UNIVERSITY OF CALIFORNIA, SAN DIEGO

A cross species metabolomics analysis on the effect of biological stressors such as depression and sleep loss on the host metabolome

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Biology

by

Fernando Vargas

Committee in charge:

Professor Pieter C. Dorrestein, Chair
Professor Rachel Dutton, Co-Chair
Professor Susan S. Golden
Professor Rob Knight
Professor Jose Pruneda-Paz

2019

The dissertation of Fernando Vargas is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2019

DEDICATION

I dedicate this to Arpa, my wife, Patricia, my mother, my family and my friends who have helped me get this far in this journey called life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FILES

1. Table 3.S1.pdf
2. Table 6.S1.pdf
3. Table 6.S3.pdf
4. Table 6.S4.pdf

ACKNOWLEDGEMENTS

It would be next to impossible to list the many people who have supported and helped me get to where I am today. My trajectory from birth to graduate school has not been linear. I have always had a natural curiosity about the world around me; starting from my time growing up as an illegal immigrant in the city of Los Angeles, to my time served as rifleman in the United State Marine Corps, and my current position as a scientist. The richness of these experiences has had a profound influence on the way I approach life. Through it all, Patricia, my mother, sacrificed everything and made this possible. Cristina, my younger sister, has been a constant reminder that I have an obligation as an older brother to make her proud. Arpa, my love and joy, has walked this crucible called graduate school next to me, and has been a constant source of love and encouragement. To the three most important people in my life, thank you for everything.

My success as a scientist can be attributed to the various scientist I can call mentors and friends. Professor Jamil Momand, my undergraduate advisor and research mentor, introduced me to the field of biochemistry and encourage me to pursue my Ph.D.. Thank you for your guidance and support. Professor Pieter C. Dorrestein, my Ph.D mentor, provided me with a once in a life time experience graduate education in mass spectrometry and its application in life sciences. I cannot express how thankful I am for your mentorship, guidance, and support. The environment that you have cultivated in the Dorrestein lab provided me with the tools and opportunities to grow and succeed as a researcher. I will always appreciate everything that you have done for me. I would also like to thank my committee members for their role in my

graduate education, thank you Professor Rob Knight, Professor Rachel Dutton, Professor Susan S. Golden, and Professor Jose Pruneda-Paz.

Chapter 2, in full, is a reprint of the material as it appears in Journal of The American Society for Mass Spectrometry 2019. da Silva, R. R; Vargas, F; Ernst, M; Nguyen, N. H; Bolleddu, S; del Rosario, K. K; Tsunoda, S. M; Dorrestein, P. C; Jarmusch, A. K.; Springer US, 2018. The dissertation author was a primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication of the material as it may appear in Rapid Communications in Mass Spectrometry, 2019, Vargas, F; Weldon, K. C; Sikora, N; Wang, M; Zhang, Z; Gentry, E. C; Panitchpakdi, M. W; Caraballo, M; Dorrestein, P. C; Jarmusch, A. K.; John Wiley & Sons, 2019. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, has been submitted for publication of the material as it may appear in Frontiers in Psychiatry, 2019, Vargas, F; Dethloff, F; Emmanuel, E; Quinn, R; Park, D. I; Herzog, D. P, Müller, M. B; Gentry, E. C; Knight, R; Gonzalez, A; Dorrestein, P. C; Turck, C. W.; Frontiers Media, 2019. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, has been submitted for publication of the material as it may appear in Scientific Reports, 2019, Thompson, R. S; Vargas, F; Dorrestein, P. C; Chichlowski, M; Berg, B. M; Fleshner, M.; Nature Publishing Group, 2019. The dissertation author was a primary investigator and author of this paper.

Chapter 6, in full, has been submitted for publication of the material as it may appear in Sleep, 2019, Bowers, S. J; Vargas, F; Gonzalez, A; He, S; Jiang, P; Dorrestein, P. C; Knight,

R; Wright, K. P; Lowry, C. A; Fleshner, M; Vitaerna, M. H; Turek, F. W.; Oxford University Press, 2019. The dissertation author was a primary investigator and author of this paper.

VITA

2014 Bachelors of Science, Chemistry, California State University, Los Angeles

2019 Doctorate of Philosophy, Biology, University of California, San Diego

PUBLICATIONS

Mohimani H, *et al*. MetaMiner: A peptidogenomics approach for the discovery of ribosomally synthesized and post-translationally modified peptides from microbial communities. *Cell Systems* (accepted July 2019)

Gauglitz MJ, *et al*. Untargeted mass spectrometry-based metabolomics approach unveils molecular changes in raw and processed foods and beverage. *Food Chemistry*. **302**, 125290 (2020).

Bolyen E, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. **37**, 852-857 (2019).

da Silva RR, **Vargas F**, *et al*. Computational removal of undesired mass spectral features possessing repeat units via a kendrick mass filter. *J Am Soc Mass Spectrom*. **30**, 268-277 (2019).

Jarmusch A, Elijah EO, **Vargas F**, *et al*. Initial development toward non-invasive drug monitoring via untargeted mass spectrometric analysis of human skin. *Anal Chem*. **91**, 8062-8069 (2019).

McCall LI, Tripathi A, **Vargas F**, *et al*. Experimental Chagas disease-induced perturbations of the fecal microbiome and metabolome. *PLoS Negl Trop Dis*. **12**, e0006344 2018).

Melnik AV, *et al*. Coupling targeted and untargeted mass spectrometry for metabolome-microbiome-wide association studies of human fecal samples. *Anal Chem*. **89**, 7549-7559 (2017).

Petras D, et al. Mass spectrometry-based visualization of molecules associated with human habitats. *Anal Chem*. **88**, 10775-10784 (2016).

Vargas F, Weldon K, Sikora N, *et al*. Open-Source Protocol for Community-Created Public MS/MS Reference Library. *RCM*. bioRxiv: https://doi.org/10.1101/804401. **(In review)**

Vargas F, *et al*. Paroxetine administration in mice alters bile acid levels. **(In review)**

Thompson RS, **Vargas F**, *et al*. The fecal metabolome is altered by stress and dietary prebiotics: Relationships between metabolomic features, sleep architecture, and microbiome alpha diversity. **(In review)**

Bowers SJ, **Vargas F**, *et al*. Repeated sleep disruption in mice leads to shifts in the fecal microbiome and metabolome that persist multiple days into recovery sleep. (**In review**).

Labarata-Bajo L, Gramalla-Schmitz A, **Vargas F**, *et al*. CD8-T-cell-induced anorexia explains chemical adaptations and blooming of an immunosuppressive commensal after chronic viral infection. *Science*. **(In review)**

Wang M, Jarmusch AK, **Vargas F**, *et al*. MASST: A web-based basic mass spectrometry search tool for molecules to search public data. *Nat Biotechnol*. bioRxiv: https://doi.org/10.1101/591016. **(In review)**

Aron A, *et al*. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat Protoc*. ChemRxiv: https://doi.org/10.26434/chemrxiv.9333212.v1. **(In review)**

Jarmusch AK, *et al*. Repository-scale Co- and Re-analysis of Tandem Mass Spectrometry Data. *Nat Protoc*. bioRxiv: https://doi.org/10.1101/750471. **(In review)**

Quinn RA, *et al*. Global Chemical Impacts of the Microbiome Include Unique Bile Acid Conjugates that Stimulate FXR. *Nature*. bioRxiv:https://doi.org/10.1101/654756. **(In review)**

Nothias LF, *et al*. Feature-based Molecular Networking in the GNPS Analysis Environment. *Nat. Methods*. bioRxiv: https://doi.org/10.1101/812404. (**In review**)

Ludwig M, *et al*. Zodiac: database-independent molecular formula annotation using Gibbs sampling reveals unknown small molecules. *Nat. Methods*. bioRxiv: https://doi.org/10.1101/842740 **(In review)**

FIELDS OF STUDY

Major Field: Biology

Studies in Metabolomics

Professor Pieter C. Dorrestein

ABSTRACT OF THE DISSERTATION

A cross species metabolomics analysis on the effect of biological stressors such as depression and sleep loss on the host metabolome

by

Fernando Vargas

Doctor of Philosophy in Biology

University of California, San Diego, 2019

Professor Pieter C. Dorrestein, Chair

Professor Rachel Dutton, Co-Chair

My dissertation research focused on the development of robust, high-throughput liquid-chromatography tandem mass spectrometry (LC-MS/MS) platforms for the rapid characterization of the gut/host metabolome. Metabolomic variations in abundance and structural information were captured via MS^1 and MS^2 spectra. Chemical differences in the metabolome between healthy and stressed subjects may serve as a diagnostic marker therefore, we aimed to develop a rapid LC-MS/MS platform to facilitate a more thorough understanding

of chemical changes within a host when stressed.

Chapter II address the challenges associated with unwanted chemical background when using swabs as a sampling device. This process required the development of the Kendrick Mass Filter (KMF) for the computational removal of undesirable background. The KMF is intended to assist in situations where chemical background removal is not possible.

Chapter III describes the development of the computational workflow MSMS-Chooser. A major hurdle in identifying chemicals in mass spectrometry experiments is the availability of MS/MS reference spectra in public databases. MSMS-Chooser enables the rapid generation of chemical standards reference spectra.

Chapter IV presents a study investigating the effect Paroxetine, a selective serotonin reuptake inhibitor (SSRI), has on the microbiome and metabolome. LC-MS/MS and 16S rRNA profiling showed several primary and secondary bile acid level, and alpha diversity differences between Paroxetine and vehicle treated mice. These results highlight the need for more analysis on the side effects of drugs like SSRI's.

Chapter V presents a study testing whether stress and/or dietary prebiotics alter the fecal metabolome. LC-MS/MS and 16S rRNA gene sequencing, showed that both stress and Test diet altered the fecal metabolome/microbiome. These results reveal novel microbial-dependent metabolites that may modulate stress physiology and sleep.

Chapter VI presents a study investigating the effect repeated sleep disruption have on the fecal microbiome and metabolome. We found global shifts in both the microbiome and metabolome in the sleep-disrupted group on the second day of recovery sleep, when most sleep parameters had recovered to baseline levels. Thus, repeated sleep disruption causes changes in

key features of the fecal microbiome and metabolome, some of which last for days after recovery of objective sleep measures.

