor knock-out mice. Focus on immunohistochemical localization. *Neuroscience* **2002,** 111, 443-532.

32. Leyte, A.; van Schijndel, H. B.; Niehrs, C.; Huttner, W. B.; Verbeet, M. P.; Mertens, K.; van Mourik, J. A., Sulfation of Tyr1680 of human blood coagulation factor VIII is essential for the interaction of factor VIII with von Willebrand factor. *J. Biol. Chem.* **1991,** 266, 740-746.

33. Wilkins, P. P.; Moore, K. L.; McEver, R. P.; Cummings, R. D., Tyrosine sulfation of P-selectin glycoprotein ligand-1 is required for high affinity binding to P-selectin. *J. Biol. Chem.* **1995,** 270, 22677-22680.

34. Hortin, G. L.; Farries, T. C.; Graham, J. P.; Atkinson, J. P., Sulfation of tyrosine residues increases activity of the fourth component of complement. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** 86, 1338-1342.

35. Gao, J.-m.; Xiang, R.-l.; Jiang, L.; Li, W.-h.; Feng, Q.-p.; Guo, Z.-j.; Sun, Q.; Zeng, Z.-p.; Fang, F.-d., Sulfated tyrosines 27 and 29 in the N-terminus of human CXCR3 participate in binding native IP-10. *Acta Pharmacol. Sin.* **2009,** 30, 193-201.

36. Costagliola, S.; Panneels, V.; Bonomi, M.; Koch, J.; Many, M. C.; Smits, G.; Vassart, G., Tyrosine sulfation is required for agonist recognition by glycoprotein hormone receptors. *EMBO J.* **2002,** 21, 504-513.

37. Farzan, M.; Schnitzler, C. E.; Vasilieva, N.; Leung, D.; Kuhn, J.; Gerard, C.; Gerard, N. P.; Choe, H., Sulfated tyrosines contribute to the formation of the C5a docking site of the human C5a anaphylatoxin receptor. *J. Exp. Med.* **2001,** 193, 1059-1066.

38. Bettelheim, F. R., Tyrosine O-sulphate in a peptide from fibrinogen. *J. Am. Chem. Soc.* **1954,** 76, 2838-2839.

39. Paxinos, G.; Watson, C., *The Rat Brain in Stereotaxic Coordinates*. 7th ed.; Academic Press: Cambridge, MA, 2013; p 472.

40. Yang, N.; Anapindi, K. D. B.; Romanova, E. V.; Rubakhin, S. S.; Sweedler, J., Improved identification and quantitation of mature endogenous peptides in the rodent hypothalamus using a rapid conductive sample heating system. *Analyst* **2017**, in press. 10.1039/C1037AN01358B.

41. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2017,** 45, D158-D169.

42. Wang, Y.; Wang, M.; Yin, S.; Jang, R.; Wang, J.; Xue, Z.; Xu, T., NeuroPep: a comprehensive resource of neuropeptides. *Database (Oxford)* **2015,** 2015, bav038.

43. Yu, Q.; Canales, A.; Glover, M. S.; Das, R.; Shi, X.; Liu, Y.; Keller, M. P.; Attie, A. D.; Li, L., Targeted mass spectrometry approach enabled discovery of O-glycosylated insulin and related signaling peptides in mouse and human pancreatic islets. **2017,** 89, 9184-9191.

44. Shumake, J.; Gonzalez-Lima, F., Brain systems underlying susceptibility to helplessness and depression. *Behav. Cogn. Neurosci. Rev.* **2003,** 2, 198-221.

45. Luo, D. D.; An, S. C.; Zhang, X., Involvement of hippocampal serotonin and neuropeptide Y in depression induced by chronic unpredicted mild stress. *Brain Res. Bull.* **2008,** 77, 8-12.

46. Wang, Y.; Yang, Y.; Hui, L.; Tie, C.; Li, F.; Xu, Z.-Q. D.; Wang, C., A neuropeptide Y variant (rs16139) associated with major depressive disorder in replicate samples from Chinese Han population. *PLoS One* **2013,** 8, e57042.

47. Merrer, J. L.; Becker, J. A. J.; Befort, K.; Kieffer, B. L., Reward processing by the opioid system in the brain. *Physiol. Rev.* **2009,** 89, 1379-1412.

48. Southey, B. R.; Lee, J. E.; Zamdborg, L.; Atkins, N.; Mitchell, J. W.; Li, M.; Gillette, M. U.; Kelleher, N. L.; Sweedler, J. V., Comparing label-free quantitative peptidomics approaches to characterize diurnal variation of peptides in the rat suprachiasmatic nucleus. *Anal. Chem.* **2014,** 86, 443-452.

49. Reuss, S.; Hurlbut, E. C.; Speh, J. C.; Moore, R. Y., Neuropeptide Y localization in telencephalic and diencephalic structures of the ground squirrel brain. *Am. J. Anat.* **1990,** 188, 163-174.

50. Pompei, P.; Severini, R.; Costa, G.; Massi, M.; Fattoretti, P.; Bertoni-Freddari, C., In situ hybridization analysis of preprotachykinin-A mRNA levels in young and old rats. *Mol. Brain Res.* **1999,** 64, 132-136.

51. Moore, J. P., Jr.; Burger, L. L.; Dalkin, A. C.; Winters, S. J., Pituitary adenylate cyclase activating polypeptide messenger RNA in the paraventricular nucleus and anterior pituitary during the rat estrous cycle. *Biol. Reprod.* **2005,** 73, 491-499.

52. Wagner, F.; French, L.; Veh, R., Transcriptomic-anatomic analysis of the mouse habenula uncovers a high molecular heterogeneity among neurons in the lateral complex, while gene expression in the medial complex largely obeys subnuclear boundaries. *Brain Struct. Funct.* **2014**, 1-20.

53. Rubakhin, S. S.; Sweedler, J. V., Characterizing peptides in individual mammalian cells using mass spectrometry. *Nat. Protocols* **2007,** 2, 1987-1997.

54. Lein, E. S.; Hawrylycz, M. J.; Ao, N.; Ayres, M.; Bensinger, A.; Bernard, A.; Boe, A. F.; Boguski, M. S.; Brockway, K. S.; Byrnes, E. J.; Chen, L.; Chen, L.; Chen, T. M.; Chin, M. C.; Chong, J.; Crook, B. E.; Czaplinska, A.; Dang, C. N.; Datta, S.; Dee, N. R.; Desaki, A. L.; Desta, T.; Diep, E.; Dolbeare, T. A.; Donelan, M. J.; Dong, H. W.; Dougherty, J. G.; Duncan, B. J.; Ebbert, A. J.; Eichele, G.; Estin, L. K.; Faber, C.; Facer, B. A.; Fields, R.; Fischer, S. R.; Fliss, T. P.; Frensley, C.; Gates, S. N.; Glattfelder, K. J.; Halverson, K. R.; Hart, M. R.; Hohmann, J. G.; Howell, M. P.; Jeung, D. P.; Johnson, R. A.; Karr, P. T.; Kawal, R.; Kidney, J. M.; Knapik, R. H.; Kuan, C. L.; Lake, J. H.; Laramee, A. R.; Larsen, K. D.; Lau, C.; Lemon, T. A.; Liang, A. J.; Liu, Y.; Luong, L. T.; Michaels, J.; Morgan, J. J.; Morgan, R. J.; Mortrud, M. T.; Mosqueda, N. F.; Ng, L. L.; Ng, R.; Orta, G. J.; Overly, C. C.; Pak, T. H.; Parry, S. E.; Pathak, S. D.; Pearson, O. C.; Puchalski, R. B.; Riley, Z. L.; Rockett, H. R.; Rowland, S. A.; Royall, J. J.; Ruiz, M. J.; Sarno, N. R.; Schaffnit, K.; Shapovalova, N. V.; Sivisay, T.; Slaughterbeck, C. R.; Smith, S. C.; Smith, K. A.; Smith, B. I.; Sodt, A. J.; Stewart, N. N.; Stumpf, K. R.; Sunkin, S. M.; Sutram, M.; Tam, A.; Teemer, C. D.; Thaller, C.; Thompson, C. L.; Varnam, L. R.; Visel, A.; Whitlock, R. M.; Wohnoutka, P. E.; Wolkey, C. K.; Wong, V. Y.; Wood, M.; Yaylaoglu, M. B.; Young, R. C.; Youngstrom, B. L.; Yuan, X. F.; Zhang, B.; Zwingman, T. A.; Jones, A. R., Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **2007,** 445, 168-176.

55. Liu, Q.; Guan, X. M.; Martin, W. J.; McDonald, T. P.; Clements, M. K.; Jiang, Q.; Zeng, Z.; Jacobson, M.; Williams, D. L., Jr.; Yu, H.; Bomford, D.; Figueroa, D.; Mallee, J.; Wang, R.; Evans, J.; Gould, R.; Austin, C. P., Identification and characterization of novel mammalian neuropeptide FF-like peptides that attenuate morphine-induced antinociception. *J. Biol. Chem.* **2001,** 276, 36961-36969.

56. Atkins, N., Jr.; Mitchell, J. W.; Romanova, E. V.; Morgan, D. J.; Cominski, T. P.; Ecker, J. L.; Pintar, J. E.; Sweedler, J. V.; Gillette, M. U., Circadian integration of glutamatergic signals by little SAAS in novel suprachiasmatic circuits. *PLoS One* **2010,** 5, e12612.

57. Sandweiss, A. J.; Vanderah, T. W., The pharmacology of neurokinin receptors in addiction: prospects for therapy. *Subst. Abuse Rehabil.* **2015,** 6, 93-102.

58. Dao, D. Q.; Perez, E. E.; Teng, Y.; Dani, J. A.; De Biasi, M., Nicotine enhances excitability of medial habenular neurons via facilitation of neurokinin signaling. *J. Neurosci.* **2014,** 34, 4273-4284.

59. Petruzziello, F.; Falasca, S.; Andren, P. E.; Rainer, G.; Zhang, X., Chronic nicotine treatment impacts the regulation of opioid and non-opioid peptides in the rat dorsal striatum. *Mol. Cell. Proteomics* **2013,** 12, 1553-1562.

60. Schwartz, M. W.; Woods, S. C.; Porte, D.; Seeley, R. J.; Baskin, D. G., Central nervous system control of food intake. *Nature* **2000,** 404, 661-671.

61. Morales-Medina, J. C.; Dumont, Y.; Quirion, R., A possible role of neuropeptide Y in depression and stress. *Brain Res.* **2010,** 1314, 194-205.

62. Secher, A.; Kelstrup, C. D.; Conde-Frieboes, K. W.; Pyke, C.; Raun, K.; Wulff, B. S.; Olsen, J. V., Analytic framework for peptidomics applied to large-scale neuropeptide identification. *Nat. Comm.* **2016,** 7, 11436.

63. Hwang, B. H.; Suzuki, R.; Lumeng, L.; Li, T. K.; McBride, W. J., Innate differences in neuropeptide Y (NPY) mRNA expression in discrete brain regions between alcohol-preferring (P) and -nonpreferring (NP) rats: a significantly low level of NPY mRNA in dentate gyrus of the hippocampus and absence of NPY mRNA in the medial habenular nucleus of P rats. *Neuropeptides* **2004,** 38, 359-368.

64. Gibbs, J.; Young, R. C.; Smith, G. P., Cholecystokinin decreases food intake in rats. *J. Comp. Physiol. Psychol.* **1973,** 84, 488-495.

65. Wei, Z.; Baggerman, G.; J. Nachman, R.; Goldsworthy, G.; Verhaert, P.; De Loof, A.; Schoofs, L., Sulfakinins reduce food intake in the desert locust, Schistocerca gregaria. *J. Insect Physiol.* **2000,** 46, 1259-1265.

66. Ye, H.; Wang, J.; Tian, Z.; Ma, F.; Dowell, J. A.; Bremer, Q.; Lu, G.; Baldo, B.; Li, L., Quantitative mass spectrometry reveals food intake-induced neuropeptide level changes in rat brain: Functional assessment of selected neuropeptides as feeding regulators. *Mol. Cell. Proteomics* **2017,** 16, 1922-1937.

67. Sakhi, K.; Wegner, S.; Belle, M. D. C.; Howarth, M.; Delagrange, P.; Brown, T. M.; Piggins, H. D., Intrinsic and extrinsic cues regulate the daily profile of mouse lateral habenula neuronal activity. *J. Physiol.* **2014,** 592, 5025-5045.

68. Lee, J. E.; Zamdborg, L.; Southey, B. R.; Atkins, N.; Mitchell, J. W.; Li, M.; Gillette, M. U.; Kelleher, N. L.; Sweedler, J. V., Quantitative peptidomics for discovery of circadian-related peptides from the rat suprachiasmatic nucleus. *J. Proteome Res.* **2013,** 12, 585-593.

69. Appelbaum, L.; Wang, G. X.; Maro, G. S.; Mori, R.; Tovin, A.; Marin, W.; Yokogawa, T.; Kawakami, K.; Smith, S. J.; Gothilf, Y.; Mignot, E.; Mourrain, P., Sleep–

wake regulation and hypocretin–melatonin interaction in zebrafish. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, 21942-21947.

70. Mikkelsen, J. D.; Hauser, F.; DeLecea, L.; Sutcliffe, J. G.; Kilduff, T. S.; Calgari, C.; Pévet, P.; Simonneaux, V., Hypocretin (orexin) in the rat pineal gland: a central transmitter with effects on noradrenaline-induced release of melatonin. *Eur. J. Neurosci.* **2001,** 14, 419-425.

71. Sakurai, T.; Amemiya, A.; Ishii, M.; Matsuzaki, I.; Chemelli, R. M.; Tanaka, H.; Williams, S. C.; Richardson, J. A.; Kozlowski, G. P.; Wilson, S.; Arch, J. R.; Buckingham, R. E.; Haynes, A. C.; Carr, S. A.; Annan, R. S.; McNulty, D. E.; Liu, W. S.; Terrett, J. A.; Elshourbagy, N. A.; Bergsma, D. J.; Yanagisawa, M., Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **1998,** 92, 573-585.

72. Hannibal, J., Pituitary adenylate cyclase-activating peptide in the rat central nervous system: An immunohistochemical and in situ hybridization study. *J. Comp. Neurol.* **2002,** 453, 389-417.

73. Jolivel, V.; Basille, M.; Aubert, N.; de Jouffrey, S.; Ancian, P.; Le Bigot, J. F.; Noack, P.; Massonneau, M.; Fournier, A.; Vaudry, H.; Gonzalez, B. J.; Vaudry, D., Distribution and functional characterization of pituitary adenylate cyclase–activating polypeptide receptors in the brain of non-human primates. *Neuroscience* **2009,** 160, 434-451.

74. Meloni, E. G.; Venkataraman, A.; Donahue, R. J.; Carlezon, W. A., Jr., Bi-directional effects of pituitary adenylate cyclase-activating polypeptide (PACAP) on fear-related behavior and c-Fos expression after fear conditioning in rats. *Psychoneuroendocrinology* **2015,** 64, 12-21.

75. Mathuru, A. S.; Jesuthasan, S., The medial habenula as a regulator of anxiety in adult zebrafish. *Front. Neural Circuits* **2013,** 7, 99.

76. Holden, J. E.; Jeong, Y.; Forrest, J. M., The endogenous opioid system and clinical pain management. *AACN Clin. Issues* **2005,** 16, 291-301.

77. Zeilhofer, H. U.; Calò, G., Nociceptin/orphanin FQ and its receptor--potential targets for pain therapy? *J. Pharmacol. Exp. Ther.* **2003,** 306, 423-429.

78. Carlton, S. M.; Du, J.; Zhou, S.; Coggeshall, R. E., Tonic control of peripheral cutaneous nociceptors by somatostatin receptors. *J Neurosci* **2001,** 21, 4042-4049.

79. Beaumont, A.; Hughes, J., Biology of opioid peptides. *Annu. Rev. Pharmacol. Toxicol.* **1979,** 19, 245-267.

80. Shelton, L.; Becerra, L.; Borsook, D., Unmasking the mysteries of the habenula in pain and analgesia. *Prog. Neurobiol.* **2012,** 96, 208-219.

81. Benedum, U. M.; Lamouroux, A.; Konecki, D. S.; Rosa, P.; Hille, A.; Baeuerle, P. A.; Frank, R.; Lottspeich, F.; Mallet, J.; Huttner, W. B., The primary structure of human secretogranin I (chromogranin B): comparison with chromogranin A reveals homologous terminal domains and a large intervening variable region. *EMBO J.* **1987,** 6, 1203-1211.

82. Gasnier, C.; Lugardon, K.; Ruh, O.; Strub, J.-M.; Aunis, D.; Metz-Boutigue, M.-H., Characterization and location of post-translational modifications on chromogranin B from bovine adrenal medullary chromaffin granules. *Proteomics* **2004,** 4, 1789-1801.

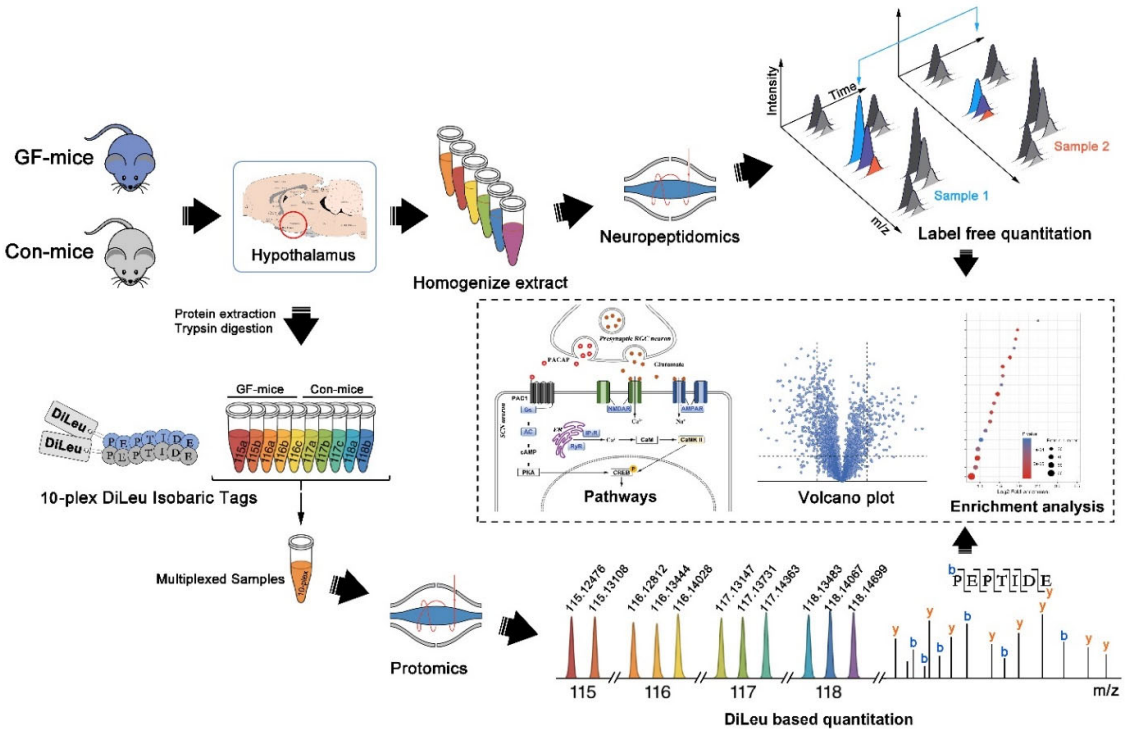
83. Rosa, P.; Hille, A.; Lee, R. W.; Zanini, A.; De Camilli, P.; Huttner, W. B., Secretogranins I and II: two tyrosine-sulfated secretory proteins common to a variety of cells secreting peptides by the regulated pathway. *J. Cell Biol.* **1985,** 101, 1999-2011.

84. Klement, É.; Hunyadi-Gulyás, É.; Medzihradszky, K. F., Biological significance and analysis of tyrosine sulfation. In *Analysis of Protein Post-Translational Modifications by Mass Spectrometry*, John Wiley & Sons, Inc.: 2016; pp 333-349.

85. Ahnoff, M.; Cazares, L. H.; Skold, K., Thermal inactivation of enzymes and pathogens in biosamples for MS analysis. *Bioanalysis* **2015,** 7, 1885-1899.

Appendix V

Integrated Label-Free and Ten-Plex DiLeu Tag Quantitative Methods for Profiling Changes in the Mouse Hypothalamic Neuropeptidome and Proteome under Different Gut Microbiota Environments



Adapted from Rui Liu, **Pingli Wei,** Caitlin Keller, Nicola Salvatore Orefice, Yatao Shi, Zihui Li, Junfeng Huang, Yusi Cui, Dustin C. Frost, Shuying Han, Tzu-Wen L. Cross, Federico E. Rey, Lingjun Li. "Integrated Label-Free and Ten-Plex DiLeu Tag Quantitative Methods for Profiling Changes in the Mouse Hypothalamic Neuropeptidome and Proteome under Different Gut Microbiota Environments." *To be submitted*. Pingli Wei provided expertise on the neuropeptidomics.

Abstract

Gut microbiota can regulate host physiological and pathological status through gut-brain communications or pathways. However, although they are important signaling molecules, the neuropeptides and proteins involved in regulating brain functions or behaviors under different gut microbiota environments are still not clearly known. To address the problem, integrated label-free and ten-plex DiLeu isobaric tag-based quantitative methods were implemented to compare the profiling of neuropeptides and proteins in the germ-free (GF) and conventional (Con) mice hypothalamus. A total of 2,943 endogenous peptides from 63 precursorsneuropeptide preprohormones? and 3,971 proteins in the mouse hypothalamus were identified. Among these 368 significantly changed peptides (fold changes over 1.5, and p-value < 0.05), 73.6% of the peptides showed higher levels in GF-mice than in Con-mice, and 26.4% of the peptides had higher levels in Con-mice than in GF-mice. These peptides were mainly from Secretogranin-2, Phosphatidylethanolamine-binding protein-1, ProSAAS, and Proenkephalin-A. DiLeu-labeled quantification showed that 282 proteins were significantly up- or down-regulated (fold changes over 1.5, and p-value < 0.05) among the 3,277 quantified proteins. These neuropeptides and proteins were mainly involved in behaviors, transmitter release, signaling pathways, and synapses. Interestingly, pathways including long-term potentiation, long-term depression, and circadian entrainment were involved. In the present study, a label-free and DiLeu-based quantitative method reveals comprehensively and significantly changed neuropeptides and proteins in the hypothalamus under different gut microbiota environments, suggesting that these neuropeptides and proteins might exert key roles in mouse behavior, brain development, memory, and learning.

Keywords: Gut microbiota; Hypothalamus; Neuropeptidomics; Proteomics; Quantitation

Introduction

Neuropeptides are a class of endogenous peptides that act as long-lasting cell-to-cell signaling neurotransmitters in the nervous system and other target organs. Neuropeptides and their receptors play important roles in several key processes[1], such as modulating neurobehavior[2], social development[3], affecting energy harvest from the diet and energy storage[4], inducing systemic insulin resistance, and modifying glucose homeostasis and the immune response[5,6]. Neuropeptides and their precursors are important mediators both within the nervous system and between neurons and other cell types. When a neuron releases neuropeptides, the binding of the neuropeptide to its receptor on a receiving cell causes conformational changes within the receptor that, depending on the type of receptor, either open ion channels or activate coupled G-proteins that can cause a series of downstream effects within the cell[7].

Furthermore, neuropeptides have been implicated in the regulation of normal biological functions such as circadian rhythm[8], feeding regulation[9], temperature fluctuation[10], depression, stress, anxiety[11], cognition[12], and other behaviors[13]. For instance, pituitary adenylate cyclase-activating polypeptide (PACAP) has neuroprotective and neurotrophic properties that could slow down Alzheimer's disease-like pathology; increasing or restoring PACAP and its receptor function might provide Alzheimer's disease treatment benefits[14]. A ProSAAS-derived neuropeptide, Big LEN, could regulate body weight by binding G-protein-coupled receptor 171 (GPR171), which might be associated with food intake and metabolism[15]. Studies have also shown that the Big LEN-GPR171 system could regulate anxiety-

like behavior and fear conditioning[16]. In addition, proteins in the brain, such as receptors and kinases, are also important and influence CNS development and behaviors. Compared to normal mice, Neufeld reported that GF-mice showed a decrease in the N-methyl-D-aspartate receptor subunit NR2B and in serotonin receptor 1A, as well as an increase in a brain-derived neurotrophic factor[17]. It has been well established that gut microbiota can alter the expression of cAMP-responding element-binding protein (CREB) and protein kinase C (PKC), which might facilitate neurodevelopment via the PKC-CREB signaling pathway[18].

Over the past decade, various evidence has supported the assertion that gut microbiota can affect the host's physiological status through "brain-gut axis" pathways[19-21]. It has been demonstrated that central nervous system (CNS) neurotransmission can be profoundly affected by the absence of normal gut microbiota[22]. Gut microbiota can regulate mouse behavior through the hippocampal glucocorticoid receptor pathway[23]. Investigations have indicated the effects of gut microbiota and how they exert their functions in regulating host physiological and pathological status through gut-brain communications or pathways. Microbiota leading to CNS influences might be associated with second messenger pathways and synaptic long-term potentiation (LTP) in brain regions[23,24]. Bercik reported that gut microbiota influence behavior and brain chemistry through the autonomic nervous system, gastrointestinal-specific neurotransmitters, or inflammation[25].

The gut-brain axis uses four major information carriers for communication between the gut and the brain: vagal and spinal afferent neurons, immune mediators, gut hormones, and gut microbiota-derived signaling molecules[26]. These brain chemistries, including neurotransmitters, neuropeptides, hormones, proteins, and receptors, can be totally different under different gut microbiota environments[26-28]. The brain is arguably the most fascinating organ in the body and also the most

complex, as reflected in its hundred billion neurons and their several hundred trillions of interconnections[29]. Among these gut-influenced brain chemistries, neuropeptides and proteins are the most important in behavior, neural interconnection, neurodegeneration, neuron development, and neuroimmunomodulation[30]. However, to date, there has been little research on determining the effects of gut microbiota on brain proteomics and neuropeptidomics, or on possible pathways of gut-brain communications at a protein or neuropeptide level.

Quantitative MS-based proteomics or peptidomics can be used to characterize relative peptide abundance across different groups and can normally be implemented using label-based or label-free methods[1,31,32]. In this study, neuropeptides and proteins were simultaneously extracted from the same hypothalamus region, and then the relative quantitative changes of neuropeptides and proteins in different groups were identified and compared based on precursor ion intensity from MS, MS/MS spectra, or the report ions of isobaric tags. Therefore, the integrated label-free and ten-plex DiLeu-tag quantitative method showed the profiling changes of neuropeptides and proteins in the mouse hypothalamus under different gut microbiota environments. This approach provided further evidence on how gut microbiota regulate brain signaling molecules (peptides and proteins), which might reveal the possible communications or pathways that gut microbiota could bring about through these crucial neuropeptides or proteins.

Experimental

Animal Experiment and Tissue Extraction

Animal care and study protocols were approved by the UW-Madison Animal Care and Use Committee. Mice were housed in the Microbial Sciences Building vivarium. Germ-free mice (GF-mice) and conventional mice (Con-mice) were bred at UW-Madison. GF-mice were housed in separate plastic flexible vinyl gnotobiotic isolators. Mice were group-housed by colonization status and diet (five mice/cage) under standard conditions (12 h light: dark, temperature- and humidity-controlled conditions) and received *ad-libitum* access to water and food. Five GF-mice and five Con-mice were all sacrificed, and their brains were dissected and immediately rapid-heated via Denator™ to minimize postmortem degradation, after which the hypothalamus region was isolated. A detailed description of tissue extraction, protein digestion, and instrument operation methods can be found in Supporting Information and a previous report[31].

Results and Discussion

LFQ of endogenous peptides in the hypothalamus region in GF- and Con-mice

An integrated LFQ (label-free quantification) and ten-plex DiLeu-based strategy was developed; the workflow of this study is illustrated in Figure 1. In total, 2,943 endogenous peptides derived from 63 precursors from the hypothalamus region of both GF- and Con-mice were identified. A complete summary of all the identified peptides is provided in Table S1. High-confidence peptide identification was achieved using the high-resolution, high-accuracy nanoLC Orbitrap MS/MS (Figure S1).

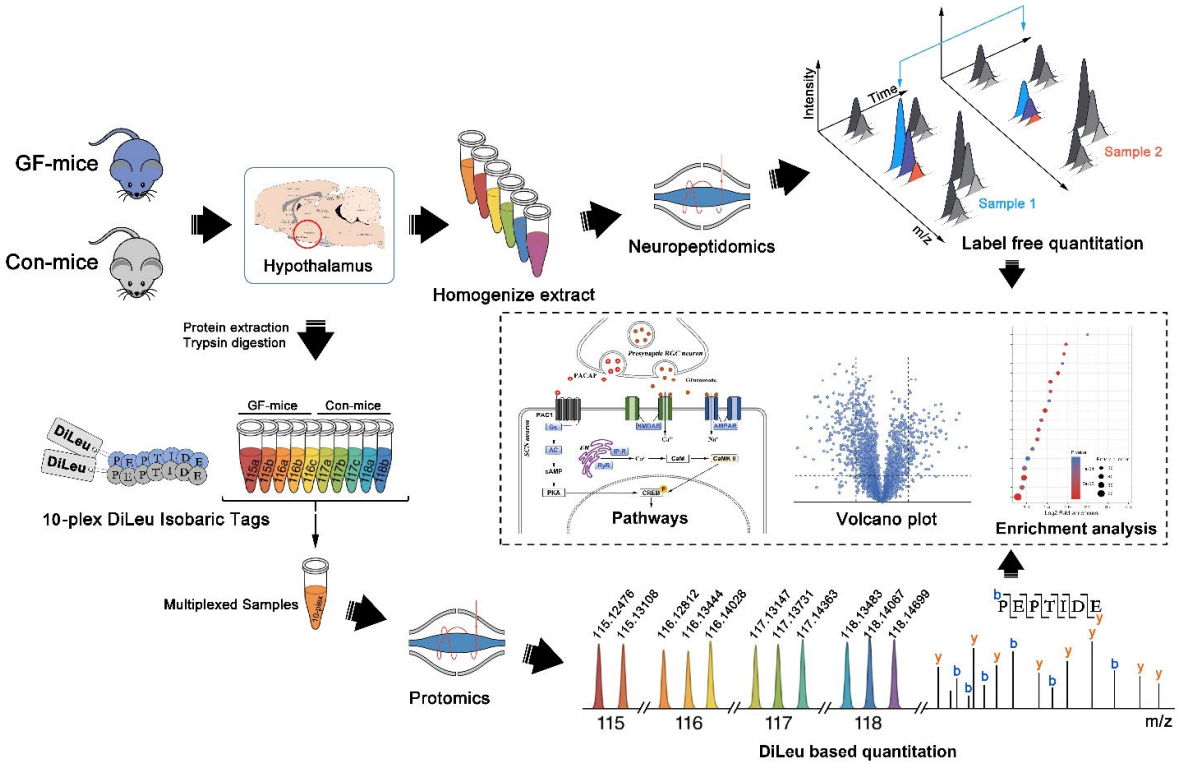


Figure 1. Workflow for the integrated LFQ and the ten-plex DiLeu-based strategy.

Among the 2,943 identified endogenous peptides, as shown in Figure 2A, 880 peptides were detected only in the GF-mice part and were mainly derived from well-characterized neuropeptide precursors, including SCG2, pro-opiomelanocortin (POMC), ProSAAS (PCSK1N), phosphatidylethanolamine-binding protein 1 (PEBP1), neuroendocrine protein (SCG5), vasopressin-neurophysin 2-copeptin (AVP), pro-thyrotropin-releasing hormone (TRH), secretogranin-1 (SCG1), and somatostatin (SST) (Figure 2D). A total of 387 peptides were detected only in the Con-mice part and included mainly SCG2, PEBP1, SCG1, proenkephalin-A (PENK), SST, PCSK1N, cholecystokinin (CCK), and VIP peptides (VIP) (Figure 2C). Over half the total, or 1,676 peptides, were identified in the overlapping set of GF- and Con-mice, in which the SCG2 precursor accounted

for the highest number of peptides (535), followed by PEBP1, PCSK1N, SCG1, PENK, SCG5, and SST(Figure 2E). As shown in Figure 2B, the average length of the endogenous peptides in the Con-mice part was 28.2 amino acids, whereas it was 24.0 and 23.0 in the GF-mice part and the overlapping part respectively.

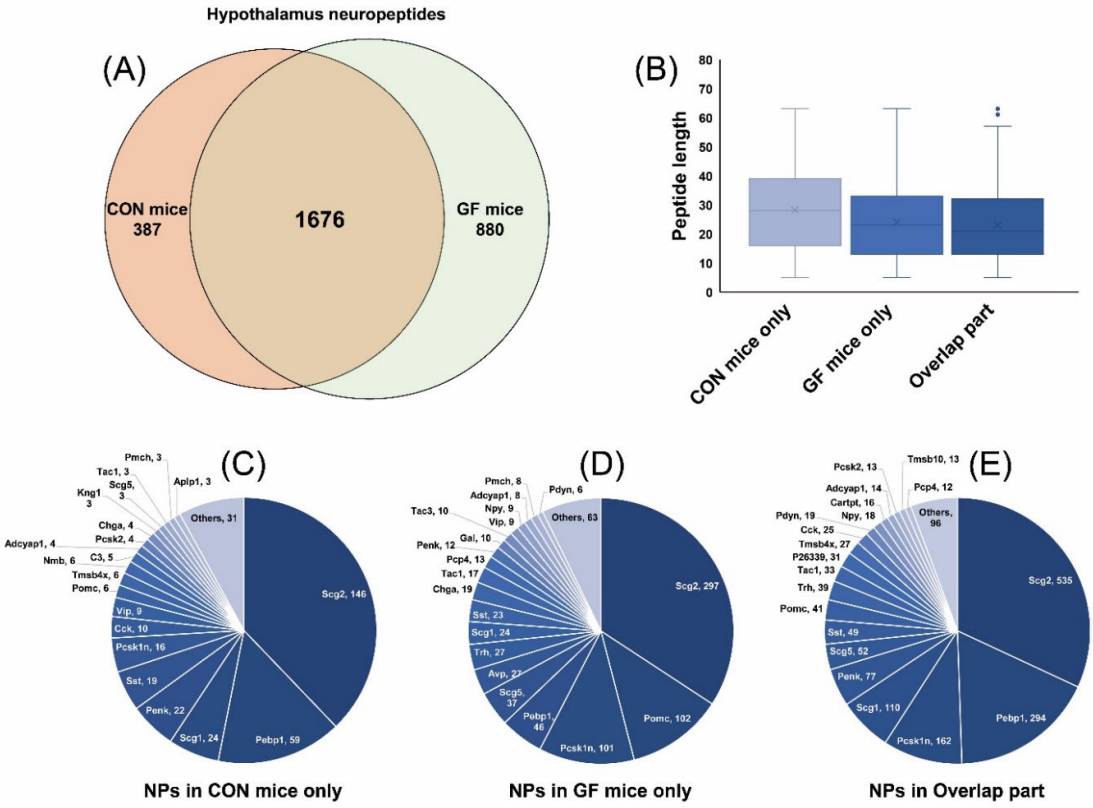


Figure 2. Distribution of the number of peptides per protein for the top 20 most abundant precursors from the hypothalamus region.

LFQ of endogenous peptides was carried out in this study to compare the relative quantitation of those peptides in the GF- and Con-mice hypothalamus. Peptides showed 1.5-fold changes, and a Student's *t*-test *p*-value lower than 0.05 was obtained, as shown in Figure S2A. In total, 368 peptides

that were significantly changed were selected (Table S2); representative neuropeptides and prohormone-derived peptides identified from the GF- and Con-mice hypothalamus are shown in Table S3. Among the 368 peptides, 271 peptides (73.6%) showed higher levels in GF-mice than in Con-mice, whereas 97 peptides (26.4%) showed higher levels in Con-mice than in GF-mice. These significantly changed peptides were mainly from SCG2 (82, 22.3%), PEBP1 (69, 18.8%), PCSK1N (72, 19.6%), SCG1 (20, 5.4%), and PENK (18, 4.9%). Peptides derived from SCG1, SCG2, and PCSK1N generally showed higher levels in GF-mice, whereas peptides derived from PEBP1 and PENK usually showed higher levels in Con-mice (Figure S3). Known bioactive neuropeptides were detected in the present study and quantified as well, such as orexin-B, joining peptide (J-peptide), corticotropin-like intermediate peptide (CLIP), γ-lipotropin, β-endorphin, α-MSH, neurokinin A, neurokinin B, galanin, substance P, nociceptin, Big LEN, Little LEN, Big SAAS, Little SAAS, and cerebellin. All these bioactive neuropeptides showed relatively low levels in Con-mice. As shown in Figure S2B.

As previously reported, neuropeptides in the hypothalamus region play important roles in affecting food-intake behavior[33], anxiety-like behavior and fear conditioning[16], body weight regulation[34], cardiac function, blood pressure regulation[35], circadian function[8], control of the sleep-wakefulness cycle, and energy homeostasis[36]. PCSK1N-derived peptides, such as Big SAAS, Little SAAS, PEN, and Big LEN, were significantly decreased after food intake, and some other neuropeptides, such as galanin and SCG2-derived peptides, were decreased as well. Chromogranin-A (CHGA)-, protachykinin-1 (TAC1)-, and CART prepropeptide (CARTPT)-derived peptides significantly increased after food intake[33]. Most of these food-intake involved neuropeptides were also detected in the present study, as shown in Figure S2C. Several mature neuropeptides derived from the PCSK1N

precursor, including Little LEN, Big LEN, Little SAAS, and Big SAAS, and also some truncated PCSK1N-derived neuropeptides, such as GAV [1-20], Big LEN [1-15], PEN [1-18], and Little SAAS [1-16], were quantified. These peptides were all down-regulated when microbiota existed in the mouse gut. PCSK1N-derived peptides have previously been reported and expressed at a relatively high level in the mouse hypothalamus region. PCSK1N can produce body weight changes through feeding control and fat deposition[34]. As one of the most abundant neuropeptides in the brain, PCSK1N-derived Big LEN and its receptor GPR171, which compose a Big LEN-GPR171 system, were involved in mood-related behaviors, such as anxiety-like and fear behaviors[16].

Galanin (which decreased by 51.4% in Con-mice) is derived from the precursor galanin detected in the present study, which was reported to be involved in the regulation of food intake and energy homeostasis[37]. The neuropeptides nociceptin and nociceptin [1-11] detected in the present study decreased by 46.4% and 58.1% in Con-mice, respectively. It has been previously reported that Nociceptin and its receptor can regulate learning and memory[38]. Decreasing the activation of the nociceptin receptor by using a specific endogenous agonist could enhance memory[38], but nociceptin or nociceptin receptor agonists might impair memory consolidation[39]. Therefore, the present results implied that the relatively low nociceptin level in Con-mice might help the mice enhance their learning and memory.

Besides the well-known mature neuropeptides, many peptides identified in the present study were truncated or longer forms. It is well known that neuropeptides are normally cleaved from a precursor molecule; a single precursor can be differentially cleaved into different sets of peptides to form mature neuropeptides[40]. These truncated peptides with differential cleavage levels might be related to distinct specificity and activity of the prohormone convertases involved, but they showed similar change

trends in the GF- and Con-mice hypothalamus. Take the neuropeptide orexin-B as an example. Orexin-B decreased by 54.4% in the Con-mice hypothalamus compared with GF-mice; similarly, truncated orexin-B forms, such as orexin-B [1-12], orexin-B [3-28], and orexin-B [14-28], were decreased by 56.4%, 51.7%, and 62.0% respectively. This trend was observed by the authors in truncated Somatostatin-28, truncated Big LEN, and Little SAAS as well. Notably, C-terminal amidation was observed in serval mature neuropeptides, such as orexin-B and C-terminal-amidated orexin-B, neurokinin B, and its C-terminal amidated form. Nevertheless, for the J-peptide, α-MSH, neurokinin A, substance P, and galanin neuropeptides, only their C-terminal amidated forms were identified in the present study. The level of peptide E (fragment) series derived from the precursor PENK was significantly elevated by over 150% in Con-mice. Besides, peptides derived from the precursor SCG1 (DPLQWKN, LFNPYFDPLQWKNSD, DGVAELDQLLHY, YDGVAELDQLLHY, etc.) were elevated from 52.4% to 266.3%. Precursor CCK-derived peptides APSGRMSVLKNLQSLDPSHRIS and APSGRMSVLKNLQSLDPS were elevated by 62.6% and 148.9%, whereas a peptide from SST, APSDPRLRQFLQKSLAAATGKQELAKYFLAE, was elevated by 263.1%.