Chapter I

Introduction

A. Biological Stressors and the Microbiome

The gut microbiota has been shown to influence physiology, health, behavior, and stress (1-14); however, the particular microbes or microbial communities associated with physiological and behavior host response to stress remain largely unknown. Changes in the gut microbiome may affect the MGB axis through modulation of metabolite signals and associated changes in the host metabolome (15-23). Sleep deprivation and circadian misalignment, common during military operations, have been associated with microbiome dysbiosis, gastrointestinal distress, altered host metabolome, and impaired cognitive function, making them ideal conditions to study the MGB axis (23-30). Prebiotic dietary supplementation has been shown to change gut microbiota composition and confer health benefits, demonstrating its potential in gut microbiome therapy (16-19). One approach to gain access to the functional role of microbes is to understand the chemistry they produce, the chemistry they influence or how the chemistry enables the formation of a community A metabolomic perspective of the gut microbiota's role in host stress response and host-microbiota interactions is critical for the development of effective strategies to increases stress robustness during military operations.

B. Metabolomics

Metabolomics focuses on the identification and characterization of small molecules, or metabolites, within a biological system. The ability to screen, diagnose, and monitor neonatal metabolic disorders in clinics (e.g. fatty acid oxidation disorders, amino acids disorders, and organic acidemias) has been one of the most important and successful applications of metabolomics in medical history. The two main methods used in metabolomic studies are

nuclear magnetic resonance (NMR) and mass spectrometry (MS). Despite great improvements in sensitivity, NMR still has orders of magnitude less dynamic range and sensitivity compared to MS. Therefore, NMR is seeing limited adoption by the larger microbiome community. MS is a powerful analytical instrument that measures molecules as charged ions and reports back their relative abundance. Its high sensitivity and sensitivity have made it an invaluable tool in metabolomic studies.

Liquid chromatography mass spectrometry (LC-MS), has revolutionized the way molecules are analyzed. When liquid chromatography is applied to complex samples, physical separation of the molecules is based on the their affinity for the mobile phase and stationary phase of the column. LC-MS is frequently used to study the metabolome due to its ability to chromatographically separate and detect a large array of molecules in a high throughput fashion. If one is interested in specific molecules like host derived or microbially modified bile acids, microbially derived short chain fatty acids, or host-dependent tryptophan derivatives a targeted analysis is preferred. Targeted mass spectrometry is 20-1,000 fold more sensitive when comparing compared to an untargeted analysis. However, targeted mass spectrometry requires *a priori* knowledge of the molecules of interest, thereby limiting it utility in explorative studies. In untargeted metabolomics, one takes a broader look at the molecules present in the sample, and one does not have to have *a priori* knowledge of what molecules are going to be observed. One of the goals of untargeted mass spectrometry, is to develop chemical profiles of biological important phenotypes, and identify molecules or mass shifts within the sample via multivariant analysis. Such annotations can then be leveraged for targeted analysis in future experiment; thereby, increasing our understanding of the underlying biology of an observed phenotype.

C. Research Objectives

The development of a robust LC-MS/MS platform for the rapid analysis of the host/gut metabolome stress response was the overarching objective of my dissertation research. The chemical changes that occurred when a host is stressed may inform the functional role microbes play during the host stress response. We sought to acquire the metabolomic profile of the host/gut before, during, and after three different stressors: drug exposure, inescapable tail shock, and repeated sleep disruptions. Two computational methods, Kendrick Mass Filter (KMF) and MSMS-Chooser, were developed to assist in this goal. The KMF workflow provides users with a method for the computation removal of unwanted chemical backgrounds, a common issues when samples are collected on swab sampling devices. The MSMS-Chooser workflow was developed to help address the annotation challenge that has prevented the identification of ~95% of molecules in a typical LC-MS/MS run by creating an easy to use platform for the rapid generation of public high quality reference MS^2 spectra. To measure the impact of prescription medication on gut homeostasis, one of the studies measured the effect paroxetine had on the gut metabolome. LC-MS/MS analysis showed that several bile acids were altered when mice were exposed to paroxetine, an unintended side effect that has never before been observed. To investigate the role prebiotics play in modulating stress, another study tested if stress and/or dietary prebiotics (Test diet) altered the fecal metabolome and explored if these changes were related to sleep and/or gut microbial alpha diversity. Network propagation analysis revealed that stress increased members of a neuroactive steroidal molecular family; and that the Test diet reduced this effect. To assess the impact of repeated sleep disruption on the host/gut metabolome, the last study exposed mice to five days of sleep disruption and assed the impact

on the gut metabolome. This study identified multiple classes of fecal metabolites that were differentially abundant in sleep-disrupted mice, which can potential be targeted to increase stress resilience in the context of sleep disruption. The culmination of my research is the development and implementation of an robust LC-MS/MS platform and data analysis workflow for the data acquisition and analysis of the host/gut metabolome after exposures to stressors like: sleep disruptions, drug exposure and inescapable physical trauma.

D. References

1. Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, Kau AL, Rich SS, Concannon P, Mychaleckyj JC, Liu J, Houpt E, Li JV, Holmes E, Nicholson J, Knights D, Ursell LK, Knight R, Gordon JI; Gut Microbiomes of Malawian Twins Pairs Discordant for Kwashiorkor. Science, 2013. 339(6119): p. 548-54.

2. Nelson DB, Rockwell LC, Prioleau MD, Goetzl L; The role of the bacterial microbiota on reproductive and pregnancy health. Anaerobe, 2016. 42: p. 67-73.

3. Cong X, Xu W, Romisher R, Poveda S, Forte S, Starkweather A, Henderson WA; Gut Microbiome and Infant Health: Brain-Gut-Microbiota Axis and Host Genetic Factors. Yale J Biol Med, 2016. 89(3): p. 299-308.

4. Zapata HJ, Quagliarello VJ; The microbiota and microbiome in aging: potential implications in health and age-related diseasess. J Am Geriatr Soc, 2015. 63(4): p. 776-81.

5. Althani AA, Marei HE, Hamdi WS, Nasrallah GK, El Zowalaty ME, Al-Khodor S, Al-Asmakh M, Abdel-Aziz H, Cenciarelli C; Human Microbiome and its Association With Health and Disease. J Cell Physiol, 2016. 231(8): p. 1688-94.

6. McHardy IH, Goudarzi M, Tong M, Ruegger PM, Schwager E, Weger JR, Graeber TG, Sonnenburg JL, Horvath S, Huttenhower C, McGovern DP, Fornace AJ Jr, Borneman J, Braun J; Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. Microbiome, 2013. 1(17).

7. Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, Jansson JK, Dorrestein PC, Knight R; Microbiome-wide association studies link dynamic microbial consortia to disease. Nature, 2016. 535: p. 94-103.

8. Antharam VC, McEwen DC, Garrett TJ, Dossey AT, Li EC, Kozlov AN, Mesbah Z, Wang GP; An Integrated Metabolomic and Microbiome Analysis Identified Specific Gut Microbiota Associated with Fecal Cholesterol and Coprostanol in Clostridium difficile Infection. PLoS one, 2016. 11(2): e0148825.

9. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, DuGar B, Feldstein AE, Britt EB, Fu X, Chung Y, Wu Y, Schauer P, Smith, JD, Allayee H, Tang WHW, DiDonato JA, Lusis AJ, Hazen SL; Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature, 2011. 472: p. 57-63.

10. Lapthorne S, Pereira-Fantini PM, Fouhy F, Wilson G, Thomas SL, Dellios NL, Scurr M, O'Sullivan O, Ross RP, Stanton C, Fitzgerald GF, Cotter PD, Bines JE; Gut

microbial diversity is reduced and is associated with colonic inflammation in a piglet model of short bowel syndrome. Gut Microbes, 2013. 4(3): p. 212-21.

11. 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; A core gut microbiome in obese and lean twins. Nature, 2009. 457(7228): p. 480-4.

12. Dicksved J, Halfvarson J, Rosenquist M, Jarnerot G, Tysk C, Apajalahti J, Engstrand L, Jansson JK; Molecular analysis of the gut microbiota of identical twins with Crohn's disease. ISME J, 2008. 2(7): p. 716-27.

13. Bosscher D, Breynaert A, Pieters L, Hermans N; Food-based strategies to modulate the composition of the intestinal microbiota and their associated health effects. Journal of physiology and pharmacology : an official journal of the Polish Physiological Society, 2009. 60 Suppl 6: p. 5-11.

14. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, Rappaport SM, van Der Hooft JJJ, Wishart DS; The food metabolome: a window over dietary exposure. Am J Clin Nutr, 2014. 99(6): p. 57-63P: p. 1286-1308.

15. Dorrestein PC, Mazmanian SK, Knight R; Finding the missing links among metabolites, microbes, and the host. Immunity, 2014. 40(6): p. 824-32.

16. Lamichhane S, Yde CC, Forssten S, Ouwehand AC, Saarinen M, Jensen HM, Gibson GR, Rastall R, Fava F, Bertram HC; Impact of dietary polydextrose fiber on the human gut metabolome. Journal of agricultural and food chemistry, 2014. 62(40): p. 9944-51.

17. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JI; Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science, 2013. 341(6150): p. 1241214.

18. Ursell LK, Haiser HJ, Van Treuren W, Garg N, Reddivari L, Vanamala J, Dorrestein PC, Turnbaugh PJ, Knight R. The intestinal metabolome: an intersection between microbiota and host; The intestinal metabolome: an intersection between microbiota and host. Gastroenterology, 2014. 146(6): p. 1470-6.

19. Fallucca F, Porrata C, Fallucca S, Pianesi M; Influence of diet on gut microbiota, inflammation and type 2 diabetes mellitus. First experience with macrobiotic Ma-Pi 2 diet. Diabetes/metabolism research and reviews, 2014. 30 Suppl 1: p. 48-54.

20. Rook GA, Lowry CA, Raison CL; Microbial 'Old Friends', immunoregulation and stress resilience. Evolution, medicine, and public health, 2013. 2013(1): p. 46-64.

21. 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; The Human Urine Metabolome. PLoS one, 2013. 8(9): e73076

22. Larsen PE, Dai Y; Metabolome of human gut microbiome is predictive of host dysbiosis. Gigascience, 2015. 4(42).

23. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites; 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, 2008. 106(10): p. 3698-3703.

24. Thaiss CA, Zeevi D, Levy M, Zilberman-Schapira G, Suez J, Tengeler AC, Abramson L, Katz MN, Korem T, Zmora N, Kuperman Y, Biton I, Gilad S, Harmelin A, Shapiro H, Halpern Z, Segal E, Elinav E; Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. Cell, 2014. 159(3): p. 514-29.

25. Laposky AD, Bass J, Kohsaka A, Turek FW; Sleep and circadian rhythms: key components in the regulation of energy metabolism. FEBS Lett., 2008. 582(1): p. 142-151.

26. Laposky AD, Turek FW; Physiologic and health consequences of circadian disruption (in animal models). Sleep Med Clin, 2009. 4: p. 127-142.