Neuropeptides derived from POMC were decreased greatly in expression levels as well in response to the existence of gut microbiota, such as J-peptide (decreased by 86.1%), CLIP (decreased by 83.0%), γ-lipotropin (decreased by 80.2%), β-endorphin (decreased by 77.0%), and α-MSH (decreased by 74.8%). POMC-derived peptides in the present study, including γ-lipotropin, β-endorphin, and α-MSH, were decreased significantly in Con-mice. The POMC neuron works in association with energy homeostasis and regulating feeding behavior through releasing peptides[41]: α-MSH mediates the anorexigenic effect on feeding, whereas on the contrary, β-endorphin can produce

food-intake effects. These two neuropeptides might maintain a dynamic balance to regulate feeding behavior. It has also been reported that POMC-derived neuropeptides with feeding and obesity-suppressing functions were increased in the hypothalamus of Con-mice compared with GF-mice, which was consistent with the present investigation[28].

In addition, a large number of active neuropeptides, truncated neuropeptides, or endogenous peptides with unknown activities were detected and quantified. In total, 84 SCG2-derived peptides were quantified, and SCG2-derived peptides are known for their reproduction, food intake, and dopamine release functions[42]. Neuropeptides derived from TAC1 and TAC3, including neurokinin A, neurokinin B, substance P, and neuropeptide K [1-23], decreased by 58.6%, 48.8%, 49.0%, and 40.5%, respectively. TAC1-derived peptides, including neurokinin A, substance P, and neurokinin B (TAC3-derived) are involved with the suprachiasmatic nucleus and are associated with circadian rhythm[8]. In addition, in Con-mice, nociceptin, galanin, catestatin, and truncated MCH were decreased by 46.4%, 51.4%, 40.4%, and 40.9%, respectively.

It has already been reported that that GF-mice gain less weight than Con-mice even though their food intake and total fecal calories are similar[43]. GF-mice were completely diet-induced obesity-resistant when fed a cholesterol-rich lard-based high-fat diet. The mechanism protecting GF-mice from diet-induced obesity might involve the influence of microbiota on both sides of the energy homeostasis[44]. Gut microbiota are an important environmental factor that affects energy harvest from the diet and energy storage in the host[4]. Compared with Con-mice, GF-mice could protect against obesity after being fed a high-fat, sugar-rich diet, which might be associated with increasing skeletal muscle and liver levels of phosphorylated AMP-activated protein kinase (AMPK) and its downstream targets[44]. Therefore, in the present study, hypothalamic mature active neuropeptides showed different

levels in GF- and Con-mice, which might indicate one of the pathways by which gut microbiota affect food intake and energy homeostasis.

GF-mice actually showed increased spontaneous motor activity and decreased anxiety-like behavior compared with Con-mice[24]. Based on the changing profiles of neuropeptides under different gut microbiota environments, it could be speculated that these changes in brain development and behavior-related neuropeptides might act as signal molecules or might be associated with second messenger pathways and synaptic LTP in brain regions, which would lead to behavioral change. Furthermore, proteins, including receptors, kinases, and cytokines, are also important in responding to gut microbiota environment change.

Ten-plex isobaric DiLeu-tag labeled quantitative proteomic analysis in GF- and Con-mice

An example tandem MS spectrum of a DiLeu-labeled peptide sequence is shown in Figure S4. Using the DiLeu-labeled proteomic approach, a total of 3,971 proteins were identified, and 3,277 proteins were quantified, as shown in Table S5. Volcano plotting (Figure S5) was used to compare the fold changes ($\log_2$ ratio GF-/Con-) and the statistical significance ($-\log_{10}$ of p-value) for all these identifications. A p-value of < 0.05 and a 1.2 magnitude of change were set as thresholds for statistically significant identifications. Among these, the 282 up- or down-regulated proteins are listed in Table S6. Biological process, cellular component, molecular function enrichment analysis, and COG/KOG protein category analysis are shown in Figure S6.

Interestingly, biological process enrichment results showed that the significantly changed proteins were related to nervous system development and signal transduction, including CNS development,

intracellular signal transduction, trans-synaptic signaling, G-protein coupled receptor signaling pathways, and modulation of synaptic transmission; they may even be related to learning and memory (Figure S7A). The cell periphery, cell projection, neuron projection, and synapses were all involved in cellular component enrichment (Figure S7B). In molecular function enrichment (Figure S7C), proteins were enriched in various functions, including transporter activity, cation or metal ion trans-membrane transporter activity, and ligand-gated ion channel activity.

KEGG pathway enrichment analysis suggested that proteins were involved in different pathways. The main pathways that attracted attention were some signaling pathways (cGMP-PKG, oxytocin, cAMP, GnRH) and synapses (glutamatergic and dopaminergic synapses). Although some hormone secretions were obtained, such as bile, insulin, and pancreatic secretions, normally these secretions take place in the gut, not in the brain, and therefore what was observed should have been those pathways taking place in the brain. In addition, LTP, circadian entrainment pathways were obtained. This showed that proteins were involved in these important pathways. Considering the circadian rhythm pathway, for instance, as shown in Figure 3A, proteins including glutamate receptor (AMPAR), glutamate receptor ionotropic, NMDA (NMDAR), inositol 1,4,5-trisphosphate receptor (IR_3P), ryanodine receptor (RyR), guanine nucleotide-binding proteins, adenylate cyclase (AC), 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase (PLC), protein kinase C (PKC), and calcium/calmodulin-dependent protein kinase (CaMKII) were involved in this pathway. Interestingly, in the circadian rhythm pathway, PACAP was also identified in the neuropeptide investigation. PACAP and glutamate play a co-regulated role in modulating the circadian clock[45]. In addition, neuropeptides have been reported to change significantly during the natural circadian rhythm. For example, the PEN-LEN region derived from PCSK1N was significantly more abundant at nighttime,

whereas the truncated GAV peptide was significantly more abundant in the daytime. When the neuropeptides and proteins detected in the present study are combined, it becomes evident that gut microbiota can affect proteins and neuropeptides in the brain, which may be associated with circadian rhythm-related physiological changes. In another KEGG pathway instance, LTP and long-term depression (LTD) attracted attention. As shown in Figure 3B, significantly changed proteins in the hypothalamus were also involved in the LTP and LTD pathways. LTP is the process of strengthening the connections between two neurons, which is associated with learning and memory. LTD has long been considered to be an important contributor to motor learning and memory.

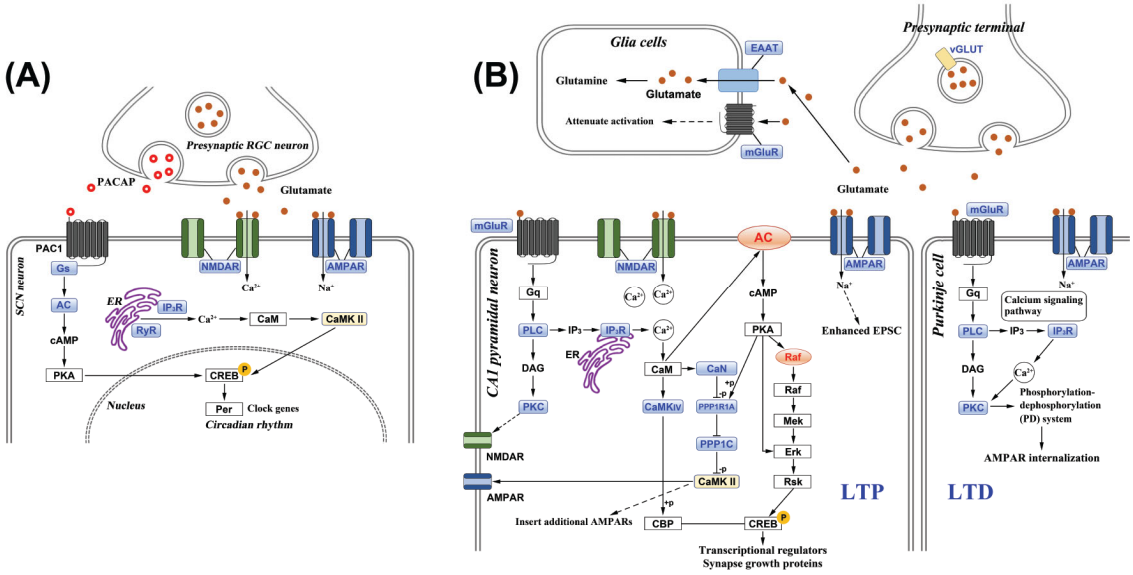


Figure 3. Neuropeptides and proteins involved in pathways: (A) circadian rhythm pathway, (B) glutamatergic synapse, LTP, and LTD pathways.

In summary, although it has been reported that gut microbiota can affect food intake, anxiety-like behavior, memory, and learning in mice, what leads to these physiological phenomena is still

attracting scientists' attention. Gut microbiota are associated with impaired learning and memory. GF-mice display normal anxiety levels, but exhibit an absence of non-spatial and working memory, proving that gut microbiota are important for memory development[46]. GF-mice showed more motor activity than Con-mice[24]. GF-mice had lower body and adipose tissue weights compared with age-matched specific pathogen-free (SPF) mice; GF-mice exhibited decreased food intake levels at 4 weeks old, but exhibited increased food intake levels compared with SPF-mice at the age of 8 months[47]. The present study has shown that under different gut microbiota environments, mice exhibit different, or even opposite, behaviors, which might be associated with changes in hypothalamic proteins and neuropeptides.

Conclusions

LFQ-based neuropeptidomics analysis and DiLeu-label-based proteomics analysis have been used to compare the relative changes in neuropeptides and proteins in the hypothalamus region of GF- and Con-mice. Interestingly, neuropeptides and proteins that significantly changed were identified, and the functions of the neuropeptides and the pathways of the proteins involved were related with mouse behaviors, brain development, memory, and learning. A label-free ten-plex DiLeu integrated quantitative strategy combined with high-resolution MS has been used to simultaneously identify and quantify a surprisingly rich neuropeptide and protein presence. The fact that many of these peptides and proteins change levels suggests that multiple peptides and proteins are likely to be associated with food intake, the circadian clock, memory, and learning. The present analysis confirmed the importance of gut microbiota in regulating brain component changes. As an analogy, based on the

present method, signaling molecules including neuropeptides and proteins in other brain regions can be comprehensively identified and quantified, and even the relationship of important signaling molecules among these different regions can be elucidated as well. Further studies should focus on validation of these neuropeptide and protein changes and even on changes in these components in other brain regions.

Acknowledgement

LL acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. This study was also funded by the National Natural Science Foundation of China (No. 81973450), Jiangsu Qinglan Project, Jiangsu "333" Project, Young Researchers Training Project of China Association of Traditional Chinese Medicine (QNRC2-C14).

References

(1) DeLaney, K.; Buchberger, A. R.; Atkinson, L.; Grunder, S.; Mousley, A.; Li, L. *J Exp Biol* **2018**, *221*.

(2) Soto, M.; Herzog, C.; Pacheco, J. A.; Fujisaka, S.; Bullock, K.; Clish, C. B.; Kahn, C. R. *Mol Psychiatry* **2018**, *23*, 2287-2301.

(3) Desbonnet, L.; Clarke, G.; Shanahan, F.; Dinan, T. G.; Cryan, J. F. *Mol Psychiatry* **2014**, *19*, 146-148.

(4) Backhed, F.; Ding, H.; Wang, T.; Hooper, L. V.; Koh, G. Y.; Nagy, A.; Semenkovich, C. F.; Gordon, J. I. *Proc Natl Acad Sci U S A* **2004**, *101*, 15718-15723.

(5) Cani, P. D.; Bibiloni, R.; Knauf, C.; Waget, A.; Neyrinck, A. M.; Delzenne, N. M.; Burcelin, R. *Diabetes* **2008**, *57*, 1470-1481.

(6) Fujisaka, S.; Ussar, S.; Clish, C.; Devkota, S.; Dreyfuss, J. M.; Sakaguchi, M.; Soto, M.; Konishi, M.; Softic, S.; Altindis, E.; Li, N.; Gerber, G.; Bry, L.; Kahn, C. R. *Journal of Clinical Investigation* **2016**, *126*, 4430-4443.

(7) van den Pol, A. N. *Neuron* **2012**, *76*, 98-115.

(8) Southey, B. R.; Lee, J. E.; Zamdborg, L.; Atkins, N., Jr.; Mitchell, J. W.; Li, M.; Gillette, M. U.; Kelleher, N. L.; Sweedler, J. V. *Anal Chem* **2014**, *86*, 443-452.

(9) Shao, W.; Lam, H. *Mass Spectrom Rev* **2017**, *36*, 634-648.

(10) Chen, R.; Xiao, M.; Buchberger, A.; Li, L. *J Proteome Res* **2014**, *13*, 5767-5776.

(11) Kormos, V.; Gaszner, B. *Neuropeptides* **2013**, *47*, 401-419.

(12) Frohlich, E. E.; Farzi, A.; Mayerhofer, R.; Reichmann, F.; Jacan, A.; Wagner, B.; Zinser, E.; Bordag, N.; Magnes, C.; Frohlich, E.; Kashofer, K.; Gorkiewicz, G.; Holzer, P. *Brain Behav Immun* **2016**, *56*, 140-155.

(13) Farzi, A.; Reichmann, F.; Holzer, P. *Acta Physiol (Oxf)* **2015**, *213*, 603-627.

(14) Rat, D.; Schmitt, U.; Tippmann, F.; Dewachter, I.; Theunis, C.; Wieczerzak, E.; Postina, R.; van Leuven, F.; Fahrenholz, F.; Kojro, E. *FASEB J* **2011**, *25*, 3208-3218.

(15) Gomes, I.; Aryal, D. K.; Wardman, J. H.; Gupta, A.; Gagnidze, K.; Rodriguiz, R. M.; Kumar, S.; Wetsel, W. C.; Pintar, J. E.; Fricker, L. D.; Devi, L. A. *Proc Natl Acad Sci U S A* **2013**, *110*, 16211-16216.

(16) Bobeck, E. N.; Gomes, I.; Pena, D.; Cummings, K. A.; Clem, R. L.; Mezei, M.; Devi, L. A. *Neuropsychopharmacology* **2017**, *42*, 2527-2536.

(17) Neufeld, K. M.; Kang, N.; Bienenstock, J.; Foster, J. A. *Neurogastroenterol Motil* **2011**, *23*, 255-264, e119.

(18) Zeng, L.; Zeng, B.; Wang, H.; Li, B.; Huo, R.; Zheng, P.; Zhang, X.; Du, X.; Liu, M.; Fang, Z.; Xu, X.; Zhou, C.; Chen, J.; Li, W.; Guo, J.; Wei, H.; Xie, P. *Sci Rep* **2016**, *6*, 29998.

(19) Valdes, A. M.; Walter, J.; Segal, E.; Spector, T. D. *BMJ* **2018**, *361*, k2179.

(20) Levy, M.; Kolodziejczyk, A. A.; Thaiss, C. A.; Elinav, E. *Nature Reviews Immunology* **2017**, *17*, 219-232.

(21) Rothschild, D.; Weissbrod, O.; Barkan, E.; Kurilshikov, A.; Korem, T.; Zeevi, D.; Costea, P. I.; Godneva, A.; Kalka, I. N.; Bar, N.; Shilo, S.; Lador, D.; Vila, A. V.; Zmora, N.; Pevsner-Fischer, M.; Israeli, D.; Kosower, N.; Malka, G.; Wolf, B. C.; Avnit-Sagi, T., et al. *Nature* **2018**, *555*, 210-215.

(22) Clarke, G.; Grenham, S.; Scully, P.; Fitzgerald, P.; Moloney, R. D.; Shanahan, F.; Dinan, T. G.; Cryan, J. F. *Mol Psychiatry* **2013**, *18*, 666-673.

(23) Luo, Y.; Zeng, B.; Zeng, L.; Du, X.; Li, B.; Huo, R.; Liu, L.; Wang, H.; Dong, M.; Pan, J.; Zheng, P.; Zhou, C.; Wei, H.; Xie, P. *Transl Psychiatry* **2018**, *8*, 187.

(24) Diaz Heijtz, R.; Wang, S.; Anuar, F.; Qian, Y.; Bjorkholm, B.; Samuelsson, A.; Hibberd, M. L.; Forssberg, H.; Pettersson, S. *Proc Natl Acad Sci U S A* **2011**, *108*, 3047-3052.

(25) Bercik, P.; Denou, E.; Collins, J.; Jackson, W.; Lu, J.; Jury, J.; Deng, Y.; Blennerhassett, P.; Macri, J.; McCoy, K. D.; Verdu, E. F.; Collins, S. M. *Gastroenterology* **2011**, *141*, 599-609, 609 e591-593.

(26) Holzer, P.; Farzi, A. *Adv Exp Med Biol* **2014**, *817*, 195-219.

(27) Luan, H.; Wang, X.; Cai, Z. *Mass Spectrom Rev* **2019**, *38*, 22-33.

(28) Schele, E.; Grahnemo, L.; Anesten, F.; Hallen, A.; Backhed, F.; Jansson, J. O. *Endocrinology* **2013**, *154*, 3643-3651.

(29) Vij, R.; Reddi, S.; Kapila, S.; Kapila, R. *Food Chem* **2016**, *190*, 681-688.

(30) Wong, M. L.; Inserra, A.; Lewis, M. D.; Mastronardi, C. A.; Leong, L.; Choo, J.; Kentish, S.; Xie, P.; Morrison, M.; Wesselingh, S. L.; Rogers, G. B.; Licinio, J. *Mol Psychiatry* **2016**, *21*, 797-805.

(31) Frost, D. C.; Greer, T.; Li, L. *Anal Chem* **2015**, *87*, 1646-1654.

(32) Zhang, Y.; DeLaney, K.; Hui, L.; Wang, J.; Sturm, R. M.; Li, L. *J Am Soc Mass Spectrom* **2018**, *29*, 948-960.

(33) Ye, H.; Wang, J.; Tian, Z.; Ma, F.; Dowell, J.; Bremer, Q.; Lu, G.; Baldo, B.; Li, L. *Molecular & Cellular Proteomics* **2017**, *16*, 1922-1937.

(34) Morgan, D. J.; Wei, S.; Gomes, I.; Czyzyk, T.; Mzhavia, N.; Pan, H.; Devi, L. A.; Fricker, L. D.; Pintar, J. E. *J Neurochem* **2010**, *113*, 1275-1284.

(35) Loh, Y. P.; Cheng, Y.; Mahata, S. K.; Corti, A.; Tota, B. *J Mol Neurosci* **2012**, *48*, 347-356.

(36) Nuñez, A.; Rodrigo-Angulo, M. L.; Andrés, I. D.; Garzón, M. *Current Neuropharmacology* **2009**, *7*, 50-59.

(37) Leibowitz, S. F. *Neuropeptides* **2005**, *39*, 327-332.

(38) Andero, R. *Prog Neuropsychopharmacol Biol Psychiatry* **2015**, *62*, 45-50.

(39) Roozendaal, B.; Lengvilas, R.; McGaugh, J. L.; Civelli, O.; Reinscheid, R. K. *Learn Mem* **2007**, *14*, 29-35.

(40) Li, C.; Kim, K. *WormBook* **2008**, 1-36.

(41) Wei, Q.; Krolewski, D. M.; Moore, S.; Kumar, V.; Li, F.; Martin, B.; Tomer, R.; Murphy, G. G.; Deisseroth, K.; Watson, S. J., Jr.; Akil, H. *Proc Natl Acad Sci U S A* **2018**, *115*, E9489-E9498.