27. Tang Y, Preuss F, Turek FW, Jakate S, Keshavarzian A; Sleep deprivation worsens inflammation and delays recovery in a mouse model of colitis. Sleep medicine, 2009. 10(6): p. 597-603.

28. Voigt RM, Forsyth CB, Green SJ, Mutlu E, Engen P, Vitaterna MH, Turek FW, Keshavarzian A; Circadian disorganization alters intestinal microbiota. PloS one, 2014. 9(5): p. e97500.

29. Kim Y, Laposky AD, Bergmann BM, Turek FW; Repeated sleep restriction in rats leads to homeostatic and allostatic responses during recovery sleep. Proceedings of the National Academy of Sciences of the United States of America, 2007. 104(25): p. 10697-702.

30. Leproult R, Holmback U, Van Cauter E; Circadian misalignment augments markers of insulin resistance and inflammation, independently of sleep loss. Diabetes, 2014. 63(6): p. 1860-9.

Chapter II

Computational Removal of Undesired Mass Spectral Features Possessing Repeat Unit via a Kendrick Mass Filter

A. Abstract

Polymers are a common component of chemical background which complicates data analysis and can impair interpretation. Undesired chemical background cannot always be addressed via pre-analytical methods, chromatography, or existing data processing methods. The Kendrick Mass Filter (KMF) is presented for the computational removal of undesired signals present in MS^1 spectra. The KMF is analogous to mass defect filtering but utilizes homology information via Kendrick mass scaling in combination with chromatographic retention time and the number of observed signals. The KMF is intended to assist in situations in which current data processing methods to remove background, e.g. blank subtraction, are either not possible or effective. The major parameters affecting KMF were investigated using PEG 400 and NIST standard reference material 1950 (metabolites in human plasma). Further exploration of the KMF performance was tested using an extract of a swab known to contain polymers. An illustrative real-world example of skin analysis with polymeric signal is discussed. The KMF is also able to provide a high-level view of the compositionality of data regarding the presence of signals with repeat units and indicate the presence of different polymers.

B. Introduction

Mass spectrometry (MS) studies are prone to undesired chemical background. One source of undesired chemical background is polymers, such as polyethylene glycol (PEG), which are ubiquitous. Undesired polymer background can often be avoided by a trained scientist

under controlled laboratory conditions; however, the task of avoiding undesired polymer background is more difficult when collecting samples outside of the laboratory (e.g. sample collection by citizen-scientists). The typical mass spectrum that results from the presence of polymers is complex and consists of oligomers signals separated by the mass-to-charge (*m/z*) of the polymer unit repeat, viz. PEG spectra will contain ions of $C_{2n}H_{4n+2}O_{n+1}$. The presence of polymers and similar undesired chemical background can be so impactful as to preclude data interpretation, therefore methods to remove such interferences are needed. Pre-analytical methods, e.g. solid phase extraction, is a common means by which to negate the effect of unwanted background; however, such methods can be costly, time consuming, and often modify the molecule composition of the sample. Another possibility is to remove interfering chemical background through chromatographic methods. Compensation for chemical background can also be performed by data processing methods, e.g. blank subtraction; however, this approach relies on co-analyzing samples which faithfully recapitulate the source of the undesired chemical background. When the source is unknown or not anticipated in the experimental design, removal of undesired chemical background is challenging. Here, we propose using the Kendrick Mass Filter (KMF) to assist in computationally removing undesired polymer signals. The KMF is intended to address the following gap: (i) the polymer background cannot be removed by pre-analytical methods (or modification of the molecular composition is unwanted, e.g. untargeted metabolomics), (ii) data processing via blank subtraction is not possible as the source cannot be faithfully recapitulated, or (iii) cases in which data has already been collected but rendered useless by undesired chemical background.

The Kendrick mass filter combines Kendrick mass scaling with mass defect filtering (MDF) (1). The Kendrick mass is calculated by rescaling the *m/z* of each ion to an integer value of the unit repeat, differing from the IUPAC definition (i.e. ^{12}C is equal to 12 unified atomic mass units). The defect, i.e. Kendrick mass defect (KMD), between the Kendrick scaled *m/z* and the integer Kendrick mass value (i.e. rounded Kendrick scaled *m/z*) is similar between homologous compounds. Homologous compounds can be readily identified by plotting of integer Kendrick values versus the KMD, providing an interpretable scatterplot, in which homologous compounds are horizontally aligned. Kendrick mass plots and similar visualizations, e.g. Van Krevelen diagrams, have been applied in the fields of petroleomics (2), dissolved organic matter (3), and other complex mixtures. Improvement to the visualization of Kendrick mass plot continues, most recently with the introduction of fractional base units which improves the visual resolving power of polymers (4). Mass Defect Filtering (MDF) has been used to perform selection and removal of data centered around a user-defined mass defect, calculated using the IUPAC mass scale (5). MDF has been applied in the study of drug metabolism (6), removal of salt clusters in LC-MS metabolomics data (7), and natural product chemistry (8). The KMF is rooted in MDF analysis but utilizes additional information via the KMD that can be used to determine homology. Here, we report the proof-of-concept for the computational removal of undesired mass spectral features possessing repeat units by KMF.

C. Materials and Methods

1. Sample Preparation and Data Acquisition

NIST standard reference material 1950 metabolites in frozen human plasma (9), polyethylene glycol 400 (PEG 400), swab extracts, and human skin samples collected using swabs were analyzed using liquid chromatography – mass spectrometry using a quadrupole-Orbitrap mass spectrometer (Q Exactive, Thermo Scientific) or quadrupole time-of-flight (ToF) mass spectrometer (maXis Impact, Bruker). Sample preparation and instrumental parameters are detailed in the Supplementary Information. QExactive files (.raw) were converted to .mzXML via MSConvert (10). The qToF files (.d) were exported using DataAnalysis (Bruker) as .mzXML files after lock mass correction. MS^1 feature finding was performed subsequently in MZmine2 (11), with parameters described in the Supplementary Information. The MS^1 feature matrix for NIST plasma, swab extract, PEG 400, and swab spiked plasma was split and individual matrices were compared. Shared variables as well as variables for which all peak area values were zero were excluded. The same operations were performed on data collected at 17,500 and 70,000 mass resolution on the QExactive.

2. Kendrick Mass Filtering and Data Processing

Kendrick mass filtering is performed by importing the MS^1 feature matrix and rescaling the m/z values according to the unit repeat (Equation S1), viz. when filtering for PEG the measured m/z are scaled by 44 divided by 44.0262. The rescaled data is then used to calculate the KMD via subtraction of the exact Kendrick mass from the nominal Kendrick mass (Equation S2). All possible Kendrick scaled differences in m/z are calculated pairwise. The difference between two peaks must be the integer repeat unit (e.g. PEG, 44) and the ΔKMD less than the user-defined ΔKMD. The retention time (RT) window (e.g. 60 s) is used in

determining filtering. In addition to ΔKMD and retention time exclusion criteria, the pairwise comparison of MS features, which meet ΔKMD and RT criteria is used to determine the number of observed signals (NOS) (**Figure 2.S1**). The final matrix is exported as comma-separated values (.csv).

The Kendrick Mass Filter (KMF) computationally removes ions (or MS^1 features if performing chemical separation prior to MS), regardless of spectral abundance, from mass spectral data, which meet filtering criteria. Our implementation of the KMF allows the user to customize three parameters: Kendrick Mass Defect (ΔKMD), chromatographic retention time (RT), and the number of observed signals (NOS). We tested all combinations of the following parameter values: ΔKMD, 0.001, 0.0015, 0.002, 0.0025, 0.005, 0.0067, 0.0075, 0.01, 0.0125, 0.015, 0.02, 0.025, 0.033, 0.05, 0.067, 0.1; RT, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00, 1.25, 1.50, 1.75, 2.00; and NOS, 2, 3, 4, 5, 6. Scaling of mass spectral data to the Kendrick mass scale and computation of the KMD provides a common visual output, the Kendrick mass plot, in which homology is represented by horizontal alignment of points. Kendrick mass scaling, or any rescaling, preserves the *m/z* differences present in the samples. The rationale for scaling in the KMF algorithm, beyond the visual benefit of the Kendrick mass plots, is that a KMD value can be defined rather than defining a slope and y-axis offset as is necessary if the non-scaled mass spectral data is used, illustrated by the Kendrick mass plot and mass defect plot of PEG 400 in **Figure 2.S2** Retention time is only relevant when chemical separation is performed prior to mass spectrometric analysis. The rationale behind inclusion of the retention time is that the probability of non-desired filtering is likely to increase as the RT criterion is relaxed. In the case of polymers, retention time is influenced by the oligomer length

and chemical composition. Fundamentally, each oligomer can be separated in time using chromatography; however, impractical and tangential when analyzing samples for non-polymeric molecules. The inclusion of the number of observed signals (NOS) criterion is intended to increase the stringency of filtering. Only spectral peaks (or MS^1 features) which have a sufficient number, defined by the user, of observed peaks (or MS^1 features) with $\Delta m/z$ equal to the repeat unit will meet the criterion. The minimum NOS value is 2, which corresponds to a pair of oligomer signals. The rationale is that if homologous molecules with repeat units are present then multiple oligomer ions will be detected as is the case in mass spectra of polymers. Fundamentally, the greater the NOS equates to more specific filtering; the cost of specificity is the potential for features to remain unfiltered (loss of sensitivity). In practice, the balance between non-specific filtering versus specific removal of undesired signals (MS^1 features) must be assessed on a case-by-case basis.

3. Data Availability

The KMF and associated plots are available on GitHub (https://github.com/DorresteinLaboratory/Kendrick_Mass_Filter). The KMF was written in R and is available as a Jupyter notebook. Data discussed in this manuscript are publically available at MassIVE (http://massive.ucsd.edu) via the following MassIVE IDs: MSV000081544 and MSV000081548. Principal component analysis (PCA) was performed in R using the pcaMethods package applying the Nonlinear Iterative Partial Least Squares algorithm, after Pareto scaling (12). Figures were generated with R standard plot function and ggplot2 package and formatted in Adobe Illustrator.