(42) Bartolomucci, A.; Possenti, R.; Mahata, S. K.; Fischer-Colbrie, R.; Loh, Y. P.; Salton, S. R. *Endocr Rev* **2011**, *32*, 755-797.

(43) Rabot, S.; Membrez, M.; Bruneau, A.; Gerard, P.; Harach, T.; Moser, M.; Raymond, F.; Mansourian, R.; Chou, C. J. *FASEB J* **2010**, *24*, 4948-4959.

(44) Backhed, F.; Manchester, J. K.; Semenkovich, C. F.; Gordon, J. I. *Proc Natl Acad Sci U S A* **2007**, *104*, 979-984.

(45) Chen, D.; Buchanan, G. F.; Ding, J. M.; Hannibal, J.; Gillette, M. U. *Proc Natl Acad Sci U S A* **1999**, *96*, 13468-13473.

(46) Gareau, M. G.; Wine, E.; Rodrigues, D. M.; Cho, J. H.; Whary, M. T.; Philpott, D. J.; Macqueen, G.; Sherman, P. M. *Gut* **2011**, *60*, 307-317.

(47) Niimi, K.; Takahashi, E. *Heliyon* **2019**, *5*, e02176.

Supplementary Information

Experimental details

Chemicals and Reagents

All isotopic reagents used for the synthesis of labels were purchased from Isotec (Miamisburg, OH). Mass spec grade trypsin and dithiothreitol (DTT) were purchased from Promega (Madison, WI). Urea, ammonium bicarbonate, ACS grade methanol (MeOH), dichloromethane (DCM), Optima UPLC grade ACN, Optima UPLC grade water, and Optima LC/MS grade formic acid were purchased from Fisher Scientific (Pittsburgh, PA). Sodium cyanoborohydride ($NaBH_3CN$), L-leucine, formaldehyde (CH_2O), hydrogen chloride gas (HCl), iodoacetamide (IAA), tris hydrochloride, trifluoroacetic acid (TFA), triethylammonium bicarbonate (TEAB), N,N-dimethylformamide (DMF), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium tetrafluoroborate (DMTMM), N-methylmorpholine (NMM), heptafluorobutyric acid (HFBA), dimethyl sulfoxide (DMSO), and

bovine serum albumin (BSA) were purchased from Sigma-Aldrich (St. Louis, MO). Hydroxylamine solution was purchased from Alfa Aesar (Ward Hill, MA). DiLeu Reporter was synthesized in our lab based on our previous work[1].

Animal Experiment and Tissue Extraction

Five GF-mice and five Con-mice were all sacrificed and brains were dissected and immediately rapid heated via Denator™ to minimize postmortem degradation, and then the hypothalamus region was isolated. The processed brain tissues were then extracted by probe sonicator into ice-cold acidified methanol (90:10:1, MeOH: water: acetic acid) as Ye's described[2]. The homogenized sample was then spun at 14,000 × g for 15 min at 4 °C. Protein concentration of the pellet was determined for each sample using bicinchoninic acid (BCA) assay from Pierce (Rockford, IL) and used to adjust the differences in the neuropeptide levels contained in different sample aliquots. The adjusted supernatant was decanted and then dried in a vacuum centrifuge. Extracts were re-suspended in 20 µl 0.1% formic acid aqueous solution by vortexing and brief sonication. Subsequently, the reconstituted samples were purified and concentrated by C18 OMIX tips (Agilent Technologies, Santa Clara, CA, USA). Tips were first wetted with ACN and equilibrated with water containing 0.1% TFA. Samples were applied onto the tip, washed with water containing 0.1% TFA, and then eluted with 50% and 80% aqueous ACN containing 0.1% TFA. Samples were dried down and loaded on LC-MS.

For proteomics analysis, pellets after neuropeptide extraction were lysed by sonication in a solution containing digest buffer (4% SDS, 100 mM Tris/Base pH 8.0). The BCA assay was applied to determine the protein concentration. Then 200 µg protein was incubated 0.5 h after adding DTT to 10 mM, and then samples were alkylated with 55 mM iodoacetamide (IAA) in the dark for 30 min,

and the reaction was quenched by 5 mM DTT. Proteins were precipitated from the soluble extract by addition of 5.5 volumes of 80% acetone overnight and rinsed with 80% acetone, and then dry in air. Dried pellet was resuspended in 8 M urea, 30 mM NaCl, 50 mM Tris (pH = 8), 5 mM $CaCl_2$, and one protease inhibitor cocktail tablet (Roche Diagnostics). The sample was sonicated in an ice-water bath for 20 min, centrifuged at 14,000 × g for 5 min, and the supernatant was collected. After that, the supernatant was diluted with 50 mM Tris solution (pH = 8) to a urea concentration of less than 1 M. Protein digestion was performed with trypsin enzyme at a 50:1 protein/enzyme ratio at 37 °C for 16 h. The digestion was quenched by adding 10% trifluoroacetic acid (TFA) to a pH lower than 3, followed by a desalting step with SepPak C18 solid-phase extraction cartridge (Waters).

Detailed syntheses of 10-Plex DiLeu tags were reported by Dustin[1], DiLeu labeling was performed by addition of labeling solution at a 20:1 label to peptide digest ratio by weight and vortexing at room temperature for 2 h. Peptides were randomly labeled with DiLeu reagents, DiLeu reagents (115a, 115b, 116a, 116b, 116c, 117a, 117b, 117c, 118a, and 118b) were used to label the digests from the different brain regions of GF-mice (labeled with 116a, 116b, 116c, 117a, and 117b) and Con-mice (labeled with 115a, 115b, 117c, 118a, and 118b), respectively. The labeling reaction was quenched by addition of hydroxylamine to a concentration of 0.25%, and the peptide mixtures were pooled in the same ratio. The pooled labeled peptide samples were dried down. The combined samples were then acidified with HFBA to a concentration of 0.5%, cleaned with SCX SpinTips (Protea Biosciences, Morgantown, WV) to remove unreacted DiLeu reagent and reaction byproducts from the labeled peptides, and then pooled samples were separated into four fractions by high pH fractionation using C18 column (Phenomenex).

Neuropeptide Analysis

A Waters nano Acquity ultraperformance liquid chromatography (UPLC) (Waters Corp., Milford, MA) was coupled to a Thermo Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) for LC-MS/MS analysis. Chromatographic separations were performed on a Waters BEH 130 Å C18 reversed-phase capillary column (150 mm × 75 μm, 1.7 μm). The mobile phases used were: mobile phase A consisted of water with 0.1% FA, and mobile phase B was composed of ACN with 0.1% FA. Samples were injected and loaded onto the Waters NanoACQ 2G-V/M Sym C18 (20 mm × 180 μm, 5 μm) using 100% A at a flow rate of 5 μl/min for 1 min. Then the peptides were separated using a solvent gradient of 0-10% B over 0.5 min and then 10-35% B over 90 min at a flow rate of 300 nL/min. Data-dependent acquisition (DDA) parameters recorded MS scans in profile mode from m/z 200–2000 at a resolution of 70,000. Automatic gain control (AGC) targets of 1×10^6 and maximum injection times (IT) of 250 ms were set. The 15 most intense precursor ions were selected for MS^2 higher-energy collisional dissociation (HCD) fragmentation with an isolation window of 2 Da and dynamic exclusion set at 40 s. An AGC target of 2×10^5 and a maximum IT of 120 ms was selected for tandem mass acquisition. The tandem MS spectra were acquired at a resolution of 17,500 in profile mode, with normalized collision energy (NCE) set at 27, and a fixed lower mass at m/z 110.

DiLeu Label Proteomic Analysis

Samples were analyzed using a Dionex Ultimate 3000 nanoLC system (Thermo Scientific) coupled to a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (QE HF, Thermo Fisher Scientific, San Jose, CA). Labeled tryptic peptide samples were dried in vacuo and dissolved in 3% ACN, 0.1% formic acid in water. Peptides were loaded onto a 75 μm inner diameter microcapillary column fabricated with an integrated emitter tip and packed with 15 cm of Bridged Ethylene Hybrid C18

particles (1.7 μm, 130 Å, Waters). Mobile phase A consisted of water with 0.1% FA, and mobile phase B was composed of ACN with 0.1% FA. Separation was performed using a gradient elution of 3% to 30% mobile phase B over 120 min at a flow rate of 300 nL/min. Survey scans of peptide precursors from 300 to 1500 m/z were performed at a resolving power of 60,000 with an AGC target of 1×10^6 and maximum injection time of 100 ms. The top 10 precursors were then selected for higher-energy C-trap dissociation tandem mass spectrometry (HCD MS^2) analysis with an isolation width of 2.0 Da, a normalized collision energy (NCE) of 30, a resolving power of 60,000, an AGC target of 1×10^5, a maximum injection time of 100 ms. Precursors were subject to dynamic exclusion for 15 s with a 10 ppm tolerance.

Data Analysis

All the raw LC-MS/MS data were imported to PEAKS Studio 8.5 (BSI, Waterloo, Ontario, Canada) for peptide identification. Data processing procedures, including peak centroiding and charge deconvolution, were conducted to refine the raw data. To perform a neuropeptide search using the PEAKS DB algorithm, the enzyme was specified as none. The peptide mass tolerance was set at 10 ppm and the MS/MS mass tolerance was set at 20 ppm. The variable modifications were set to include amidation (C-terminal), acetylation (protein N-terminal), and Gln->pyro-Glu (N-terminal). No enzyme specificity was required. The database search was conducted using SwePep neuropeptide database (download on the October 6th, 2018). The cutoff of false discovery rate (FDR) for peptide identification was set below 1%. Peptide spectral matches (PSMs) with a –10logP value cutoff of 30 in PEAKS were considered for further manual inspection. All the spectra used for identification assigned by PEAKS were subsequently manually checked. For PTM assignment, a minimum of three consecutive b or y ions must be present in the MS/MS spectra.

Label-free quantitation (LFQ) method was applied to calculate the relative expression level changes by comparing the peaks areas calculated using extracted ion chromatograms (XIC). Such differential analysis was conducted using PEAKS Studio 8.5. The acquired LC-MS/MS data were first aligned by the software. Using one of the samples as a reference, the peak areas of each identified peptides were normalized using Tubulins as internal standard proteins, and then extracted, and the differential analysis was conducted on the GF- and Con-mice samples region by region. A t-test was performed between GF- and Con-mice groups, and any pair at $p < 0.05$ was considered as statistically significant. The parameters used were as follows: m/z range: 200-2000; retention time window = 6 min; ion m/z width = 10 ppm; retention time range: 5-80 min.

For DiLeu labeled quantitative analysis, mass spectra were processed using MaxQuant (1.6.0.1). Raw files were searched in MaxQuant UniProt mouse complete database (download on the February 3rd, 2019) using Sequest HT algorithm with trypsin selected as the enzyme and two missed cleavages allowed. Searches were performed with a precursor mass tolerance of 25 ppm and a fragment mass tolerance of 0.03 Da. Static modifications consisted of DiLeu labels on peptide N-termini (+145.12801 Da) and carbamidomethylation of cysteine residues (+57.02146 Da). Dynamic modifications consisted of DiLeu labels on lysine residues, oxidation of methionine residues (+15.99492 Da), deamidation of asparagine and glutamine residues (+0.98402 Da), and methylation of C termini and aspartic acid, glutamic acid, histidine, lysine, arginine, serine, and threonine residues (+14.01565 Da). PSMs were validated based on q-values to 1% FDR (false discovery rate) using percolator. Quantitation was performed in MaxQuant with a reporter ion integration tolerance of 20 ppm for the most confident centroid. Only the PSMs that contained all 10 reporter ions were considered, and protein quantitative ratios were determined using a minimum of one quantified

peptide. Reporter ion ratio values for protein groups were exported to Excel workbook format (Redmond, CA). The UniProt Gene Ontology Annotation database (http://www.ebi.ac.uk/GOA/) was used. Quantification of neuropeptides and proteins from different samples was analyzed by Excel and SPSS 16.0 software.

Supplementary Figures

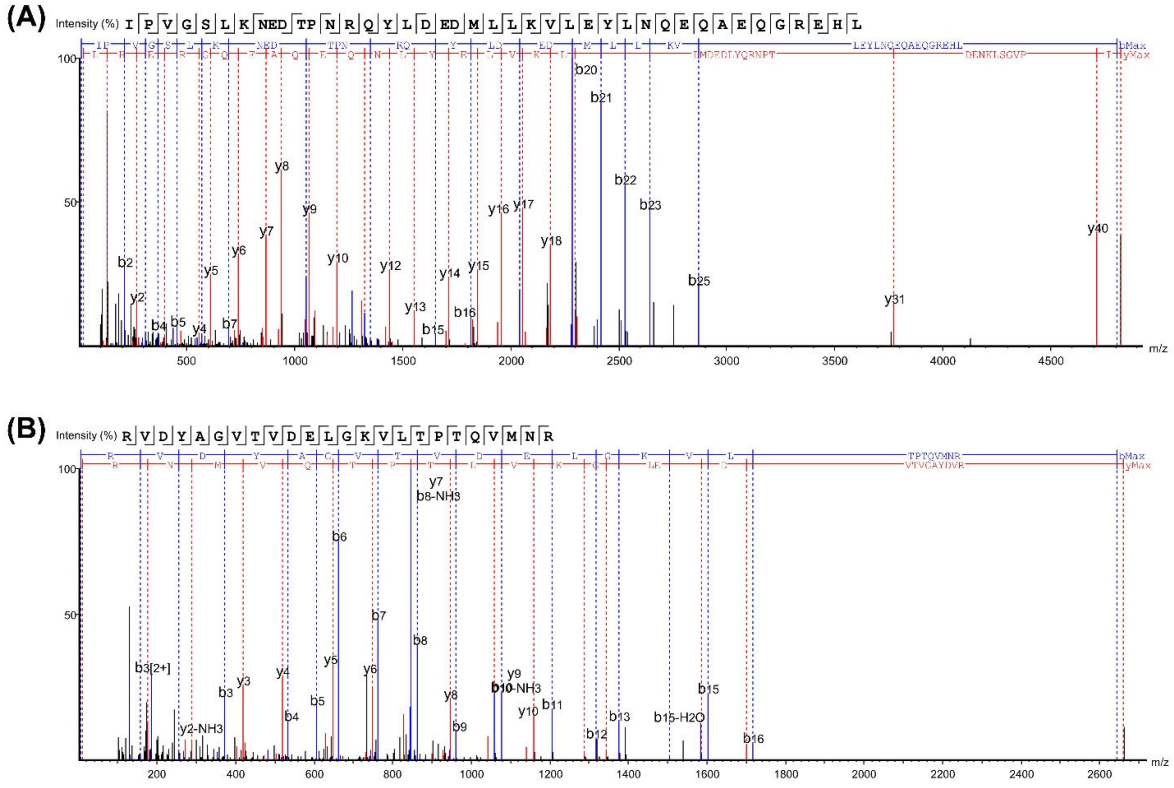


Figure S1. (A) MS/MS spectra of a 5+ charged peptide ion were detected at *m/z* 966.2924. Using a stringent search criterion, a mass tolerance for precursor ions at 10 ppm, and one for product ions at 20 ppm, the precursor ion was identified as IPVGSLKNEDTPNRQYLDEDMLLKVLEYLNQEQAEQGREHL from the neuropeptide precursor secretogranin-2 with high confidence (−10logP score of 151.09). (B) A 4+ charged ion at *m/z* 666.3583 was characterized as peptide derived from a phosphatidylethanolamine-binding protein-1 with the sequence RVDYAGVTVDELGKVLTPTQVMNR (−10logP score of 131.07).

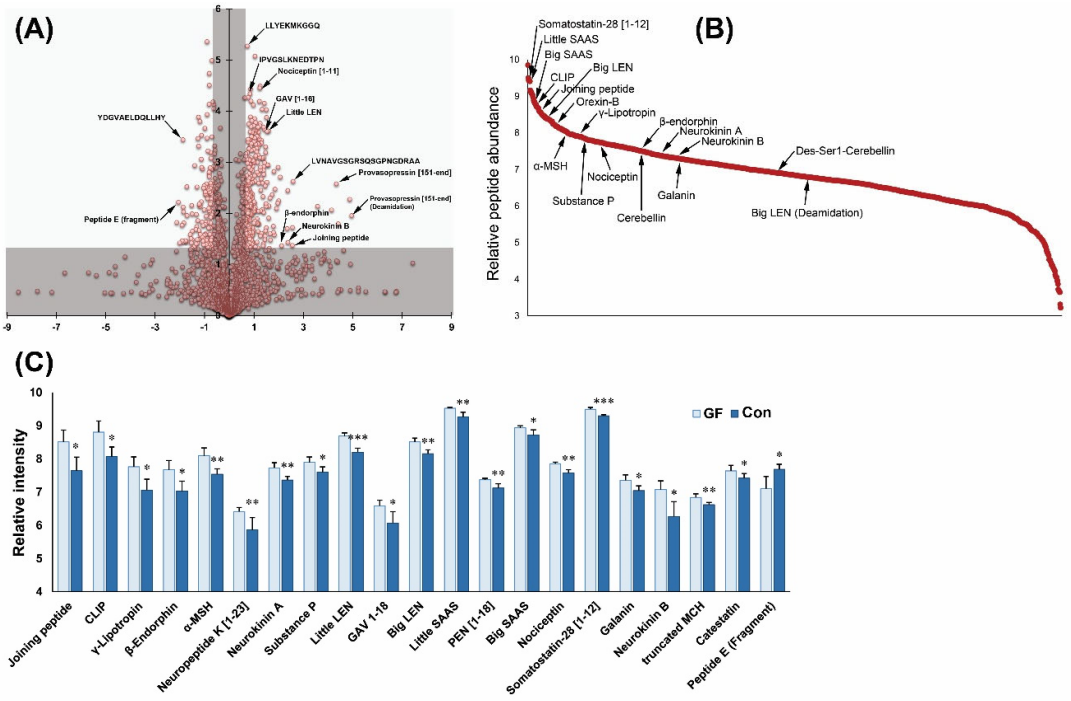


Figure S2. (A) Volcano plots showing $-\log_{10}$ *p*-values plotted against $\log_2$ ratio (GF-/Con-). The horizontal line represents the *t*-test threshold of significance assigned ($p < 0.05$). The vertical lines mark the threshold of 1.5-fold up and down-regulated peptides. Some peptides had been reported previously, such as LLYEKMKGGQ (SCG5), IPVGSLKNEDTPN (SCG2), LVNAVGSGRSQSGPNGDRAA (SCG2), and YDGVAELDQLLHY (SCG1). (B) Relative abundance of endogenous peptides from the hypothalamus (GF-/Con-). For each peptide, the peak area was calculated and $\log_{10}$ transformed, somatostatin-28 [1–12], Little SAAS, Big SAAS, CLIP, Big LEN, orexin-B, and α-MSH were present in the upper range, whereas deamidated Big LEN, galanin, neurokinin A, neurokinin B, β-endorphin, and cerebellin were present in the lower range. (C) Ratio of the relative abundance of endogenous peptides as determined by LFQ in GF- and Con-mice. For each peptide, the peak area was calculated and $\log_{10}$ transformed.

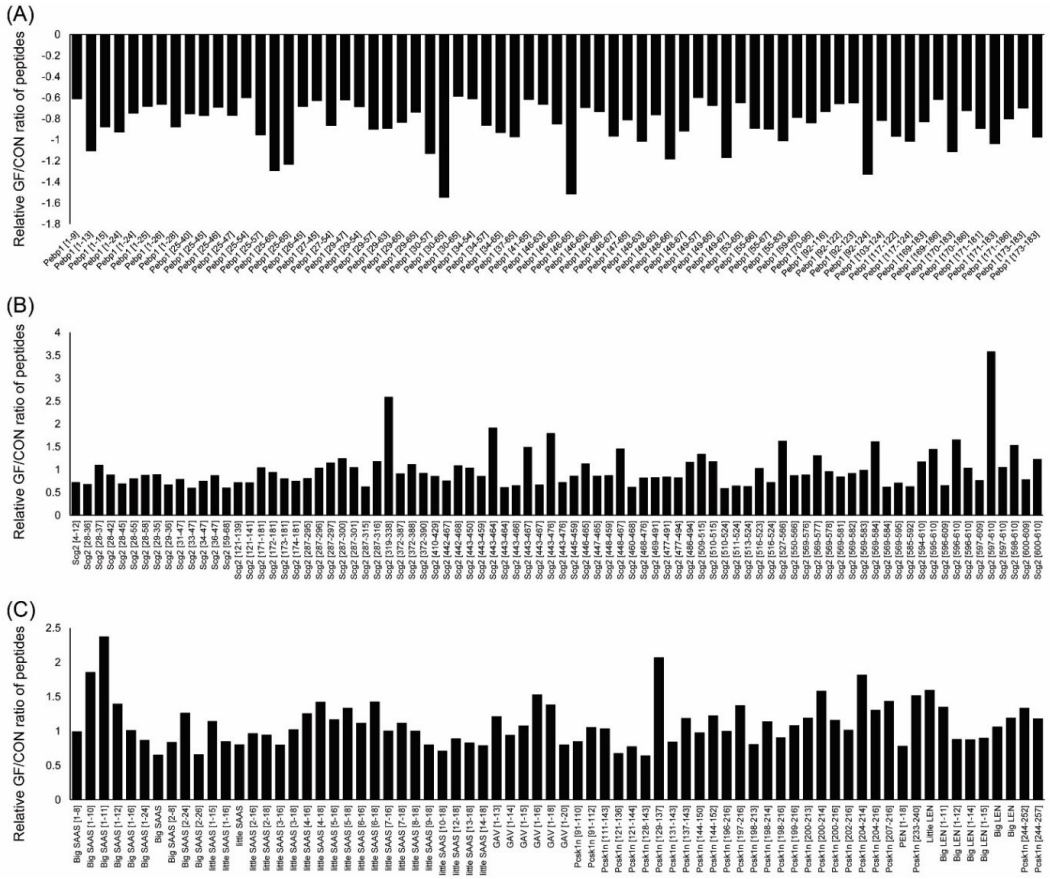


Figure S3. Relative GF/Con ratio of peptides from PCSK1N, SCG2and PEBP1. PCSK1N and SCG2-derived peptides were decreased in the Con-mice hypothalamus compared with GF-mice. In contrast, those PEBP1-derived peptides in Con-mice were significantly elevated from 50.2% to 191.9% in the hypothalamus compared with GF-mice.

Even though most of the SCG2- and PEBP1-derived peptides had not been reported yet, in the present study, some function-related SCG2-derived peptides related to food-intake behaviors were detected, such as ANQIPKVAWIPDVES, ASFQRNQLLQKEPDLRLE, ASFQRNQLLQKEPDLRLENVQKFPSPEM, and LDEDMLLK[2]. The SCG2-derived neuropeptide manserin was detected in the present study, which was identified in Zhang and Southey's paper as well[3,4]. In addition, the IPVGSLKNEDTPN and IPVGSLKNEDTPNRQYLDEDMLLKVLE peptides were also detected in Southey's paper and showed differential abundance in daytime and nighttime, which indicated that some SCG2-derived peptides might participate in circadian rhythm regulation[4]. The PEBP1-derived homologous peptides QAEWDDYVPKLYEQLSGK, QAEWDDYVPKLYEQL, and KGNDISSGTVLSDYVGSGPPSGTGL were also reported in Southey's paper[4].

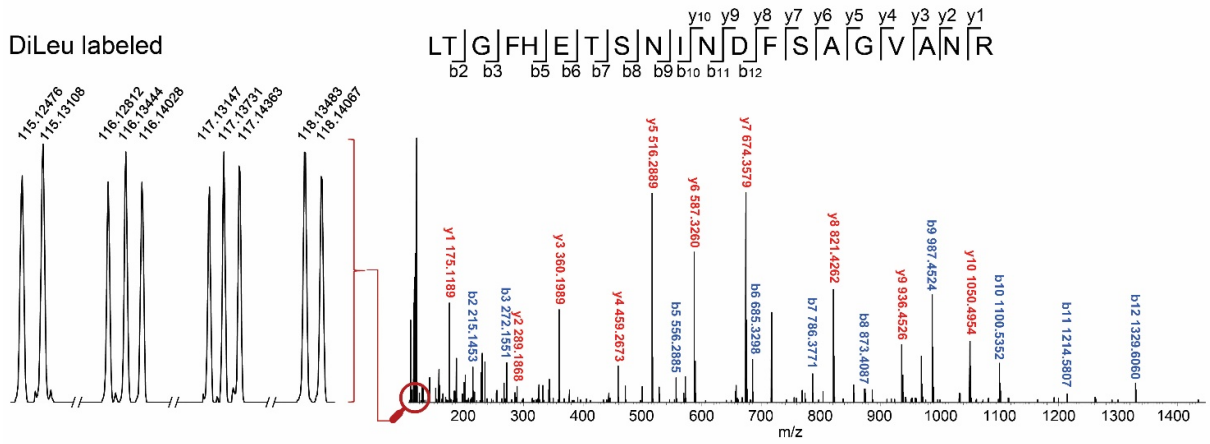


Figure S4. An example of identification and quantification of a DiLeu-labeled peptide. Ten reporter ions of DiLeu represent the relative abundance of peptide LTG…NR in ten different hypothalamus samples.

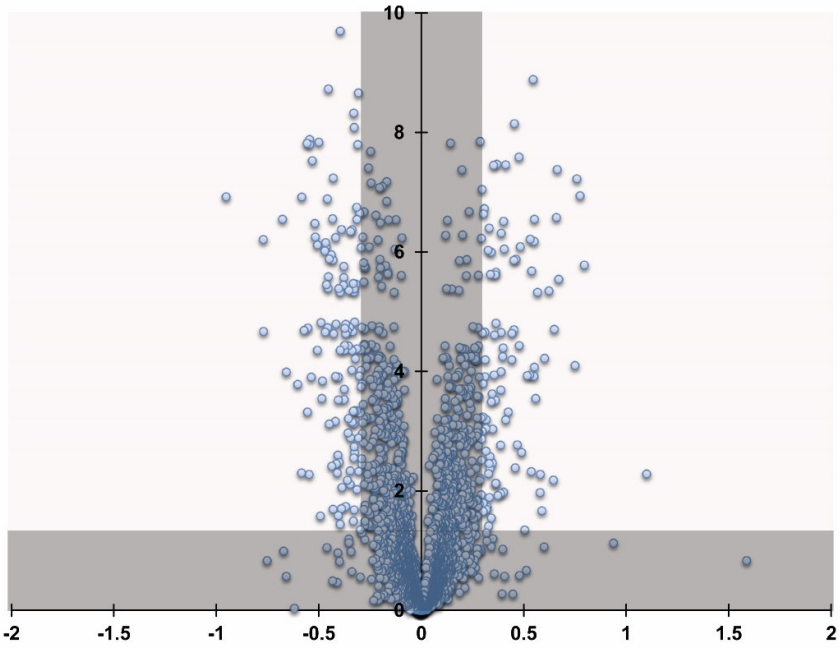


Figure S5. Volcano plots showing −log10 p-values plotted against log2 ratio (GF-/Con-). The horizontal line represents the *t*-test threshold of significance assigned ($p < 0.05$). The vertical lines mark the threshold of 1.2-fold up and down-regulated proteins.

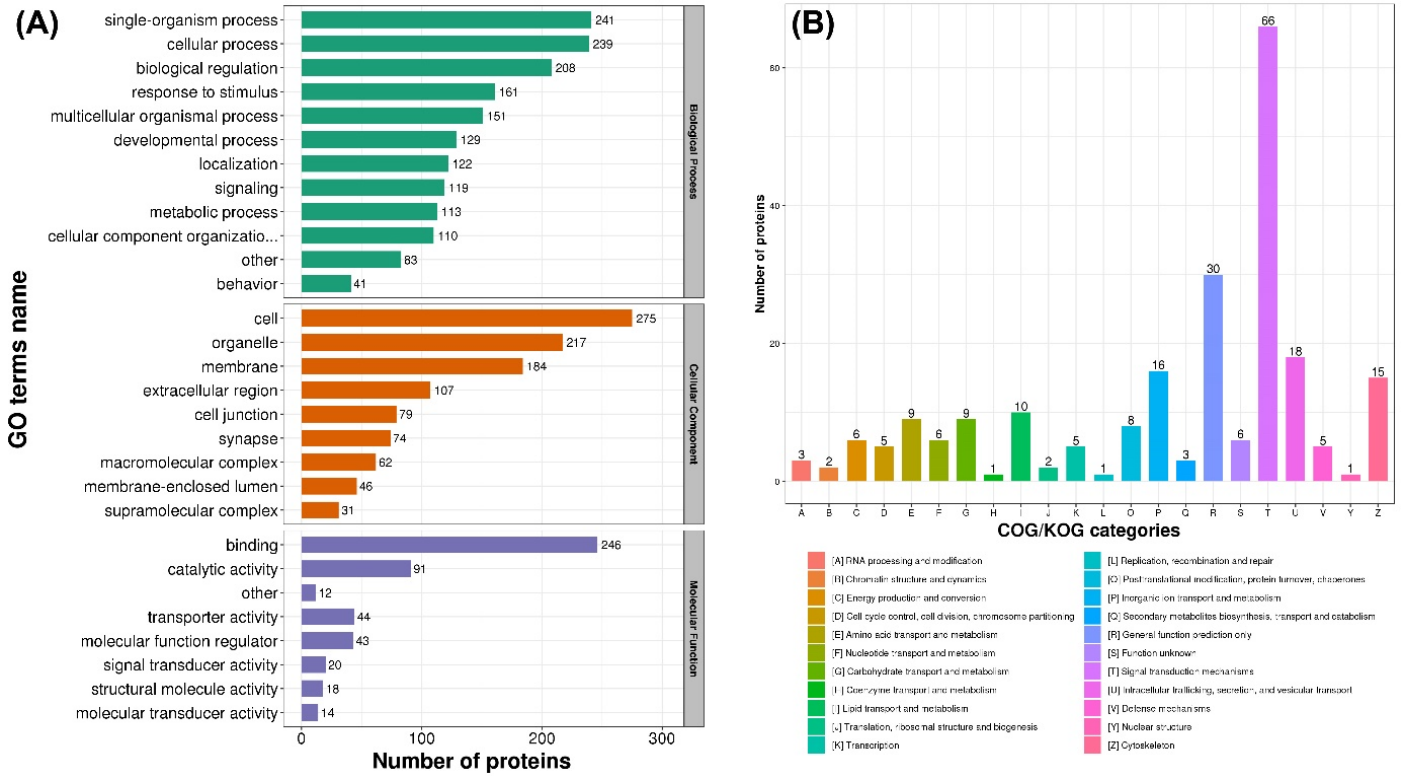


Figure S6. (A) Among these 282 significantly changed proteins, the functions and characteristics of all the detected and quantified proteins were further investigated by gene ontology (GO) analysis and classification of the proteins into three categories: biological process, cellular component, and molecular function. (B) Analysis of eukaryotic orthologous groups (KOG) and clusters of orthologous groups (COG) of proteins was performed. Among the significantly changed proteins, 66 proteins were involved in signal transduction mechanisms, 30 proteins were related to general function prediction only, 18 proteins were involved in intracellular trafficking, secretion, and vesicular transport, and 16 proteins were related to inorganic ion transport and metabolism.

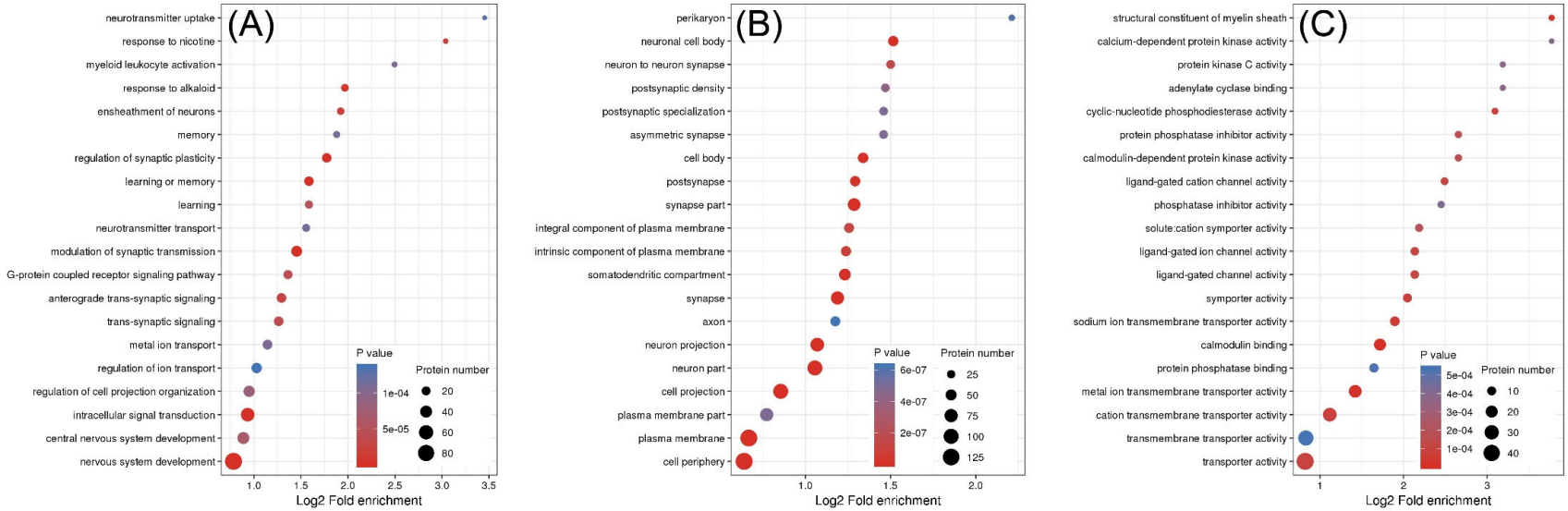


Figure S7. Enrichment results of biological process (A), cellular component (B), and molecular function (C).

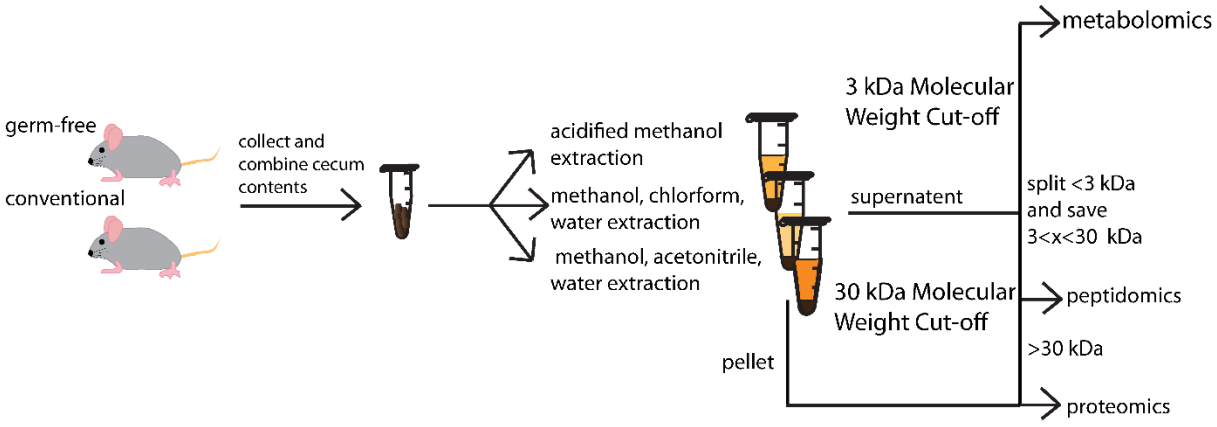
Table S1. Representative neuropeptides and preprohormone-derived peptides identified from GF- and Con- mice hypothalamus and their corresponding changes. (−0.98) denotes C-terminal amidation, (+15.99) denotes oxidation, and (+0.98) denotes deamidation. The GF/Con ratio represents the relative level changes of peptides under different gut environments.

Accession	Precursor	Gene	Name	Peptide sequence	PTM	m/z	z	Mass	ppm	GF/CON Ratio	*P* value
O55241	Orexin	*Hcrt*	Orexin-B	R.RPGPPGLQGRLQRLLQANGNHAAGILTM.G		588.1335	5	2935.6038	9.4	2.19	0.0128
P01193	Pro-opiomelanocortin	*Pomc*	Joining peptide	R.AEEEAVWGDGSPEPSPRE(-.98).G	C-term amidation	970.9355	2	1939.8547	0.9	7.21	0.0480
P01193	Pro-opiomelanocortin	*Pomc*	γ-Lipotropin	R.ELEGERPLGLEQVLESDAEKDDGPYRVEHF.R		864.9205	4	3455.6479	1.4	5.05	0.0463
P01193	Pro-opiomelanocortin	*Pomc*	CLIP	R.RPVKVYPNVAENESAEAFPLEF.K		1253.6394	2	2505.2539	4.1	5.89	0.0425
P01193	Pro-opiomelanocortin	*Pomc*	α-MSH	R.SYSMEHFRWGKPV(-.98).G	C-term amidation	811.9008	2	1621.7823	2.9	3.97	0.0497
P01193	Pro-opiomelanocortin	*Pomc*	β-endorphin	R.YGGFMTSEKSQTPLVTLFKNAIIKNAHKKGQ		688.1751	5	3435.8335	1.7	4.35	0.0431
P41539	Protachykinin-1	*Tac1*	Neurokinin A	R.HKTDSFVGLM(+15.99)(-.98).G	Oxidation C-term amidation	575.2912	2	1148.5648	2.6	2.42	0.0089

P41539	Protachykinin-1	*Tac1*	Substance P	R.RPKPQQFFGLM(+15.99)(-.98).G	C-term amidation	682.3691	2	1362.7230	0.5	1.96	0.0212
P47212	Galanin peptides	*Gal*	Galanin	R.GWTLNSAGYLLGPHAIDN(+.98)HRSFSDKHGLT(-.98).G	Deamidation C-term amidation	633.7219	5	3163.5588	4.5	2.06	0.0238
P55099	Tachykinin-3	*Tac3*	Neurokinin B	R.DM(+15.99)HDFFVGLM(-.98).G	Oxidation C-term amidation	613.7717	2	1225.5260	2.3	1.78	0.0303
Q64387	Prepronociceptin	*Pnoc*	Nociceptin	R.FGGFTGARKSARKLANQ.K		603.6692	3	1807.9805	2.9	1.86	0.0004
P60041	Somatostatin	*Sst*	Somatostatin-28 [1-12]	R.SANSNPAMAPRE.R		622.7897	2	1243.5615	2.6	1.58	0.0008
Q9QXV0	ProSAAS	*Pcsk1n*	Big SAAS	S.ARPVKEPRSLSAASAPLVETSTPLRL.R		687.3975	4	2745.5500	3.9	2.31	0.0213
Q9QXV0	ProSAAS	*Pcsk1n*	Big LEN	R.LENPSPQAPARRLLPP		586.3352	3	1755.9631	11.8	2.28	0.0026
Q9QXV0	ProSAAS	*Pcsk1n*	Little LEN	R.LENPSPQAPA.R		512.2601	2	1022.5032	2.4	3.02	0.0002
Q9QXV0	ProSAAS	*Pcsk1n*	Little SAAS	R.SLSAASAPLVETSTPLRL.R		907.0076	2	1811.9993	0.7	1.74	0.0021
Q9R171	Cerebellin-1	*Cbln1*	Des-Ser1-Cerebellin	S.GSAKVAFSAIRSTNH.E		515.9432	3	1544.8059	1.3	2.10	0.0471
Q9R171	Cerebellin-1	*Cbln1*	Cerebellin	R.SGSAKVAFSAIRSTNH.E		544.9545	3	1631.8379	2.4	2.83	0.0191

Appendix VI

Extraction Optimization for Combined Metabolomics, Peptidomics, and Proteomics Analysis of Gut Microbiota Samples



Adapted Caitlin Keller, **Pingli Wei**, Benjamin Wancewicz, Tzu-Wen L Cross, Frederico Rey, Lingjun Li. "Extraction Optimization for Combined Metabolomics, Peptidomics, and Proteomics Analysis of Cecum Gut Microbiota Samples." *To be submitted.* **Pingli Wei** prepare sample and datanalysis for the peptide and protein study.