D. Results and Discussion

1. Optimization for PEG 400 removal while minimizing feature removal from NIST plasma standard reference material

The effect of KMF parameters will vary with different data and parameters should be selected carefully, balancing removal of features with specificity. We defined the following goal: maximize the filtering of PEG 400 features while minimizing the number of plasma features filtered. We evaluated parameter selection using plasma spiked with PEG 400. The ideal parameters in this instance were determined by plotting the ratio of PEG 400 to plasma MS^1 features filtered versus the number of PEG 400 MS^1 features filtered; the points were colored by the MS^1 features filtered from the plasma spiked with PEG 400 (**Figure 2.1**). Triplicate technical measurements, of PEG 400, plasma spiked, and plasma spiked with PEG 400 were used. Each point in the plot represents a different set of KMF parameters (restricted to parameters tested). The uppermost grouping corresponded to the greatest filtering of PEG 400 features (y-axis) as well as the greatest number of filtered features in the plasma spiked with PEG 400 sample (colored light-green and yellow). The vertical groupings result from different NOS values. In fulfilling the defined goal (the smallest ratio), the leftmost points in the uppermost grouping are the most suitable sets of KMF parameters. Contrastingly, if filtering is desired with no regard to potential over filtering, then points (sets of KMF parameters) in the rightmost and uppermost grouping and their respective parameters should be used. An interactive plot was created (see Experimental section), which displayed the x and y values when hovering over a point. The values were used to look up the associated set of KMF

parameters (ΔKMD, RT, and NOS) in the optimization table (electronic supplementary information – **Table S2.1**), and displayed in **Figure 2.S3**. The parameters chosen (associated with the point 0.4285, 84 in **Figure 2.1**) were as follows: ΔKMD of 0.01, RT of 0.8 min, and NOS of 2.

The effect of KMF parameters (ΔKMD, RT, and NOS) on the number of MS^1 features filtered were systematically explored using PEG 400 and NIST 1950 plasma standard reference material (SRM) based on the previously determined KMF parameters. The effect of KMF is displayed in **Figure 2.2a-b**, isolating RT as a variable (held constant at 0.8 min). PEG 400 MS^1 features filtered, i.e. those meeting the KMF criteria, (**Figure 2.2a**) were substantially affected by the ΔKMD parameter. The number of features filtered quickly increased and arrived at a plateau resulting from the removal of all apparent PEG features. The filtering of plasma MS^1 features, not desired, were less affected by the KMF with ΔKMD values <0.015, **Figure 2.2b**, filtering increased with increasing ΔKMD. The NOS parameter, increasing from 2 (red) to 6 (purple), reduced the overall number of features filtered in PEG 400 as well as plasma which reflects the intended increase in stringency. **Figure 2.2c**, the overall number of MS^1 features filtered were less with a NOS of 6 compared to a NOS of 2 (most to least stringent, respectively), but a similar ratio of PEG 400 to plasma features is obtained at the local minimum (~0.01) which mirrors the ΔKMD parameter chosen during optimization.

The effect of RT on filtering is presented in **Figure 2.2d-e**, isolating ΔKMD (value held constant at 0.01). The filtering of PEG 400 MS^1 features increased with a larger RT parameter. The RT parameter is expected to change based on the chromatographic separation. **Figure 2.2d**, filtering increased quickly with small RT parameter (<0.25 min) increases, afterwards filtering

of features plateau with a retention time greater than 0.5 min. The filtering of plasma MS^1 features, **Figure 2.2e**, increased more slowly at small RT parameter values comparatively. This observation supports that PEG is selectively filtered, but the probability of selecting non-desired signals increases as RT increases. Increasing the NOS parameter reduces the overall number of features filtered in PEG 400 and plasma, **Figure 2.2f**, similar to the behavior observed with the ΔKMD parameter.

The ΔKMD values tested mirror the range of *m/z* accuracy from high to low mass resolution analyzers, e.g. Orbitrap and ToF to linear ion traps to quadruples. The results of the KMF, and any similar mass defect filter, are dependent on *m/z* accuracy. Mass drift, space charge, and peak symmetry could all influence KMD, not evaluated extensively here. However, the effect of acquiring data at different mass resolutions (17,500 and 70,000 using an Orbitrap mass analyzer) was briefly explored. The number of PEG 400 feature filtered were similar between data collected at 17,500 and 70,000; however, the number of plasma features filtered was generally less at 70,000 compared to 17,500. This behavior is believed to be attributed to the improved mass accuracy in data acquired at greater mass resolution. The full extent of different data acquisition parameters was not tested in this work.

The data which results from KMF of PEG 400 (**Figure 2.3a-e**) and plasma (**Figure 2.3f-j**) collected at a resolution of 17,500 are displayed; KMF results for data collected at a resolution of 70,000 can be found in **Figure 2.S4**. The Kendrick mass plot for PEG 400, **Figure 2.3a**, displays the MS^1 features originally present (black) and those which were filtered (red). A large number of the horizontally aligned features within a narrow Kendrick mass defect, suggesting homologous molecules, were filtered. The majority of the MS^1 features which we identified as

PEG 400 oligomers (MS^2 supporting putative identification is shown in **Figure 2.S5**), eluting between two and four minutes, could be removed using the defined parameters, **Figure 2.3b**. The MS^1 feature spectrum, i.e. plot of all *m/z* values and their corresponding abundance regardless of their retention time, prior to application of the KMF is displayed in **Figure 2.3c**. The spectrum of MS^1 features retained (those not filtered), **Figure 2.3d**, indicated near complete removal of PEG 400 oligomer signals. The spectrum of MS^1 features filtered by the KMF is displayed in **Figure 2.3e** which contains many oligomer signals present in the original spectrum. Upon closer inspection of the *m/z* differences of signals which were not filtered in the PEG 400 sample, differences between the apparent oligomer peaks did not match the Δ *m/z* requirement to be filtered. The MS^1 feature finding algorithm can cause the *m/z* to fluctuate, by averaging the *m/z* over the aligned samples, and is believed to be the origin of the observed mass difference rather than *m/z* measurement error. The filtering of plasma was minimal, a desired result in this instance, as evident in the Kendrick mass plot (**Figure 2.3f**), MS^1 feature plot (**Figure 2.3g**), and the MS^1 feature spectrum prior to filtering (**Figure 2.3h**), MS^1 features retained after filtering (**Figure 2.3i**), and spectrum MS^1 features removed by the KMF (**Figure 2.3j**).

In addition to visual inspection of the filtered peaks as displayed in **Figure 2.3** and parameter optimization, we recommend the use of Kendrick mass plots and fractional base units as described in Fouquet and Sato which can improve the visual resolving power of oligomeric series (4). The KMF using fractional base units is available in the supplementary code (https://github.com/DorresteinLaboratory/Kendrick_Mass_Filter) allowing one to quickly change the fractional base unit and create plots. Another visual inspection tool available in the

supplementary code is the Gaussian shape tool that takes advantage of the Gaussian shape of most oligomeric series.

2. Testing of KMF for the removal of PEG in NIST plasma standard reference material spiked with a swab extract

An aliquot of plasma was spiked with a swab extract, which was known to contain polymers, and measured in triplicate. This sample illustrates a scenario in which complex undesired chemical background is present. The MS^1 spectra indicated polymer ions, primarily PEG oligomers, which was confirmed by MS^2, **Figure 2.S6**. The KMF parameters chosen previously, maximizing the filtering of PEG 400 while minimizing the number of plasma features filtered, were used (ΔKMD = 0.01, RT = 0.8 min, and NOS = 2). The Kendrick mass plot, **Figure 2.4a**, and MS^1 feature plot, **Figure 2.4b**, display a large number of features, which met the KMF criteria. The original MS^1 feature spectrum, **Figure 2.4c**, is composed predominantly of signals, which appear to originate from plasma. However, a large number of peaks at approximately 10% of the base peak appeared to be oligomers (based on equal spacing). Those apparent polymer features are reduced in the KMF filtered MS^1 feature spectrum (**Figure 2.4d**). The KMF features which were removed, plotted in **Figure 2.4e**, indicated a large number of low abundance ions. Incidentally, one abundant ion at *m/z* 496.3400 which is not believed to be an oligomer signal was filtered.

3. KMF of axilla skin swab samples in organ transplant cohort reduces spectral complexity associated with uncontrolled deodorant use

Untargeted metabolomics analysis performed on human skin samples from organ transplant patients on immunosuppressive therapy (n = 302), sampled using moistened cotton swabs, were processed using the KMF. One initial question posed was whether the endogenous metabolomic information acquired from hand, face, and axillary skin samples were different. PCA was performed on the untargeted metabolomic data after row sum normalization and pareto scaling. The PCA score plot, principal component 1 (i.e. PC_1) vs principal component 2, for the original data is plotted in **Figure 2.5**. The molecular differences detected in hand, face, and axilla samples resulted in only very moderate separation of hand, face, and axillary samples (red, green, and black, respectively). The dispersion of sample points in the score plot, **Figure 2.5a**, along PC_1 (the PC of greatest contribution to data variance; 8.6%) was investigated and found to be partly due to the presence of polypropylene glycol (PPG) oligomer ions. Additional PCA score plots and loading values can be found in the supplementary information (**Figure 2.S7**, **Table 2.S1**, and electronic supplementary information – **Table 2.S2**). The PPG oligomer signals were characteristically separated by *m/z* 58.0419 (an illustrative example is shown in **Figure 2.6c**). These peaks were detected in the axilla samples from a number of, but not all, subjects. We hypothesize that the presence of PPG is related to deodorant use as PPG is present in the formulation of deodorant and other skin care products.

In this instance, the removal of the variance due to the PPG ions in the data was desired in order to better visualize the compositional differences between hand, face, and axillary samples independent of deodorant use. The KMF was adjusted for PPG, i.e. Kendrick mass scaling and the specification of the integer unit repeat. The parameters previously used for the filtering of PEG were applied here, i.e. ΔKMD of 0.01, RT of 0.8, and NOS of 2. The Kendrick

mass plot and MS^1 feature plot are displayed in **Figure 2.6a, b.** respectively. A large number of features meet the KMF filtering criteria, which while not surprising in this case illustrates how many MS^1 features can be linked to polymers and how such information can complicate interpretation. The KMF removed singly- and doubly-charged species (**Figure 2.S8**). The *m/z* difference and homology is preserved between the n^{th} oligomer peak equal to the charge in the oligomer distribution, for example the 2nd consecutive PPG oligomer from any signal in a doubly-charged distribution would meet the Δ*m/z* 58 filtering requirement for PPG. The accuracy of filtering for multiple charged ions was not evaluated in this work. A representative MS^1 feature spectrum from subject BF1637, right axillary, is displayed prior to KMF in **Figure 2.6c**. The MS^1 feature spectrum of retained features is shown in **Figure 2.6d**. The MS^1 feature spectrum displaying the features filtered is plotted in **Figure 2.6e** which clearly indicates the majority of PPG features in this PPG dominated example were removed.PCA was performed on the KMF data, **Figure 2.5b**. The differential grouping of axilla samples compared to skin (hands and face) is more apparent, and the PCA loadings indicate that the largest source of variance, PC_1 (5.0%), no longer is associated with PPG oligomers peaks (PCA loadings are tabulated in **Table 2.S2** and electronic supplementary information – **Table 2.S4**). PCA score plots visualizing combinations of different PCs can be found in the Supplementary Information (**Figure 2.S9**). Note, differences in the PCA score plot are anticipated when changing the number of variables. This real-world experiment exemplifies a situation in which control over the polymer content would have been problematic without control of deodorant use *a priori*, and one in which a large number of different polymer sources are possible, which makes proper controls difficult to obtain. The KMF processing of data improved interpretation and

investigation of the initial question of differentiating hands, face, and axillary samples based on the endogenous metabolome. However, it should be noted that filtering of the data to remove polymer ions does not compensate for the signal suppression caused by polymers during electrospray ionization which will influence the ions observed as well as abundance measurement.