Abstract

Multiomic studies are increasingly used to gain a deeper understanding of molecular processes occurring in a biological system, such as the complex microbial communities (i.e., microbiota) that reside the distal gut. While a combining metabolomics and proteomics is more commonly used, multiomics studies including peptidomcis characterization are less frequently undertaken. Here, we investigated three different extraction methods, chosen for their previous use in extracting metabolites, peptides, and proteins, and compared their ability to perform metabolomic, peptidomic, and proteomic analysis of mouse cecum content. The methanol/chloroform/water extraction performed the best for metabolomic and peptidomic analysis as it detected the largest number of small molecules and identified the largest number of peptides, but the acidified methanol extraction performed best for proteomics analysis as it had the highest number of protein identifications. The methanol/chloroform/water extraction was further analyzed by identifying metabolites with MS/MS analysis and by gene ontology analysis for the peptide and protein results to provide a multiomics analysis of the gut microbiota.

Keywords: Microbiome, Multiomics, Metabolomics, Peptidomics, Proteomics

Introduction

The gastrointestinal tract is host to large and dynamic communities of microbes, containing ~10^8 to 10^{10} organisms per gram in the illium and stool [1] , and encompassing the 3 domains of life: bacterial, archaea and eukaryotes [2,3]. The gut microbiota has various roles in human health, including nutrition and immune function modulation [4,5]. Furthermore, the disruption of the gut microbiota has been connected to various inflammatory chronic conditions such as inflammatory

bowel disease, obesity [6,7] and cardiovascular disease [8,9]. Due to the importance that the microbiota plays in human health, there is an abundance of studies on various aspects of the microbiota. Research includes various aspects of the microbiome such as the metagenomic [10] and proteomic content [11], but also looks at how the microbiome influences other systems, including blood and cerebrospinal fluid metabolites [12,13].

While studying individual "omics" results in a depth of information about the genomic, proteomic or metabolomic content of the desired biological specimen, the combination of multiple of these analyses can determine different connections between molecular classes in a biological system. Consequently, multiomic approaches are used to study the microbiota, as well as in other fields [14-17]. A combination of metagenomic and metabolomic approaches has been used, for example, to study the microbiota's response to infection [18], the effect of arsenic on the gut microbiome [19], and the effect of trichloroacetamide exposure on the gut microbiome and urinary metabolites [20]. Other combinations of metagenomics, metatranscriptomics, metabolomics, and (meta)proteomics can also be employed [21]. Although large amounts of data created in multiomic techniques can be challenging and time consuming to analyze, advances in software are arising to simplify the task [22,23].

While the field of peptidomics is more recent than the fields of metabolomics and proteomics, it has nevertheless risen to importance [24,25]. Well-studied signaling peptides in the brain, i.e. neuropeptides, are an important class of molecules regulating a wide variety of processes [26,27]. Similarly, peptide hormones in the endocrine system are also important endogenous peptides that have been studied due to their role in regulating metabolism [28,29]. In the gut, bioactive peptides, derived from digestion of proteins in the intestines, has resulted in a number of peptides with roles in health [30]. For example, bioactive peptides present in the gut that inhibit

the angiotensin I-converting enzyme (ACE) can reduce hypertension and improve cardiovascular health [31,32].

Although multiomic studies with a combination of genomics, metabolomics, and proteomics is common, peptidomics is much less frequently combined with other "omic" techniques. Therefore, we set out to identify an optimal extraction method ideal for combining peptidomics with metabolomics and proteomics for a multiomics approach to analyze the gut microbiota. Here, we tested three different extraction methods for combined metabolomics, peptidomics, and proteomics. The three extractions chosen have been previously used in various biological systems to achieve good results in one or more of the metabolomics, peptidomics, and proteomics fields and comprise of a variety of solvent systems. The chloroform/methanol/water extraction is a common small molecule extraction that can be used for metabolomics and proteomics multiomics studies [33], and has previously been used to study cecal metabolomics [18]. An acidified methanol extraction is commonly used in neuropeptide extractions [34,28] but has also been applied for metabolomics [35]. Finally, the 40/40/20 acetonitrile/methanol/water extraction adds acetonitrile to the extraction solvent, which has been shown to be beneficial for certain metabolites [36] and has been shown to work well in metabolomics studies [37]. The small molecules, peptides, and proteins detected for each extraction were compared and used to determine an optimal extraction for metabolomic, peptidomic, and proteomic analysis. While the acidified methanol extraction performed best for the proteomics experiments, the methanol/chloroform/water extraction performed the best in terms of the number of small molecules and peptides detected and was further evaluated with gene ontology analysis.

Materials and Methods

Cecum Collection

All animal procedures were approved by the Institutional Animal Care and Use Committee at the University of Wisconsin-Madison. Germ-free (n=1) and conventionally-raised (n=1) male C57BL/6 mice were used and fed LabDiet #5021 (Purina Mills, Inc. Richmond, IN). At 21 weeks of age, mice were euthanized via CO_2 inhalation and exsanguination. Cecal content was collected, snap frozen in liquid nitrogen immediately, and stored in the -80°C until analysis.

Sample Preparation

The two cecum samples were combined and separated into three approximately equal aliquots. On one aliquot, a methanol/chloroform/water extraction was performed in a PTFE tube by adding in order, 3 parts methanol, 1 part chloroform, and 4 parts MilliQ water (total volute 4.0 mL). The tube was vortexed and centrifuged at 3200 x g and 4^{o}C for 15 minutes. The upper aqueous layer was removed, and 4 parts methanol were added to the tube and vortexed. The tube was centrifuged again at 1500 x g and 4^{o}C for 5 minutes. The supernatant was removed (organic fraction). The aqueous fraction, organic fraction, and pellet were dried down and saved in the -80^{o}C until further processing. The second aliquot was extracted with acidified methanol (methanol/water/acetic acid 90/9/1 v/v/v) and the third aliquot with 40% methanol 40% acetonitrile 20% water. Both of these two extractions were probe sonicated for 3 cycles (8s on 15 s off) at 4^{o}C and centrifuged at 15,000 x g and 4^{o}C for 15 minutes. The supernatant and pellet were separated and then dried down in a speed vac and saved in the -80^{o}C until further processing.

A 3 kDa molecular weight cut-off (MWCO) filtration was performed on the aqueous and organic fraction of the methanol/chloroform/water extraction, the acidified methanol (AcMeOH) extraction, and the methanol/acetonitrile/water extraction. The Amicon Ultra (Millipore) MWCO device was first rinsed with 0.2 mL 0.1 M NaOH, and 0.5 mL 50% methanol. Both rinses were

centrifuged at 14,000 x g until the rinse was through the membrane. The sample was then loaded into the device and centrifuged through the device at 14,000 x g. A final rinse of 0.1 mL 50% methanol was added to the device and the MWCO was centrifuged at 14,000 x g. The content below 3 kDa was split into two aliquots (one metabolomics and one peptidomics) for each of the four samples and dried down in a speed vac. The MWCO was rinsed with 0.4 mL 50% methanol and equilibrated for 5 mins, and then flipped over and centrifuged at 14,000 x g for 2 mins to collect content above 3 kDa from the device. A 30 kDa MWCO was performed on the above three kDa fraction to separate the extract into peptidomics (3 kDa to 30 kDa) and proteomics (above 30 kDa) fractions.

The below 3 kDa and 3-30 kDa peptidomics fractions were combined. Sep-Pak C18 was used for peptide desalting, and then peptide samples were dried down in a speed vac and saved at -80^{o}C until LC-MS/MS analysis. The proteomics fractions from the supernatant contents above 30 kDa and from the pellets were combined. The aqueous and organic fraction above 30 kDa from the methanol/chloroform/water extraction method were combined together with the pellet. The protein mixture samples were dissolved in 1 mL ice-cold PBS, and debris from the pellet was removed with a low centrifuge speed (300 x g, 4^{o}C for 5 mins) [38]. The supernatant was carefully collected, and the pellets were washed another two times with PBS and all the supernatant obtained from each time was combined. The supernatant was then centrifuged at 20,000 x g for 10 min to pellet the bacterial cells and host cells. The pellet was then lysed with 8 M urea Lysis buffer with sonication (On 8 sec, Off 15 sec, 3 cycles). The total protein concentration of each pellet was determined by BCA assay and then digested with Trypsin/Lys-C mixture overnight. Then, the digested proteins were desalted, dried down in a speed vac, and saved in the -80^{o}C until LC-MS/MS analysis.

Metabolomics Data Acquisition and Analysis

The aqueous fraction was resuspended at 10 mg/mL in optima grade water with 0.1% formic acid while the other three metabolomic samples (organic fraction, AcMeOH sample, MeOH/AcN/water sample) were resuspended at 10 mg/mL in optima grade methanol with 0.1% formic acid. Any samples that were cloudy were centrifuged briefly and the supernatant used for the analysis. LC-MS/MS analysis was performed with a Dionex Ultimate 3000 UHPLC system connected to a Q Exactive mass spectrometer (Thermo Scientific). Separation occurred on a Cortecs C18 column (2.1 mm internal diameter x 100 mm length, 1.6 μm particle size; Waters), equipped with a corresponding guard column with a column temperature of 35°C, and mobile phases of optima grade water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B). A 35 minute gradient at a flow rate of 0.3 mL/minutes with the following conditions was used for separation: 0–5 min, 1% B; 5–10 min, linear gradient from 1–3% B; 10–18 min, linear gradient from 3–40% B; 18–22 min, linear gradient from 40–80% B; 22–27 min, column cleaning at 95% B; and 27–35 min, re-equilibration at 1% B. A top 5 data dependent acquisition method was used for MS/MS of small molecules in the extractions. The full MS settings were 70,000 resolution, 1e6 AGC, 100 ms max inject time, 100-1500 *m/z*. The MS/MS settings were 35,000 resolution, 1e5 AGC, 100 ms max inject time, 1.0 *m/z* isolation window, and 30 dynamic exclusion. Three technical replicates were run for each extraction, and each technical replicate used a different HCD collision energy (25, 30, 40 respectively). Samples were run in both positive and negative mode.

Compound Discoverer software was used to analyze the LC-MS/MS data for each extraction in both positive and negative ion mode (the aqueous and organic fractions were analyzed together). Individual runs were aligned with an adaptive curve model with a maximum shift of 1 minute and 5 ppm tolerance. Unknown compounds were detected with a 5 ppm mass tolerance,

30% intensity tolerance, 3 S/N ratio, and 1,000,000 minimum peak intensity. Unknown compounds were grouped with a 5 ppm mass tolerance and 0.1-minute retention time tolerance. A fill gaps step was used with 5 ppm mass error and 0.1 retention time error. Constant sum normalization and marking of background compounds were used. The MS/MS spectra were searched in the mzCloud library against all activation types and activation energies and matches were manually validated by ensuring that all the major fragment ions in the database spectra matched the experimental spectra. MS/MS spectra were also searched against the MassBank of North America MS/MS database for additional identifications.

Peptidomics Acquisition and Analysis

Peptide samples were resuspended in optima grade water with 3% acetonitrile and 0.1% formic acid. LC-MS/MS was performed on an Ultimate 3000 UPLC system coupled with the Orbitrap Fusion™ Tribrid™ Mass Spectrometer. A 75 μm × 16 cm homemade column packed with 1.7 μm, 150 Å, BEH C18 material obtained from a Waters (Milford, MA) UPLC column (part no. 186004661) was used for label-free peptide separation at a flow rate of 0.3 μl/min. Mobile phase A was 0.1% formic acid in optima water and mobile phase B was 0.1% formic acid in optima acetonitrile. The 145 min optimized gradient used was as follows: 0-18.33 min, 3% solvent B; 18.33-30 min, 3-10% B; 30-50 min, 10-20% B; 50-108 min, 20-75% B; 108-118 min, 75% B; 118-118.5 min 75%-95% B; 118.5-128 min, 95% B; 128-128.5 min, 95%-3% B; 128.5-145 min, 3% B. Full MS scans were acquired from *m/z* 300 to 1500 at a resolution of 60 K, automatic gain control (AGC) at 2×10^5, and maximum injection time (IT) of 100 ms. The top 20 precursors were then selected for higher-energy C-trap dissociation tandem mass spectrometry (HCD MS2) analysis with an isolation window of 1 m/z, a HCD collision energy (NCE) of 30, a resolving

power of 15 k, an AGC target of 5×10^4, a maximum injection time of 100 ms, and a lower mass limit of 120 m/z.

The .raw data files from the Orbitrap MS analysis were searched against a combined database which included food, 93 strains of bacteria, and mouse proteome from Uniport with PEAKS STUDIO 8.5 software. A precursor tolerance of 10 ppm and a fragment mass tolerance of 0.02 Da were allowed. Acetylation (N-term), amidation, oxidation (M), pyro-Glu from E, pyro-Glu from Q, sulfation (STY), were set as rare dynamic modifications and allowing three maximum variable PTM per peptide. Parameters for confident peptide identification were Ascore (PTM site confidence) higher than 20, FDR lower than 1%, and the presence of at least one unique peptide.

Proteomics Acquisition and Analysis

A Dionex UltiMate 3000 nanoLC system coupled with a Q Exactive HF Orbitrap MS was used for Ultra-performance LC-MS analysis. Homemade column and mobile phases were the same as mentioned above for the peptidomic analysis. The optimized gradient used was as follows: 0-16 min, 3% solvent B; 16-20 min, 3-25% B; 20-30 min, 25-45% B; 30-50 min, 45-70% B; 50-56 min, 70-95% B; 56-60 min 95% B; 60-60.5 min, 95-3% B; 60.5-70 min, 3% B. Full MS scans were acquired from *m/z* 300 to 1500 at a resolution of 60 K, AGC at 1×10^6, and maximum injection time (IT) of 100 ms. The top 15 precursors were then selected for higher-energy C-trap dissociation tandem mass spectrometry (HCD MS2) analysis with an isolation window of 1.4 *m/z*, a normalized collision energy (NCE) of 30, a resolving power of 15 K, an AGC target of 1×10^5, a maximum injection time of 100 ms, and a lower mass limit of 120 *m/z*. PEAKS software was used for protein identification. The parameters were the same as used above for peptidomics analysis, except trypsin with D&P enzyme was selected for this bottom-up proteomics study. Non-specific

cleavage at both ends of the peptide was allowed and the maximum missed cleavages per peptide was set at two.

Results

Two cecal content samples were combined and split into three approximately equal aliquots to test three different extraction protocols. The sample preparation workflow is provided in Online Resource 1 for the three extractions. The three extractions tested were a methanol/chloroform/water (MeOH/$CHCl_3$/H_2O) extraction, an acidified methanol (AcMeOH) extraction, and a methanol, acetonitrile, water (MeOH/AcN/H_2O) extraction. The methanol/chloroform/water extraction results in two liquid fractions, an aqueous fraction and an organic fraction, as well as a pellet. The other two extractions resulted in a supernatant and a pellet. The four liquid portions were processed with a 3 kDa molecular weight cut-off followed by a 30 kDa molecular weight cut-off. The <3 kDa portion was split for LC-MS metabolomics analysis and nano-LC-MS peptidomics analysis. The 3<x<30 kDa portion was saved for nano-LC-MS peptidomics analysis as well. The >30 kDa fraction was combined with the pellet for bottom up nano-LC-MS proteomics analysis.

Metabolomics

Metabolomics data was analyzed in Compound Discoverer 2.0 to detect unknown compounds and perform tandem MS (MS/MS) matching of experimental MS/MS spectra to the mzCloud high resolution/accurate mass spectral database. Each of the three extractions were analyzed separately in the software with the same parameters to test how many compounds were detected in each extraction. The aqueous and organic fractions for the methanol/chloroform/water extraction were combined in the software. **Fig. 1** shows the results for the metabolomics analysis.

In both positive and negative mode, the methanol/chloroform/water extraction detected about 2-times more compounds than the other extractions in Compound Discoverer. Compound monoisotopic molecular weights were processed in METLIN with a 5 ppm error to approximate how many of the *m/z* potentially matched to known small molecules. The methanol/chloroform/water extraction still out-performed the other two extractions in both positive and negative mode in the METLIN analysis. In order to compare the detected *m/z* across the three extractions, the METLIN searching results were used with the venny software [39] to create Venn diagrams. To be consistent, the lowest METLIN identification number was used for the *m/z* that had multiple accurate mass matches to the METLIN database. In the Venn diagrams (**Fig. 1b**, positive data, **Fig. 1d**, negative data), a majority of the compounds are detected with the methanol, chloroform water extraction, with only a small percentage unique to one of the other two extractions. Thus, the methanol/chloroform/water extraction performs the best for metabolomics analysis.

Peptidomics

Peptidomic analysis was performed on a nano-LC QE-HF system and the data analyzed in PEAKS 8.5 software. **Fig. 2** shows the results of the peptidomics experiments by showing the peptide sequences detected and by comparing the protein accession numbers detected. In **Fig. 2a**, the number of peptide sequences is shown with the number of peptide sequences shared in both technical replicates represented with the diagonal lines. The organic fraction of the methanol/chloroform/water extraction had the most peptides identified, with the aqueous fraction in second. The other two extractions showed low peptide identifications. When comparing the proteins detected from the peptide sequences that were shared between both technical replicates for the organic and aqueous fractions, most of the aqueous proteins are also in the organic fraction,

but there are proteins only in the aqueous fraction. To compare the three extractions, the combination of unique proteins from the aqueous and organic fraction were taken, along with the proteins in both technical replicates for the other two extractions. The Venn diagram shows only 2 proteins that were not detected by the methanol/chloroform/water extraction and over 100 proteins that were only detected in the methanol/chloroform/water extraction. Thus, the methanol/chloroform/water extraction performs best for the peptidomics analysis as it identifies the most proteins and covers almost all of the proteins identified in the other two extractions.

To look further at the differences between the peptides detected in the aqueous and organic fractions, peptide sequences found in each fraction from the proteins present in both technical replicates were compared with respect to their length and isoelectric point. The isoelectric point was calculated using the Peptide Property calculator (GeneScript, online tool) as it allowed the inclusion of certain posttranslational modifications (PTMs), for example acetylation. PTMs that were not present in this online tool were excluded from the isoelectric point analysis, but this was minority for each fraction (11/198 for organic fraction 2/89 for aqueous fraction). **Fig. 3** shows the results of the length and isoelectric point comparison. The length distribution in the aqueous and organic fractions seemed to be similar as the organic fraction had approximately 2-fold more peptides sequences at each length, and overall the organic fraction has about 2 times more sequences. The isoelectric point distribution, however, does appear to be different between the two fractions. While the number of peptides with very low pI's is approximately the same for both fractions, the organic fraction has more peptide sequences with pI's above 4. The organic fraction potentially extracted a larger number of peptides than the aqueous fraction due to the preference of the peptides for the methanol/chloroform solvent.

Proteomics

Bottom-up proteomics was performed on the pellets and the >30 kDa content from the molecular weight cut-off step. The results of this analysis are shown in **Fig. 4**. The acidified methanol extraction had noticeably more proteins than the other two extractions. While only proteins detected in both technical replicates were counted as identifications, it is worth noting that many proteins were only in 1 technical replicate, rather than both technical replicates (as shown by the diagonal lines). Potentially adding in a third technical replicate in the experimental analysis could increase the number of identifications by increasing the identifications seen in multiple technical replicates. The comparison of the three extractions shows that each extraction has hundreds of identifications unique to that extraction, with the acidified methanol extraction having over a thousand identifications unique to it. For searching, a combined multi-organism database was used containing the mouse genome, the genomes of approximately 90 bacterial strains for a model microbiota [18], and potential proteins from the food the mice were fed. While conventionally raised mice very likely have a more diverse microbial community [40], here a more focused database was used because increasing the database size to a more comprehensive one would make processing the MS/MS data significantly more challenging. For studies highly interested in the microbial species present, a targeted microbiome database without the mouse and food components could be used instead. **Fig 4c** shows the number of proteins that match to each of the different potential sources of the protein, namely the mouse, the microbiome, and the food. The methanol/acetonitrile/water extraction has the most proteins matching to both the mouse database and the food proteins. Where the acidified methanol extraction gets a vast majority of its protein identifications from is from microbiome proteins, as it has a much higher number of microbiome identifications compared to the other two extractions.