4. KMF for exploring data compositionality

A priori knowledge of undesired chemical background is rare, and the KMF is particularly suitable in such scenarios (e.g. unknown or unanticipated source of background). In addressing the lack of a priori knowledge, we applied the KMF to explore data compositionality prior to the filtering step. Compositional analysis in this manner is intended to be only informative; the performance of evaluating the accuracy of selection was not evaluated in this study. As the combinatorial nature of the filtering criteria can generate false positives, manual inspection of selected signals is recommended.

The data compositionality of plasma sample spiked with a swab extract (n = 3), previously discussed when optimizing KMF parameters, is displayed in **Figure 2.7**. The evaluated background ions were split into the following categories: composition, containment, and source. Note, the evaluated background was not comprehensive, but users can add background signals of interest to the freely available code. The contaminant category included common background polymers and signals including perfluorinated molecules (unit repeat of CF_2), polysiloxanes, PPG (unit repeat of $C_3H_6O_1$), and PEG (unit repeat of $C_2H_4O_1$). Compositional analysis via KMF of the plasma sample spiked with a swab extract, **Figure 2.7**,

indicated the presence of a large number of PEG MS^1 features matching expectation. Few features meet the criteria for other polymers. The composition category included unit repeats with masses associated with CH_2, C_2H_4, C_3H_6, C_4H_8, and O, similar to traditional uses of Kendrick mass analysis of data. MS^1 features associated with the composition category are only informative in this instance and should not be filtered, but reveal high-level information on sample components. Similarly, the source category is largely informative in this instance, but could be valuable in understanding ionization generated signals, intentionally or inadvertently. The accuracy in determining the presence of MS^1 features belonging to the source category is beyond the scope of this publication, but the aim of future studies.

E. Conclusions

Mass spectrometry, particularly untargeted analysis, frequently encounters some degree of chemical background such as polymers. Systematic investigation of the KMF user-defined parameters, ΔKMD, RT, and NOS, was performed using a PEG 400 standard and plasma. Testing of the KMF parameters determined using the PEG 400 and plasma were applied to plasma spiked with a swab extract, illustrative of a situation in which complex undesired chemical background is present. The MS^1 features, which were not filtered in the MS^1 feature plot were highly reminiscent of those in the plasma standard MS^1 feature plot. Further, the spectrum of plasma spiked with swab extract was very similar to that of the plasma standard. The skin samples obtained from the organ transplant recipients illustrate the effects of uncontrolled polymer background on multivariate statistical analysis – the observation of which is likely amplified by the low biomass skin samples analyzed – and how the KMF can be used

to remove such interferences and clarify interpretation. The KMF was also used to explore the composition of the data acquired providing a high-level overview of MS^1 features, which display unit repeat homology without a priori knowledge. Such an analysis might serve as a first step to determine which signals are most abundant and the target of subsequent filtering.

The parameters in this instance were determined by maximizing the filtering of PEG 400 MS^1 features while minimizing NIST plasma SRM. In general, parameters should be selected based on the data and weighing desired outcomes (e.g. the extent of filtering against inappropriate removal of non-polymer ions). Ideally, the best use case requires a pilot experiment in which the desired polymer can be spiked into the matrix allowing the ideal parameters to be determined; this process is usually not feasible due to time and resource constraints. Given the constraints, we recommend that users run the filter on a few representative samples, and use the plots provided in the supplementary Jupyter notebook to adjust the parameters and inspect the filtering results. Confirmation of MS^1 features which are retained or filtered is recommended. We further recommend the use of GNPS (13) for MS^2 queries for putative metabolite identification and contribution of background MS^2 spectra into the public spectra library.

The KMF is not intended to replace appropriate study design and careful sample handling. It is intended to be used in situations in which other options, e.g. background subtraction, are not possible or ineffective. The proposed KMF can theoretically be applied to mass spectral data obtained using any mass analyzer; however, mass accuracy will influence filtering. Additionally, the method by which MS^1 feature finding is performed will influence filtering quality. We anticipate that the KMF could also be used when chemical separation is

not performed, e.g. matrix-assisted laser desorption, nanoelectrospray, and ambient ionization techniques. We envision the use of the KMF to select only polymer ions in MS^1 data, filtering all non-polymer MS^1 features, in order to study the polymer content of samples.

F. Acknowledgements

The authors would like to acknowledge the scientific support of the Collaborative Mass Spectrometry Innovation Center and the Center for Microbiome Innovation, University of California - San Diego, and funding support from the US National Institutes of Health 1R03CA211211-01 and 5P41GM103484-07.

Chapter 2, in full, is a reprint of the material as it appears in Journal of The American Society for Mass Spectrometry 2019. da Silva, R. R; Vargas, F; Ernst, M; Nguyen, N. H; Bolleddu, S; del Rosario, K. K; Tsunoda, S. M; Dorrestein, P. C; Jarmusch, A. K.; Springer US, 2018. The dissertation author was a primary investigator and author of this paper.

G. Figures and Tables

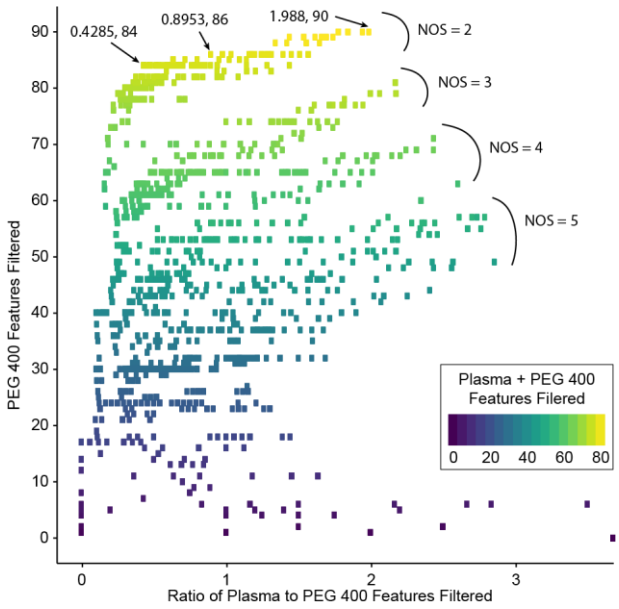


Figure 2.1. Plot displaying KMF results for all sets, represented as points, of KMF parameters (ΔKMD, RT, and NOS) tested. Data were collected at a resolution of 17,500. KMF parameters for the points highlighted are as follows: 0.4285, 84 (ΔKMD of 0.01, RT of 0.8 min, and NOS of 2); 0.8953, 86 (ΔKMD of 0.033, RT of 0.8 min, and NOS of 2); 1.988, 90 (ΔKMD of 0.1, RT of 2.0 min, and NOS of 2).

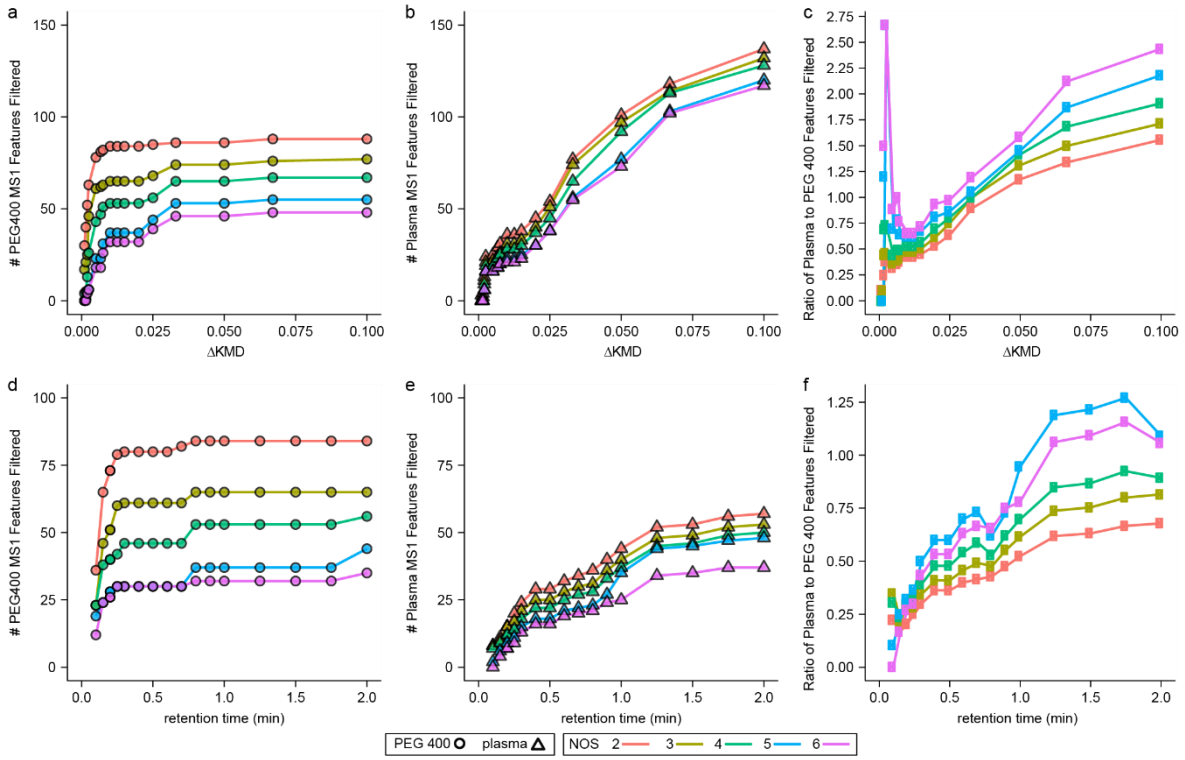


Figure 2.2. (**a**) ΔKMD versus the number of PEG 400 features filtered (circle) and (**b**) plasma features filtered (triangle). (**c**) ΔKMD versus the ratio of PEG 400 to plasma features filtered. NOS is indicated by color. RT was constant at 0.8 min. (**d**) RT versus the number of PEG 400 features filtered (circle) and (**e**) plasma features filtered (triangle). (**f**) RT versus the ratio of PEG 400 to plasma features filtered. NOS is indicated by color. ΔKMD was constant at 0.01. Data plotted were acquired at a resolution of 17,500.