Combined Metabolomics, Peptidomics, Proteomics

Fig. 5 combines the results of the three different omics by total identifications and by a source comparison for the combined peptide and protein results. The total identifications were calculated with the total number of hits to the METLIN database from the positive mode LC-MS data for the metabolomics analysis, the identified proteins from the peptidomics study shared between both technical replicates and, the protein number from the proteomics study shared between both technical replicates. When all three omics are combined, the methanol/chloroform/water and the acidified methanol extractions have similar number of total identifications. When the peptide and protein results are combined, and the source of the protein identification investigated, the methanol/chloroform/water, and methanol/acetonitrile/water extractions have similar numbers of mouse proteins. However, the acidified methanol extraction still has the most identifications due to the large number of microbial protein identifications. For microbiome studies focusing on the host response, the methanol/chloroform/water performs well for metabolomics, peptidomics, and proteomics, but if microbial proteins are desired, the acidified methanol extraction preforms best.

Discussion

In order to get a good representation of metabolites, peptides, and proteins, the methanol/chloroform/water extraction was chosen for further investigation into the compounds detected. For metabolomics identification, the mzCloud and MassBank small molecule MS/MS databases were used for spectral matching of fragment ions. Potential matches to either database were manually inspected for verification. **Online Resource 2** compares the total small molecule putative identifications in the methanol/chloroform/water extraction with the two MS/MS databases with both databases in positive and negative modes, as well as comparing the overlap

between the positive and negative putative identifications. **Online Resource 3** provides all the small molecule identifications for the methanol/chloroform/water extraction with their molecular weight, retention time, and which database and polarity they were identified from. In positive mode, both databases provided a similar number of identifications and resulted in complementary coverage for a total of 57 identifications. In negative mode, almost all the results came from the MSDial database. Overall, the positive and negative mode identification results showed complementary coverage as many identifications were made only in one of the polarities. To verified putatively identified compounds, standards could be obtained, and retention times and fragmentation patterns compared to the experimental data. More identifications are potentially possible using other databases (i.e., METLIN) or using *in silico* fragmentation software. However, these were not utilized here due to the time involved in metabolomics identification and the lack of a biological experiment that would show upregulation or downregulation of certain *m/z* in a biological condition.

The identifications for the peptidomics results for the methanol/chloroform/water extraction are in **Online Resource 4** (aqueous fraction) and **Online Resource 5** (organic fraction), and the proteomics results for the methanol/chloroform/water extraction are provided in **Online Resource 6**. Gene Ontology analysis was conducted on the protein and peptide results from the methanol/chloroform/water extraction using DAVID Bioinformatics Resources 6.8 [41,42]. The gene ontology results of the biological processes for the detected peptides and proteins are shown in **Fig. 6**. The top biological process for both the protein and peptide results was proteolysis. Other shared biological process include digestion, chromatin silencing, regulation of systemic arterial blood pressure by renin-angiotensin, and metabolic process. Overall, the many shared biological processes in the protein and peptide gene ontology results indicate that the protein and peptide

results agree well with each other. By integrating the peptide results with the protein results, a more comprehensive understanding of the biological processes can be achieved for a biological question of interest.

To look further at the peptidomcis results, the peptides belonging to the biological process of digestion, a key function of the digestive track that the cecum is a part of, were investigated. **Table 1** shows the peptides that fell under the biological process of digestion. Peptide sequences from the enzymes chymotrypsinogen B1(Ctrb1), serine protease, and trypsin were detected. Further analysis would be necessary to discover the potential role that these peptides play. Furthermore, an *in silico* study reported the potential for endogenous proteins in the gut to be digested into bioactive peptides [43]. Experimentally, peptides were observed from three proteins, Mucin-13, Chymotrypsinogen B, and pancreatic triacylglycerol lipase, that have predicted potential for bioactive peptide release after intestinal digestion of the gut endogenous protein. The predicted activity for peptides from these three proteins is ACE inhibition, which can prevent hypertension [32].

Multiomic analysis can provide a greater understanding by studying not just the one subclass of molecules, protein changes, for example, but also changes in the metagenomic, or metabolome in order to better understand the relationships between various biological systems. Here, the ability of three extractions for combined metabolomics, peptidomics, and proteomics analysis was compared. The methanol/chloroform/water extraction method enabled a more comprehensive view of all three omics in the mouse host system and performed particularly strong in metabolomics and peptidomics analysis. By including peptidomics in the multiomics experiments, a deeper understanding of the role of peptides could be obtained, for example, by characterization and discovery of bioactive peptides and their role in various pathways.

Acknowledgements

This study was supported in part by grant funding from the NIH (R01DK071801, RF1AG052324, to LL) and a grant from a Transatlantic Networks of Excellence Award from the Leducq Foundation (to FER). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531 to LL) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. T.-W. L. C. is supported by the National Institutes of Health, under Ruth L. Kirschstein National Research Service Award T32 HL 007936 from the National Heart Lung and Blood Institute to the University of Wisconsin-Madison Cardiovascular Research Center. LL acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Conflict of Interest

The authors have no conflict of interests to declare.

References

1. Sender R, Fuchs S, Milo R (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol 14 (8):e1002533. doi:10.1371/journal.pbio.1002533
2. Scanlan PD, Marchesi JR (2008) Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. Isme j 2 (12):1183-1193. doi:10.1038/ismej.2008.76
3. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, Bushman FD (2013) Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One 8 (6):e66019. doi:10.1371/journal.pone.0066019
4. Sharon G, Garg N, Debelius J, Knight R, Dorrestein PC, Mazmanian SK (2014) Specialized metabolites from the microbiome in health and disease. Cell Metab 20 (5):719-730. doi:10.1016/j.cmet.2014.10.016
5. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI (2011) Human nutrition, the gut microbiome and the immune system. Nature 474 (7351):327-336. doi:10.1038/nature10213
6. Turnbaugh PJ, Backhed F, Fulton L, Gordon JI (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe 3 (4):213-223. doi:10.1016/j.chom.2008.02.015
7. Rosenbaum M, Knight R, Leibel RL (2015) The gut microbiota in human energy homeostasis and obesity. Trends Endocrinol Metab 26 (9):493-501. doi:10.1016/j.tem.2015.07.002
8. Jie Z, Xia H, Zhong SL, Feng Q, Li S, Liang S, Zhong H, Liu Z, Gao Y, Zhao H, Zhang D, Su Z, Fang Z, Lan Z, Li J, Xiao L, Li R, Li X, Li F, Ren H, Huang Y, Peng Y, Li G, Wen B, Dong B, Chen JY, Geng QS, Zhang ZW, Yang H, Wang J, Zhang X, Madsen L, Brix S, Ning G, Xu X, Liu X, Hou Y, Jia H, He K, Kristiansen K (2017) The gut microbiome in atherosclerotic cardiovascular disease. Nat Commun 8 (1):845. doi:10.1038/s41467-017-00900-1
9. Tang WH, Kitai T, Hazen SL (2017) Gut Microbiota in Cardiovascular Health and Disease. Circ Res 120 (7):1183-1196. doi:10.1161/circresaha.117.309715
10. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI (2009) A core gut microbiome in obese and lean twins. Nature 457 (7228):480-U487. doi:10.1038/nature07540
11. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL, Jansson JK (2009) Shotgun metaproteomics of the human distal gut microbiota. Isme Journal 3 (2):179-189. doi:10.1038/ismej.2008.108
12. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G (2009) Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. Proceedings of the National Academy of Sciences of the United States of America 106 (10):3698-3703. doi:10.1073/pnas.0812874106
13. Vogt NM, Romano KA, Darst BF, Engelman CD, Johnson SC, Carlsson CM, Asthana S, Blennow K, Zetterberg H, Bendlin BB, Rey FE (2018) The gut microbiota-derived metabolite trimethylamine N-oxide is elevated in Alzheimer's disease. Alzheimer's Research & Therapy 10 (1):124. doi:10.1186/s13195-018-0451-2
14. Roume H, Muller EE, Cordes T, Renaut J, Hiller K, Wilmes P (2013) A biomolecular isolation framework for eco-systems biology. Isme j 7 (1):110-121. doi:10.1038/ismej.2012.72
15. Wettersten HI, Hakimi AA, Morin D, Bianchi C, Johnstone M, Donohoe DR, Trott JF, Abu Aboud O, Stirdivant S, Neri B, Wolfert R, Stewart B, Perego R, Hsieh JJ, Weiss RH (2015) Grade-Dependent Metabolic Reprogramming in Kidney Cancer Revealed by Combined Proteomics and

Metabolomics Analysis. Cancer Research 75 (12):2541-2552. doi:10.1158/0008-5472.can-14-1703
16. Weckwerth W, Wenzel K, Fiehn O (2004) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. Proteomics 4 (1):78-83. doi:10.1002/pmic.200200500
17. Mostafa I, Zhu N, Yoo MJ, Balmant KM, Misra BB, Dufresne C, Abou-Hashem M, Chen S, El-Domiaty M (2016) New nodes and edges in the glucosinolate molecular network revealed by proteomics and metabolomics of Arabidopsis myb28/29 and cyp79B2/B3 glucosinolate mutants. J Proteomics 138:1-19. doi:10.1016/j.jprot.2016.02.012
18. Bratburd JR, Keller C, Vivas E, Gemperline E, Li L, Rey FE, Currie CR (2018) Gut Microbial and Metabolic Responses to Salmonella enterica Serovar Typhimurium and Candida albicans. MBio 9 (6). doi:10.1128/mBio.02032-18
19. Lu K, Abo RP, Schlieper KA, Graffam ME, Levine S, Wishnok JS, Swenberg JA, Tannenbaum SR, Fox JG (2014) Arsenic exposure perturbs the gut microbiome and its metabolic profile in mice: an integrated metagenomics and metabolomics analysis. Environ Health Perspect 122 (3):284-291. doi:10.1289/ehp.1307429
20. Zhang Y, Zhao F, Deng Y, Zhao Y, Ren H (2015) Metagenomic and metabolomic analysis of the toxic effects of trichloroacetamide-induced gut microbiome and urine metabolome perturbations in mice. J Proteome Res 14 (4):1752-1761. doi:10.1021/pr5011263
21. Kaiser BLD, Li J, Sanford JA, Kim YM, Kronewitter SR, Jones MB, Peterson CT, Peterson SN, Frank BC, Purvine SO, Brown JN, Metz TO, Smith RD, Heffron F, Adkins JN (2013) A Multi-Omic View of Host-Pathogen-Commensal Interplay in Salmonella-Mediated Intestinal Infection. Plos One 8 (6):13. doi:10.1371/journal.pone.0067155
22. Fan J, Saha S, Barker G, Heesom KJ, Ghali F, Jones AR, Matthews DA, Bessant C (2015) Galaxy Integrated Omics: Web-based Standards-Compliant Workflows for Proteomics Informed by Transcriptomics. Mol Cell Proteomics 14 (11):3087-3093. doi:10.1074/mcp.O115.048777
23. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, Paley S, Subhraveti P, Karp PD (2018) The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res 46 (D1):D633-d639. doi:10.1093/nar/gkx935
24. Clynen E, Baggerman G, Veelaert D, Cerstiaens A, Van der Horst D, Harthoorn L, Derua R, Waelkens E, De Loof A, Schoofs L (2001) Peptidomics of the pars intercerebralis-corpus cardiacum complex of the migratory locust, Locusta migratoria. Eur J Biochem 268 (7):1929-1939
25. Verhaert P, Uttenweiler-Joseph S, de Vries M, Loboda A, Ens W, Standing KG (2001) Matrix-assisted laser desorption/ionization quadrupole time-of-flight mass spectrometry: an elegant tool for peptidomics. Proteomics 1 (1):118-131. doi:10.1002/1615-9861(200101)1:1<118::aid-prot118>3.0.co;2-1
26. An S, Harang R, Meeker K, Granados-Fuentes D, Tsai CA, Mazuski C, Kim J, Doyle FJ, 3rd, Petzold LR, Herzog ED (2013) A neuropeptide speeds circadian entrainment by reducing intercellular synchrony. Proc Natl Acad Sci U S A 110 (46):E4355-4361. doi:10.1073/pnas.1307088110
27. Schank JR, Ryabinin AE, Giardino WJ, Ciccocioppo R, Heilig M (2012) Stress-related neuropeptides and addictive behaviors: beyond the usual suspects. Neuron 76 (1):192-208. doi:10.1016/j.neuron.2012.09.026
28. Yu Q, Canales A, Glover MS, Das R, Shi X, Liu Y, Keller MP, Attie AD, Li L (2017) Targeted Mass Spectrometry Approach Enabled Discovery of O-Glycosylated Insulin and Related Signaling

Peptides in Mouse and Human Pancreatic Islets. Anal Chem 89 (17):9184-9191. doi:10.1021/acs.analchem.7b01926
29. Barreto SG, Carati CJ, Toouli J, Saccone GT (2010) The islet-acinar axis of the pancreas: more than just insulin. Am J Physiol Gastrointest Liver Physiol 299 (1):G10-22. doi:10.1152/ajpgi.00077.2010
30. Moller NP, Scholz-Ahrens KE, Roos N, Schrezenmeir J (2008) Bioactive peptides and proteins from foods: indication for health effects. Eur J Nutr 47 (4):171-182. doi:10.1007/s00394-008-0710-2
31. Yamamoto N, Akino A, Takano T (1994) Antihypertensive effect of the peptides derived from casein by an extracellular proteinase from Lactobacillus helveticus CP790. J Dairy Sci 77 (4):917-922. doi:10.3168/jds.S0022-0302(94)77026-0
32. Dave LA, Hayes M, Montoya CA, Rutherfurd SM, Moughan PJ (2016) Human gut endogenous proteins as a potential source of angiotensin-I-converting enzyme (ACE-I)-, renin inhibitory and antioxidant peptides. Peptides 76:30-44. doi:10.1016/j.peptides.2015.11.003
33. Sapcariu SC, Kanashova T, Weindl D, Ghelfi J, Dittmar G, Hiller K (2014) Simultaneous extraction of proteins and metabolites from cells in culture. MethodsX 1:74-80. doi:10.1016/j.mex.2014.07.002
34. Ye H, Wang J, Tian Z, Ma F, Dowell J, Bremer Q, Lu G, Baldo B, Li L (2017) Quantitative Mass Spectrometry Reveals Food Intake-Induced Neuropeptide Level Changes in Rat Brain: Functional Assessment of Selected Neuropeptides as Feeding Regulators. Mol Cell Proteomics 16 (11):1922-1937. doi:10.1074/mcp.RA117.000057
35. Cao QJW, Ouyang CZ, Zhong XF, Li LJ (2018) Profiling of small molecule metabolites and neurotransmitters in crustacean hemolymph and neuronal tissues using reversed-phase LC-MS/MS. Electrophoresis 39 (9-10):1241-1248. doi:10.1002/elps.201800058
36. Rabinowitz JD, Kimball E (2007) Acidic acetonitrile for cellular metabolome extraction from Escherichia coli. Anal Chem 79 (16):6167-6173. doi:10.1021/ac070470c
37. Bazurto JV, Dearth SP, Tague ED, Campagna SR, Downs DM (2017) Untargeted metabolomics confirms and extends the understanding of the impact of aminoimidazole carboxamide ribotide (AICAR) in the metabolic network of Salmonella enterica. Microb Cell 5 (2):74-87. doi:10.15698/mic2018.02.613
38. Zhang X, Ning Z, Mayne J, Moore JI, Li J, Butcher J, Deeke SA, Chen R, Chiang CK, Wen M, Mack D, Stintzi A, Figeys D (2016) MetaPro-IQ: a universal metaproteomic approach to studying human and mouse gut microbiota. Microbiome 4 (1):31. doi:10.1186/s40168-016-0176-z
39. Oliveros *JC* (*2007-2015*) Venny. An interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html,
40. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, Li X, Long H, Zhang J, Zhang D, Liu C, Fang Z, Chou J, Glanville J, Hao Q, Kotowska D, Colding C, Licht TR, Wu D, Yu J, Sung JJ, Liang Q, Li J, Jia H, Lan Z, Tremaroli V, Dworzynski P, Nielsen HB, Backhed F, Dore J, Le Chatelier E, Ehrlich SD, Lin JC, Arumugam M, Wang J, Madsen L, Kristiansen K (2015) A catalog of the mouse gut metagenome. Nat Biotechnol 33 (10):1103-1108. doi:10.1038/nbt.3353
41. Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37 (1):1-13. doi:10.1093/nar/gkn923

42. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4 (1):44-57. doi:10.1038/nprot.2008.211
43. Dave LA, Montoya CA, Rutherfurd SM, Moughan PJ (2014) Gastrointestinal endogenous proteins as a source of bioactive peptides--an in silico study. PLoS One 9 (6):e98922. doi:10.1371/journal.pone.0098922

Tables

Table 1. Peptide Sequences from the Digestion biological process in the peptide gene ontology results.

Identification Name	Uniptrot ID	Sequence	Mass (Da)
chymotrypsinogen B1(Ctrb1)	Q9CR35	A.GEFDQGSDEENVQVLK.I	1792.8115
		K.IAQVFKNPK.F	1043.6127
protease, serine 2(Prss2)	P07146	I.NVLEGNEQFVDSAK.I	1548.7419
trypsin 4(Try4)/ trypsin 5(Try5)	Q9R0T7/ Q9QUK9	R.TLNNDIM(+15.99)LIK.L	1189.6377

Figure Captions

Fig. 1 Metabolomics results for the three extractions. The number of compounds detected is shown for positive **(A)** and negative **(C)** mode. The m/z with hits in METLIN were compared between the three extractions in Venn Diagrams (**B**, positive and **D**, negative mode)

Fig. 2 Peptidomics results from the extraction test. In **(A)** the number of peptide sequences are shown. **(B)** compares the protein accession numbers in the aqueous and organic fractions of the methanol, chloroform, water extraction. **(C)** compares the protein accession numbers for the three extractions.

Fig. 3 Comparison of peptide length **(A)** and isoelectric point **(B)** between peptide sequences detected in organic and aqueous fractions.

Fig. 4 Proteomics results from extraction optimization. **(A)** shows the detected proteins with the diagonal lines representing the number of shared proteins in both technical replicates. **(B)** compares the shared proteins in both technical replicates for the three extractions. **(C)** shows the breakdown of the shared proteins to the source of the protein ID for each extraction

Fig. 5 Combined results of the three omics analysis. **(A)** shows the combined identifications of metabolomics, peptidomics, and proteomics results. **(B)** compares the protein source for the combined peptide and protein results.

Fig. 6 DAVID gene ontology results showing the biological processes associated with detected peptide **(A)** and protein **(B).** SABP stands for systemic arterial blood pressure.