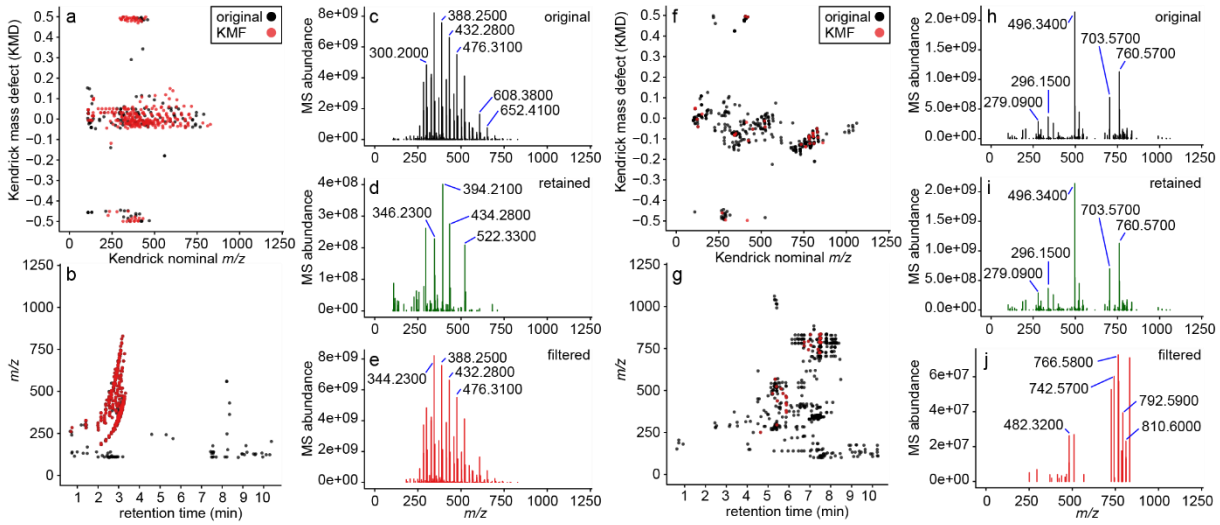


Figure 2.3. (**a**) Kendrick mass plot (original MS^1 features, black, and filtered MS^1 features, red) of PEG 400. (**b**) MS^1 feature plot of PEG 400, original MS^1 features (black), and MS^1 features filtered (red). (**c**) MS^1 feature spectrum for PEG 400 prior to KMF. (**d**) MS^1 feature spectrum of retained features (not filtered) and (**e**) MS^1 feature spectrum of MS^1 features removed via KMF. (**f**) Kendrick mass plot (original MS^1 features, black, and filtered MS^1 features, red) of plasma. (**g**) MS^1 feature plot of plasma original MS^1 features (black) and MS^1 features filtered (red). (**h**) Illustrative plasma MS^1 feature spectrum prior to KMF; (**i**) plasma MS^1 feature spectrum of MS^1 features retained; and (**j**) MS^1 feature spectrum of MS^1 features removed via KMF. KMF parameters: $\Delta KMD = 0.01$, RT = 0.8 min, and NOS = 2. PEG was filtered using the ethylene oxide unit repeat (*m/z* 44.0262). Data shown were acquired at a resolution of 17,500.

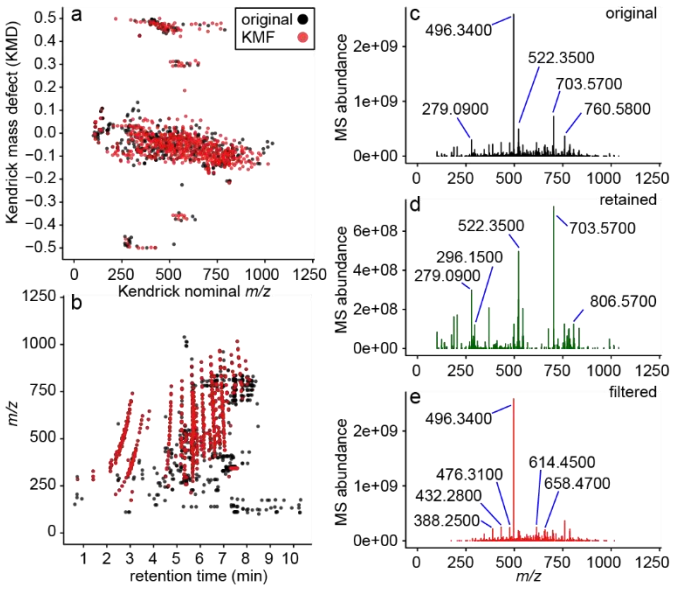


Figure 2.4. (**a**) Kendrick mass plot (original MS^1 features, black, and filtered MS^1 features, red) of plasma spiked with swab extract filtering. (**b**) MS^1 feature plot of plasma spiked with swab extract, original MS^1 features (black) and MS^1 features filtered (red). (**c**) Plasma spiked with swab extract prior to KMF, (**d**) MS^1 feature spectrum of features retained, and (e) MS^1 features removed via KMF. KMF parameters: ΔKMD = 0.01, RT = 0.8 min, and NOS = 2. PEG was filtered using the ethylene oxide unit repeat (*m/z* 44.0262). Data shown were acquired at a resolution of 17,500.

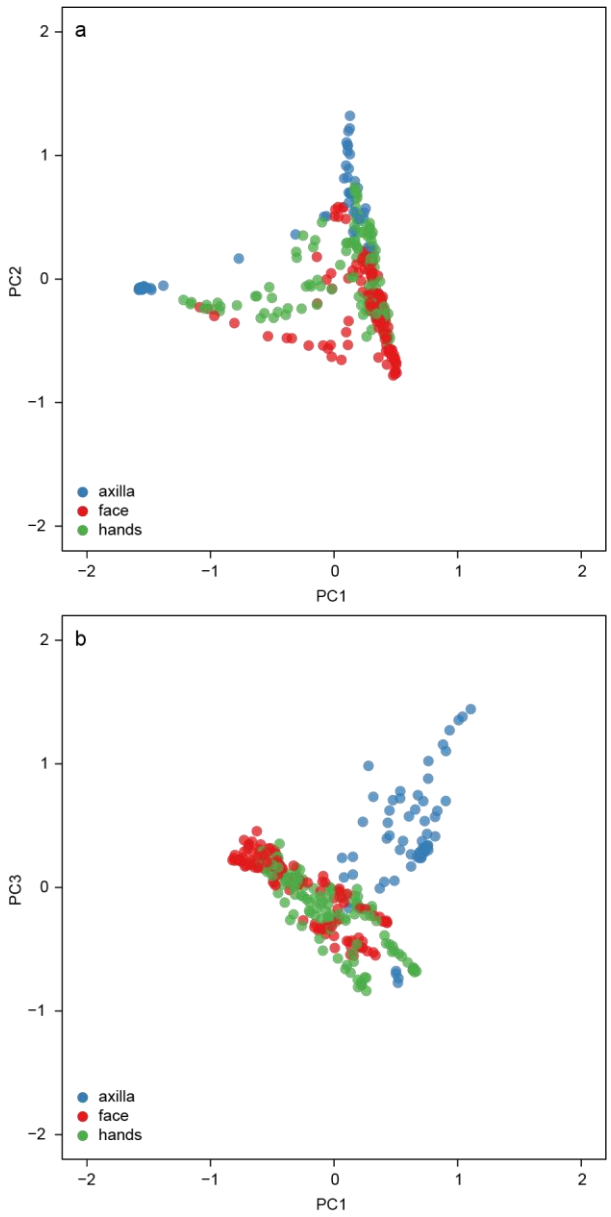


Figure 2.5. (**a**) PCA score plot of original data obtained from the skin swab samples, hands (green), face (red), and axilla (blue), of organ transplant recipients, 15 subjects. (**b**) PCA score plot of KMF data obtained from the skin swab samples, hands (green), face (red), and axilla (blue), of organ transplant recipients, 15 subjects. KMF was performed for PPG. KMF parameters: ΔKMD = 0.01,RT = 0.8 min, and NOS = 2.

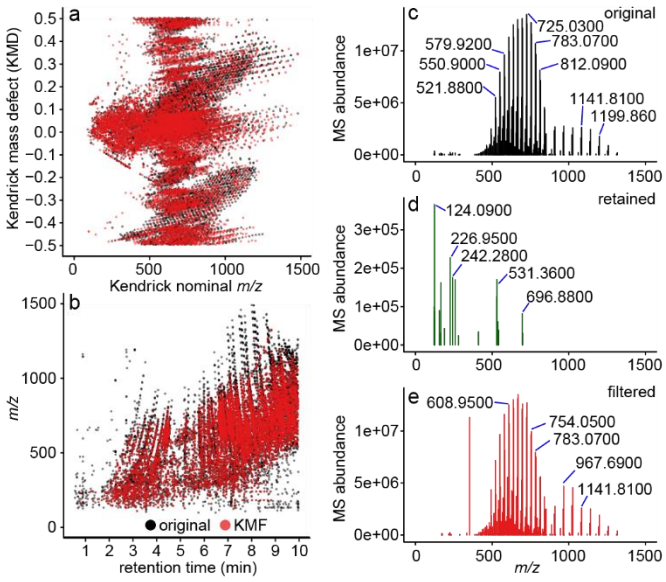


Figure 2.6. (**a**) Kendrick mass plot (original MS^1 features, black, and filtered MS^1 features, red) of skin swab samples of organ transplant recipients for PPG (original feature, black, and filtered features, red). (**b**) MS^1 feature plot, original MS^1 features (black) and MS^1 features filtered (red). (**c**) Illustrative MS^1 feature spectrum of an axilla sample from subject BF1637, right axillary, believed to contain PPG due to deodorant use. (**d**) MS^1 feature spectrum of features retained after KMF and (**e**) MS^1 feature spectrum of MS^1 features removed via the KMF. KMF parameters: ΔKMD = 0.01, RT = 0.8 min, and NOS = 2. PPG was filtered using the propylene oxide unit repeat (*m/z* 58.0419).