Figures

Fig 1

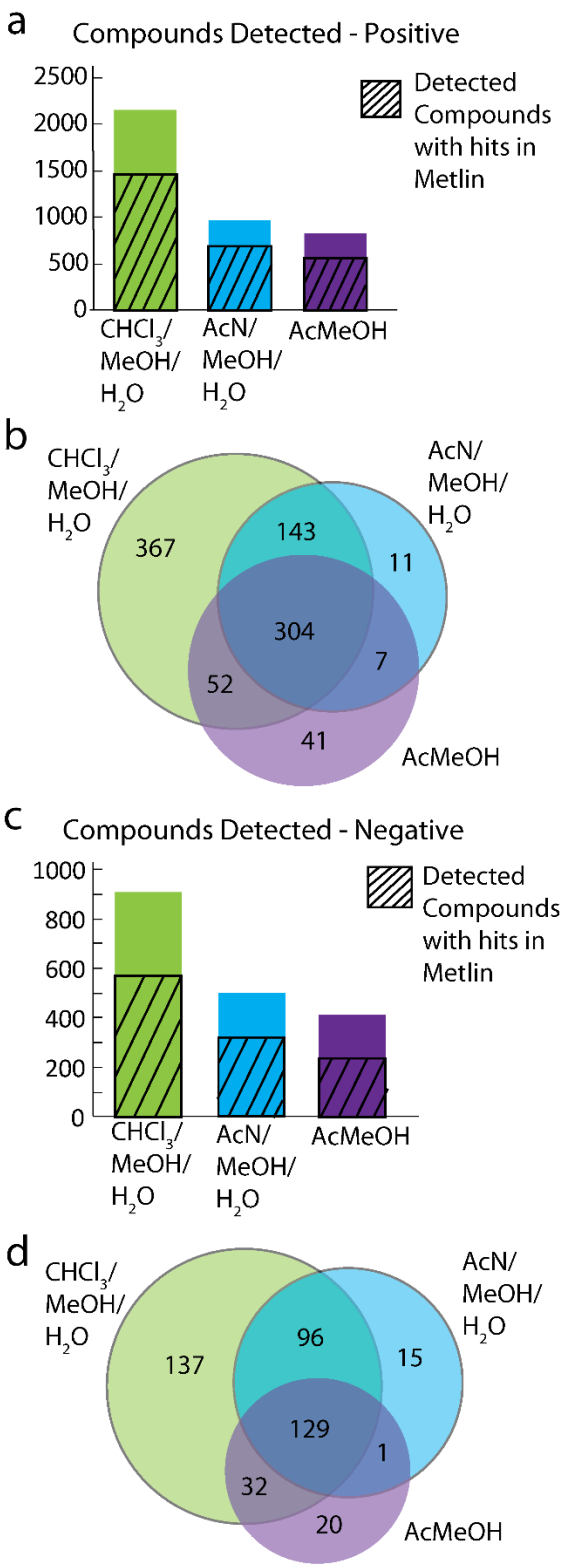


Fig 2

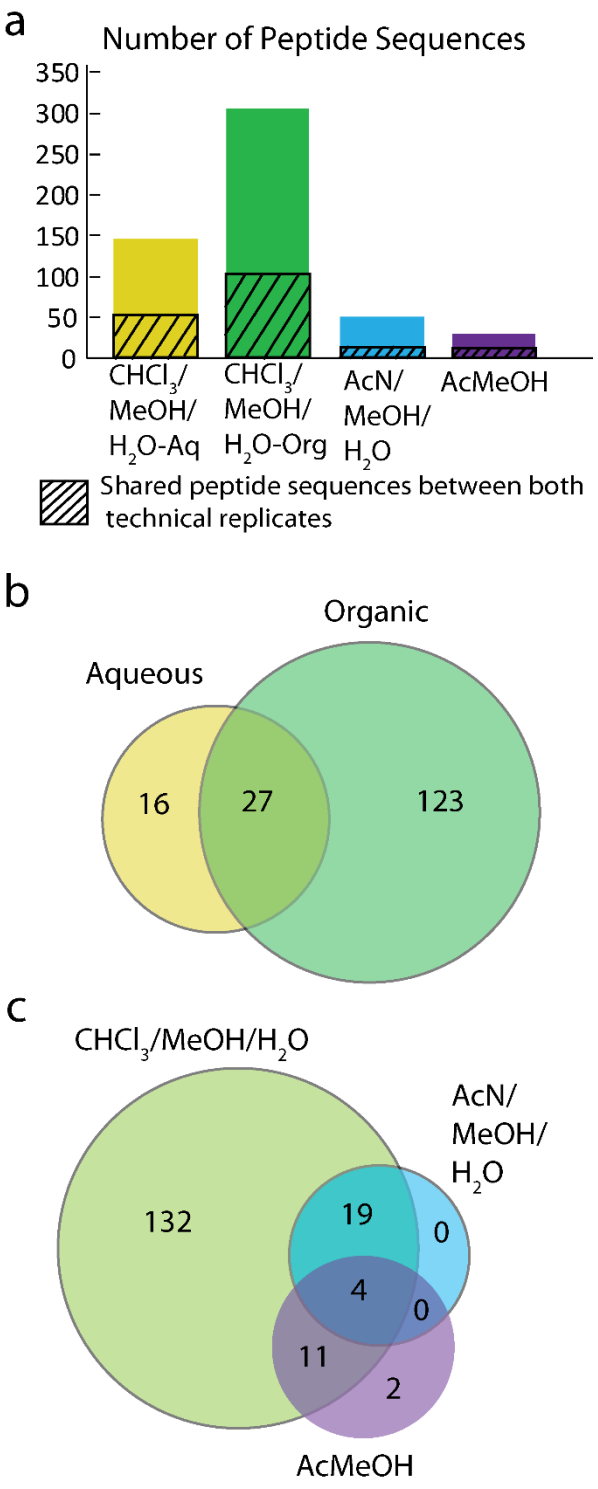

a
Number of Peptide Sequences
350
300
250
200
150
100
50
0
CHCl3/ MeOH/ H2O-Aq
CHCl3/ MeOH/ H2O-Org
AcN/ MeOH/ H2O
AcMeOH
Shared peptide sequences between both technical replicates
b
Organic
Aqueous
16
27
123
c
CHCl3/MeOH/H2O
AcN/ MeOH/ H2O
132
19
0
4
0
11
2
AcMeOH

Fig 3

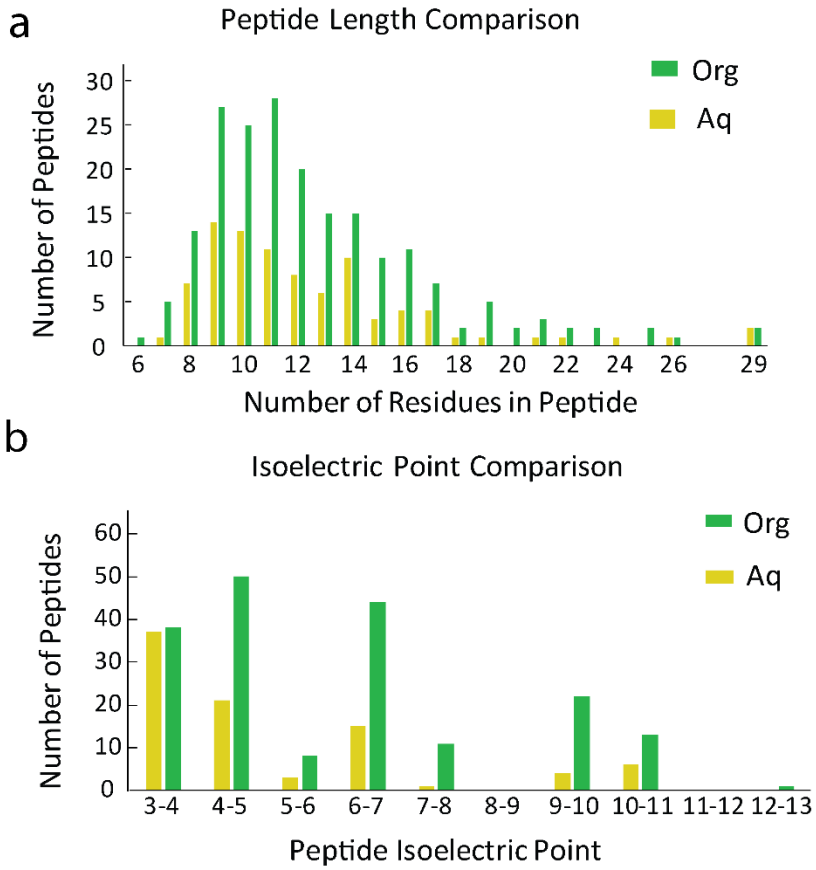

a
Peptide Length Comparison
Org
Aq
Number of Peptides
0
5
10
15
20
25
30
6
8
10
12
14
16
18
20
22
24
26
29
Number of Residues in Peptide
b
Isoelectric Point Comparison
Org
Aq
Number of Peptides
0
10
20
30
40
50
60
3-4
4-5
5-6
6-7
7-8
8-9
9-10
10-11
11-12
12-13
Peptide Isoelectric Point

Fig 4

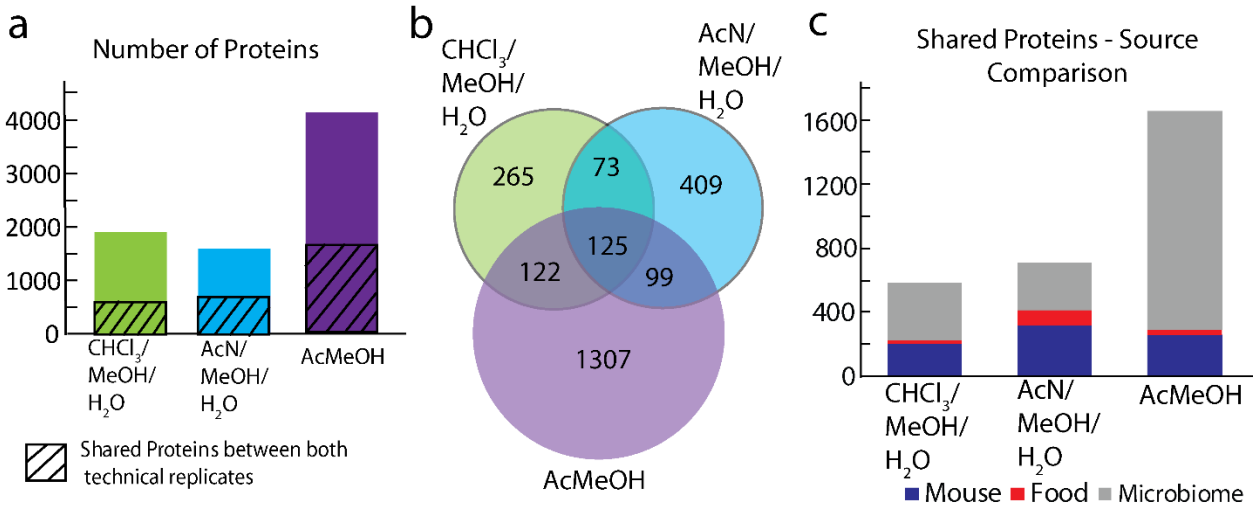

a
Number of Proteins
4000
3000
2000
1000
0
CHCl3/ MeOH/ H2O
AcN/ MeOH/ H2O
AcMeOH
Shared Proteins between both technical replicates
b
CHCl3/ MeOH/ H2O
AcN/ MeOH/ H2O
265
73
409
125
122
99
1307
AcMeOH
c
Shared Proteins - Source Comparison
1600
1200
800
400
0
CHCl3/ MeOH/ H2O
AcN/ MeOH/ H2O
AcMeOH
Mouse
Food
Microbiome

Fig 5

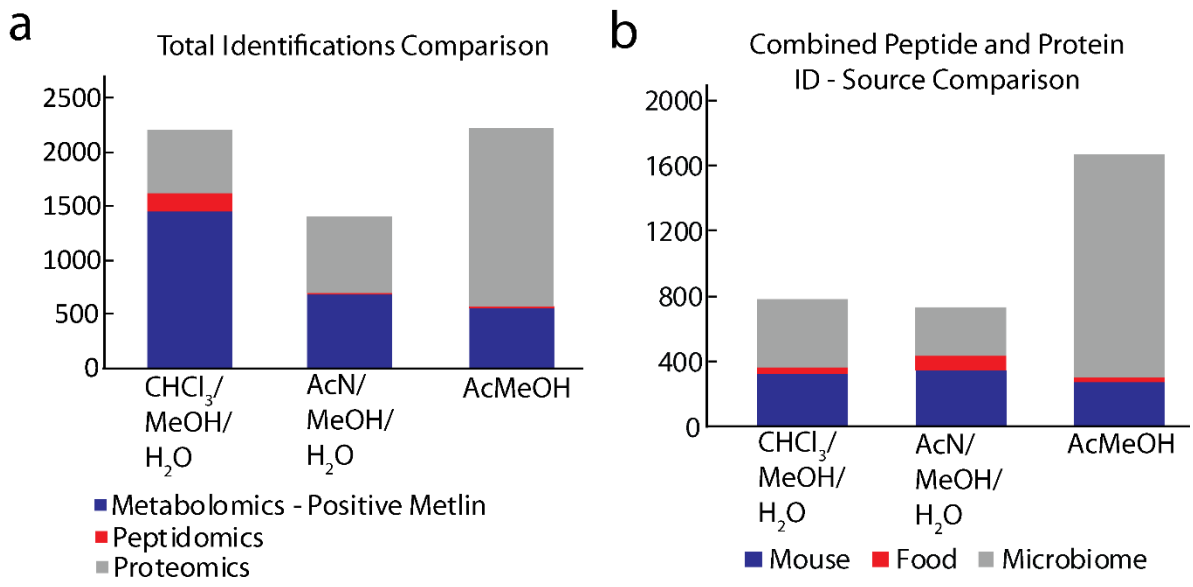
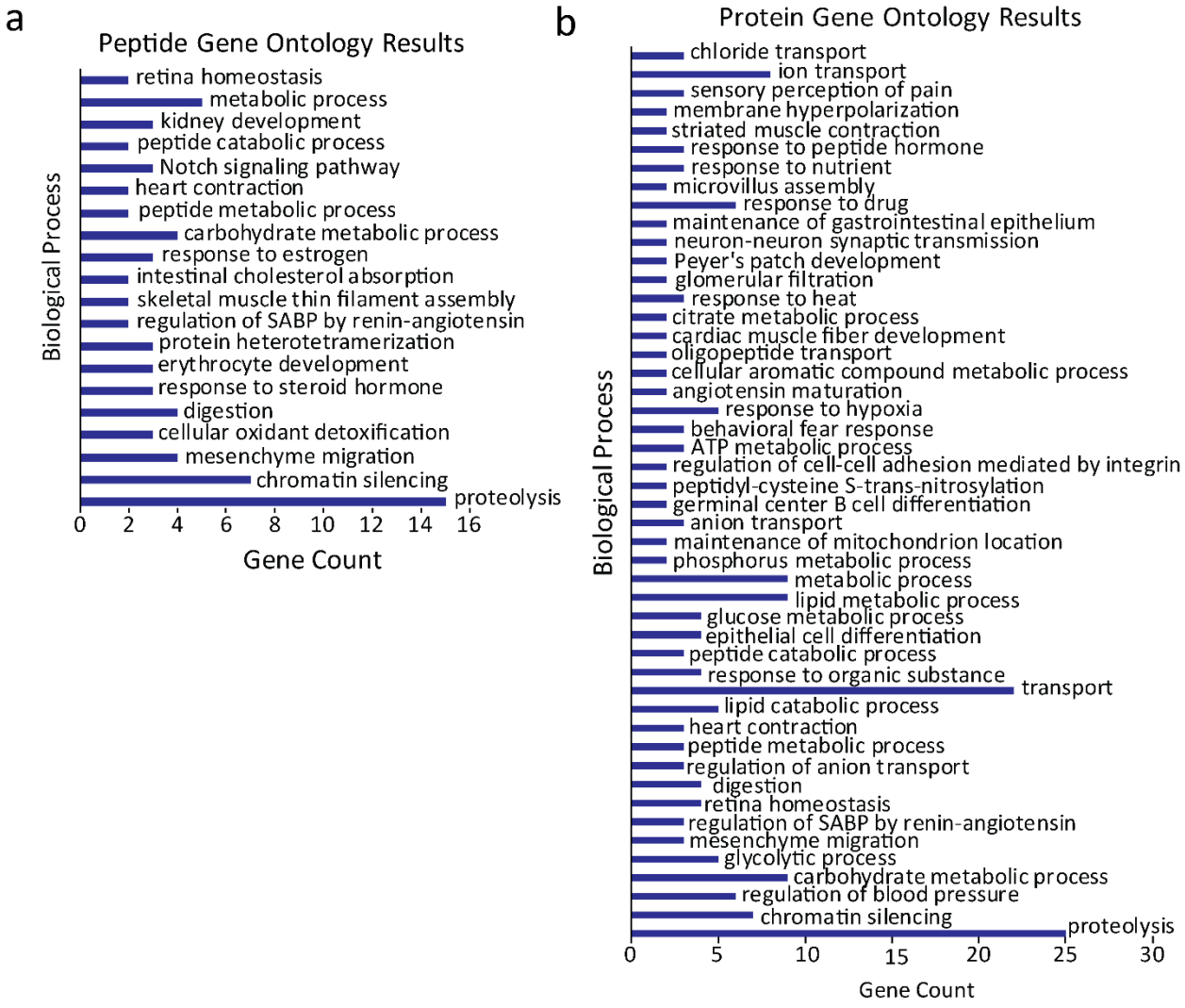
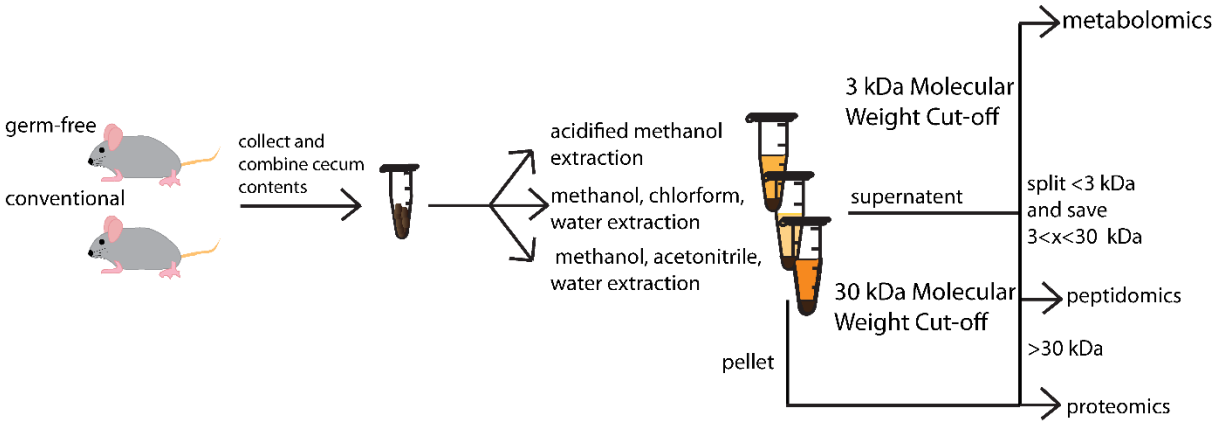

a
Total Identifications Comparison
2500
2000
1500
1000
500
0
CHCl3/ MeOH/ H2O
AcN/ MeOH/ H2O
AcMeOH
Metabolomics - Positive Metlin
Peptidomics
Proteomics
b
Combined Peptide and Protein ID - Source Comparison
2000
1600
1200
800
400
0
CHCl3/ MeOH/ H2O
AcN/ MeOH/ H2O
AcMeOH
Mouse
Food
Microbiome

Fig 6


a
Peptide Gene Ontology Results
retina homeostasis
metabolic process
kidney development
peptide catabolic process
Notch signaling pathway
heart contraction
peptide metabolic process
carbohydrate metabolic process
response to estrogen
intestinal cholesterol absorption
skeletal muscle thin filament assembly
regulation of SABP by renin-angiotensin
protein heterotetramerization
erythrocyte development
response to steroid hormone
digestion
cellular oxidant detoxification
mesenchyme migration
chromatin silencing
proteolysis
0 2 4 6 8 10 12 14 16
Gene Count
Biological Process
b
Protein Gene Ontology Results
chloride transport
ion transport
sensory perception of pain
membrane hyperpolarization
striated muscle contraction
response to peptide hormone
response to nutrient
microvillus assembly
response to drug
maintenance of gastrointestinal epithelium
neuron-neuron synaptic transmission
Peyer's patch development
glomerular filtration
response to heat
citrate metabolic process
cardiac muscle fiber development
oligopeptide transport
cellular aromatic compound metabolic process
angiotensin maturation
response to hypoxia
behavioral fear response
ATP metabolic process
regulation of cell-cell adhesion mediated by integrin
peptidyl-cysteine S-trans-nitrosylation
germinal center B cell differentiation
anion transport
maintenance of mitochondrion location
phosphorus metabolic process
metabolic process
lipid metabolic process
glucose metabolic process
epithelial cell differentiation
peptide catabolic process
response to organic substance
transport
lipid catabolic process
heart contraction
peptide metabolic process
regulation of anion transport
digestion
retina homeostasis
regulation of SABP by renin-angiotensin
mesenchyme migration
glycolytic process
carbohydrate metabolic process
regulation of blood pressure
chromatin silencing
proteolysis
0 5 10 15 20 25 30
Gene Count
Biological Process

SUPPLEMENTARY INFORMATION

Online Resource 1. Diagram showing experimental workflow



Online Resource 2. Comparison of mzCloud and MSDial spectral matching databases for metabolomics identification