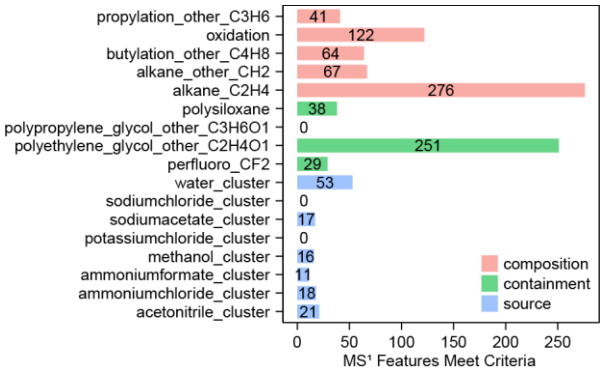


Figure 2.7. Compositional analysis. The MS1 features which meet KMF criteria associated with the composition, (red) containment, (green) and source (blue) categories are plotted.

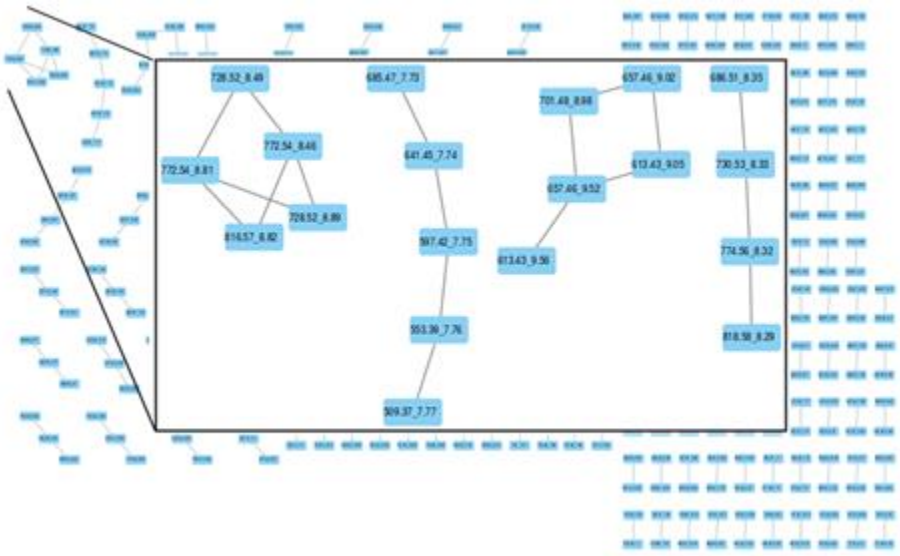


Figure 2.S1. (**a**) Putative polymer units filtered with KMD and RT window highlighting masses of possibly multiple units eluting in a small RT window.

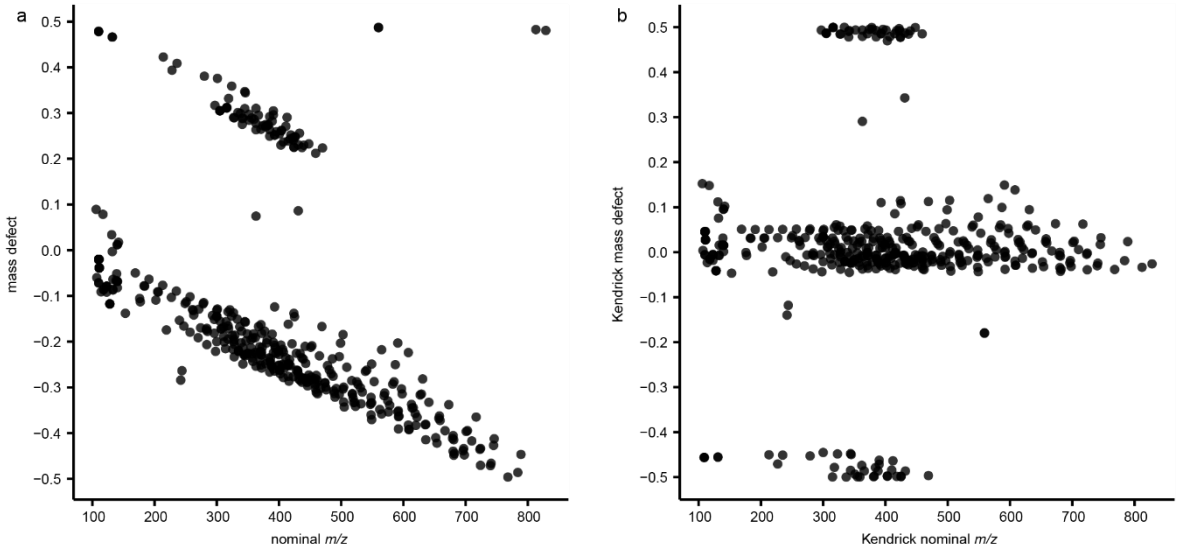


Figure 2.S2. (**a**) Putative Mass defect plot of PEG 400 data and (**b**) Kendrick mass plot of PEG 400 data. Data collected at 17,500 resolution.

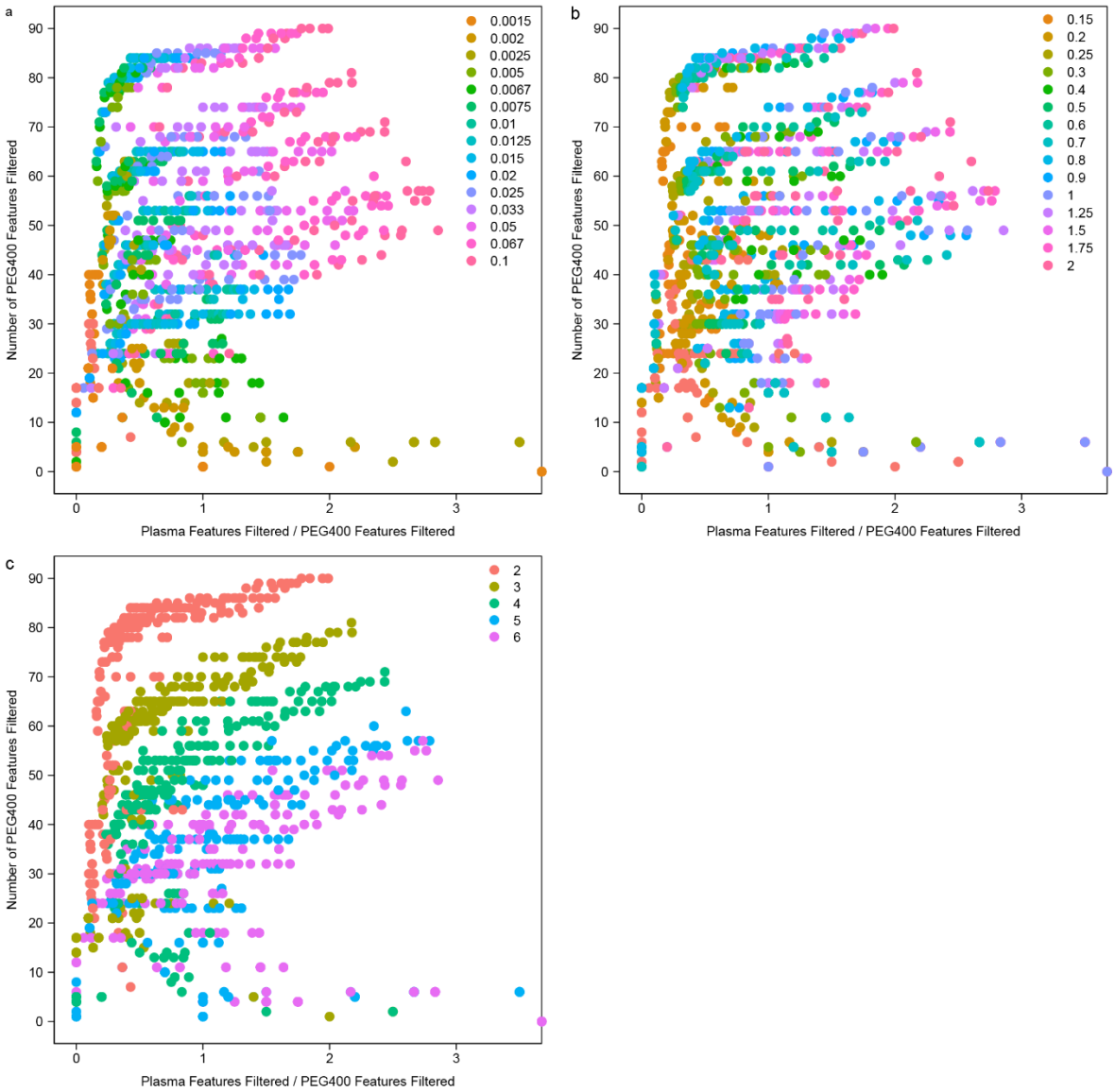


Figure 2.S3. Plot displaying KMF results for all sets, represented as points, of KMF parameters (ΔKMD, RT, and NOS) tested colored by (a) ΔKMD, (b) RT, and (c) NOS. Data were collected at a resolution of 17,500.

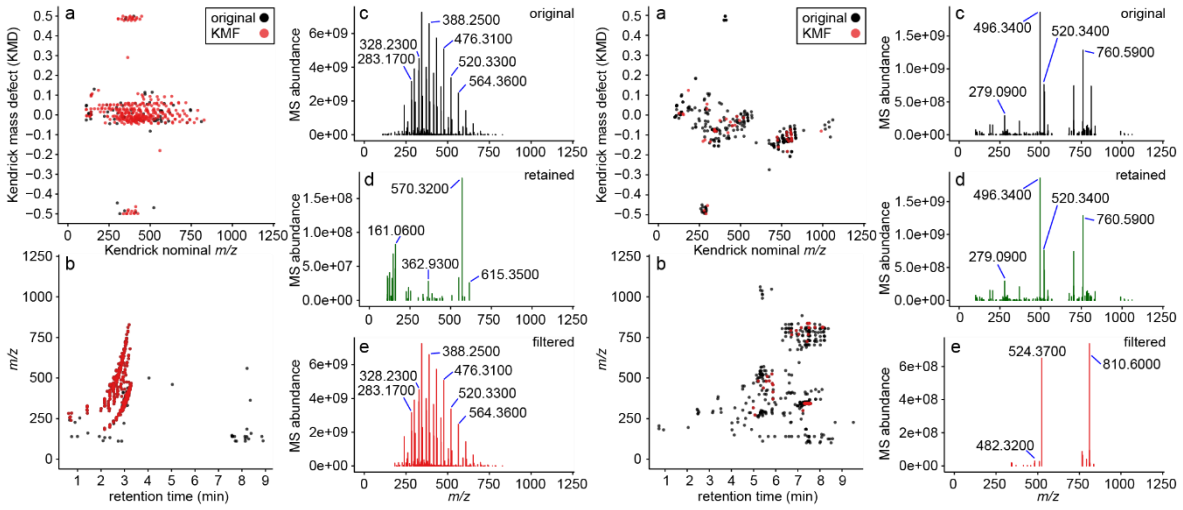


Figure 2.S4. (**a**) Kendrick mass plot (original MS^1 features, black, and filtered MS^1 features, red) of PEG 400. (**b**) MS^1 feature plot of PEG 400, original MS^1 features (black) and MS^1 features filtered (red). (**c**) MS^1 feature spectrum for PEG 400 prior to KMF. (**d**) MS^1 feature spectrum of retained features (not filtered), and (**e**) MS^1 feature spectrum of MS^1 features removed via KMF. (**f**) Kendrick mass plot (original MS^1 features, black, and filtered MS^1 features, red) of plasma. (**g**) MS^1 feature plot of plasma original MS^1 features (black) and MS^1 features filtered (red). (**h**) Illustrative plasma MS^1 feature spectrum prior to KMF; (**i**) plasma MS^1 feature spectrum of MS^1 features retained; and (**j**) MS^1 feature spectrum of MS^1 features removed via KMF. KMF parameters: KMD = 0.01, RT = 0.8 min, and NOS = 2. Data plotted were acquired at a resolution of 70,000.

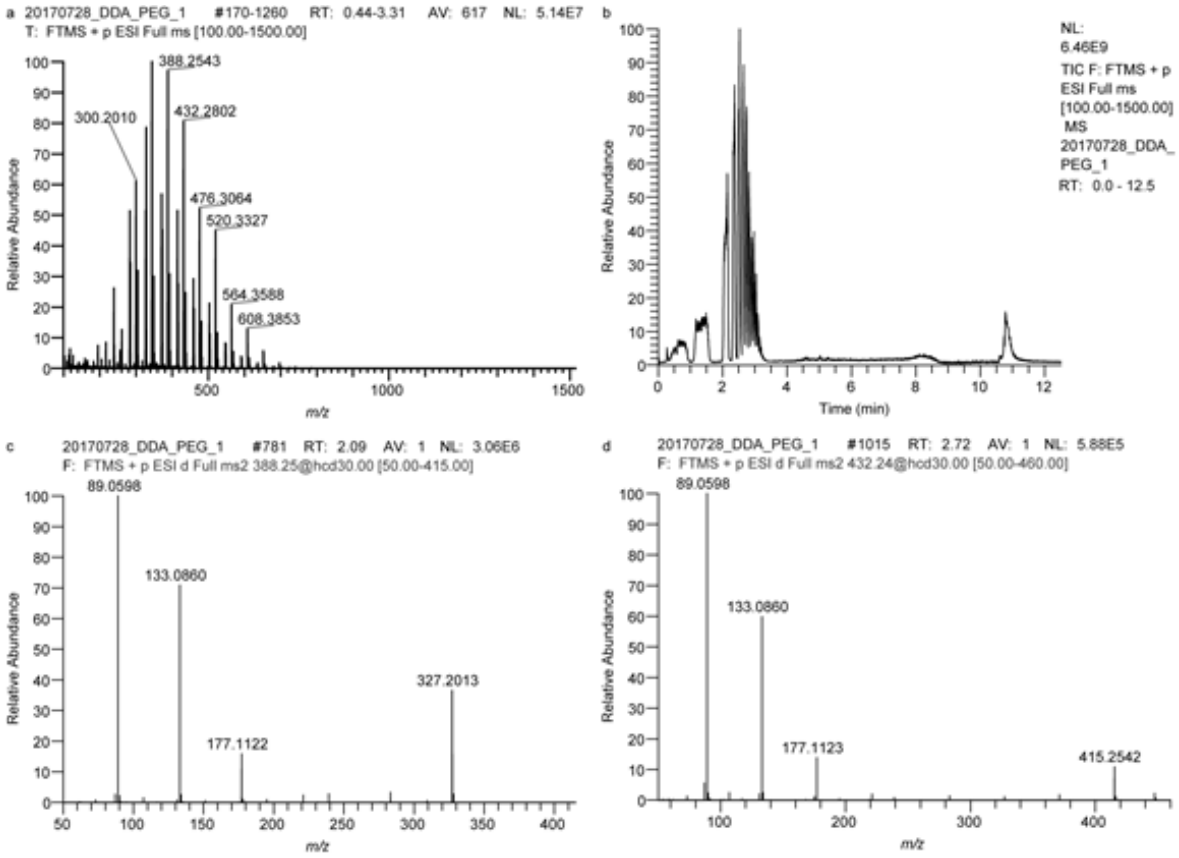


Figure 2.S5. (**a**) Average MS^1 spectrum (0.44 - 3.31 min) of a PEG 400 standard. (**b**) Total Ion Chromatogram of a PEG 400 standard. Illustrative MS^2 spectrum of PEG oligomers (**c**) *m/z* 388.25 and (**d**) *m/z* 432.24.

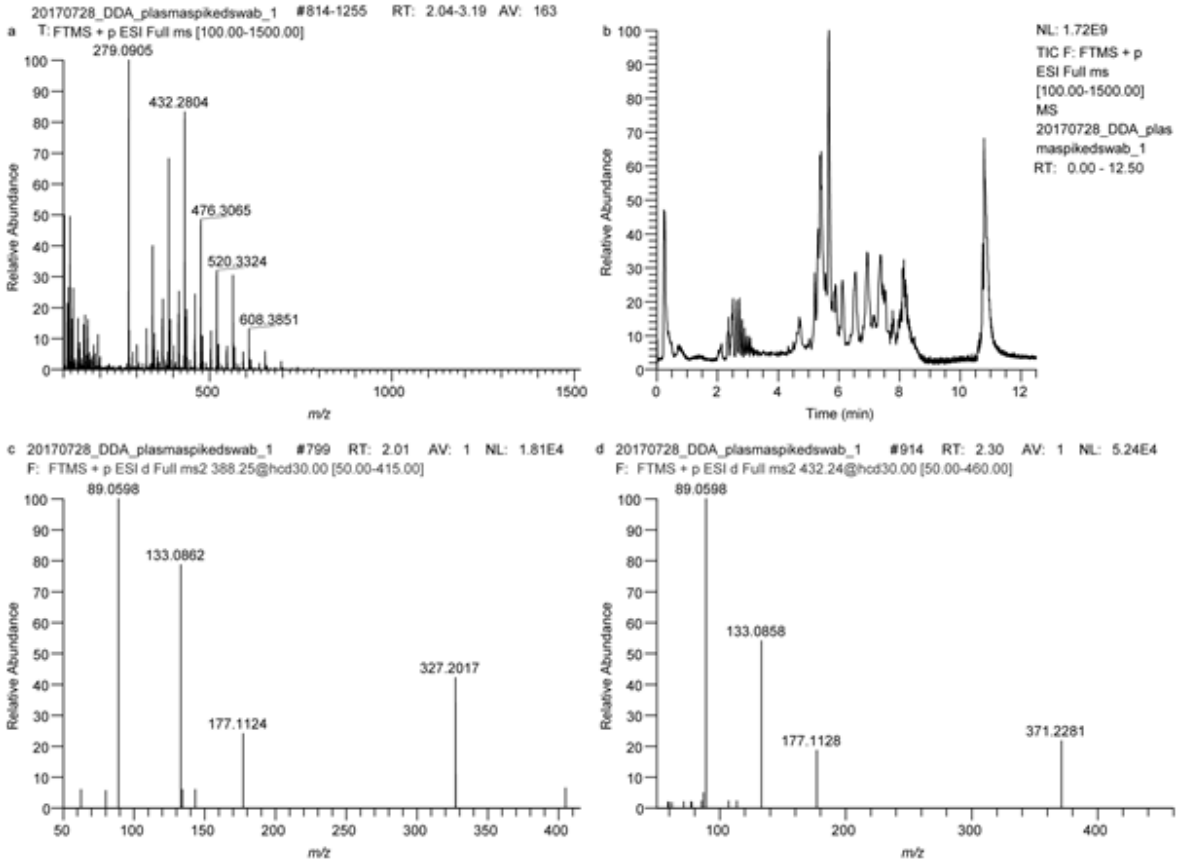


Figure 2.S6. (**a**) Average MS^1 spectrum (2.44 - 3.19 min) of plasma spiked with swab extract. (**b**) Total Ion Chromatogram of plasma spiked with swab extract. MS^2 spectrum of PEG oligomers (**c**) *m/z* 388.25 and (**d**) *m/z* 432.24.

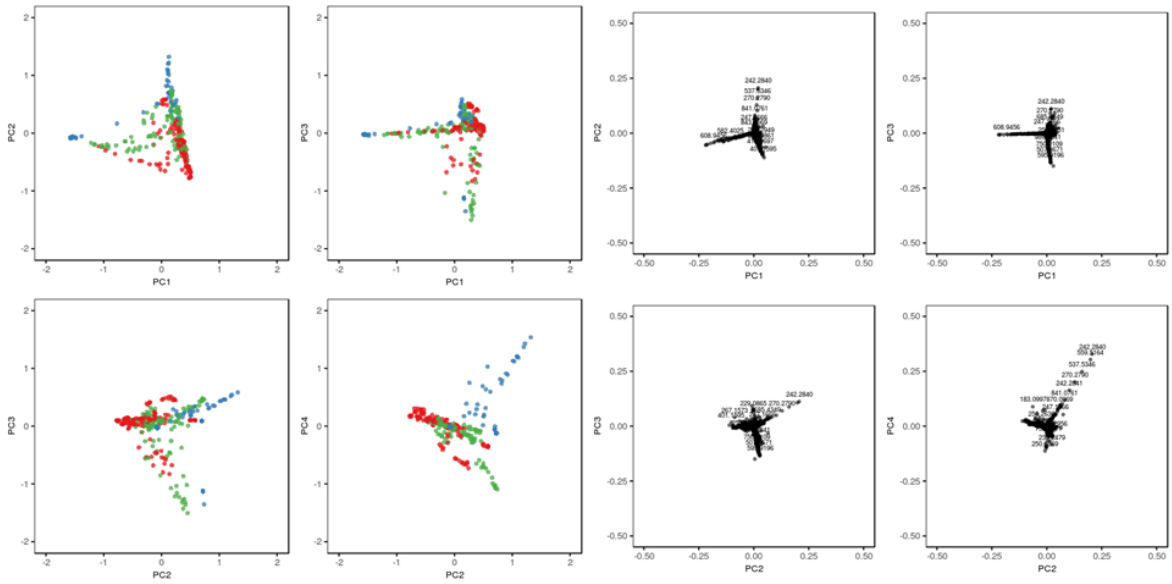


Figure 2.S7. PCA score and loading plots of original data obtained from the skin swab samples of organ transplant recipients, 15 subjects (axillary, black; face, green; and hands, red.

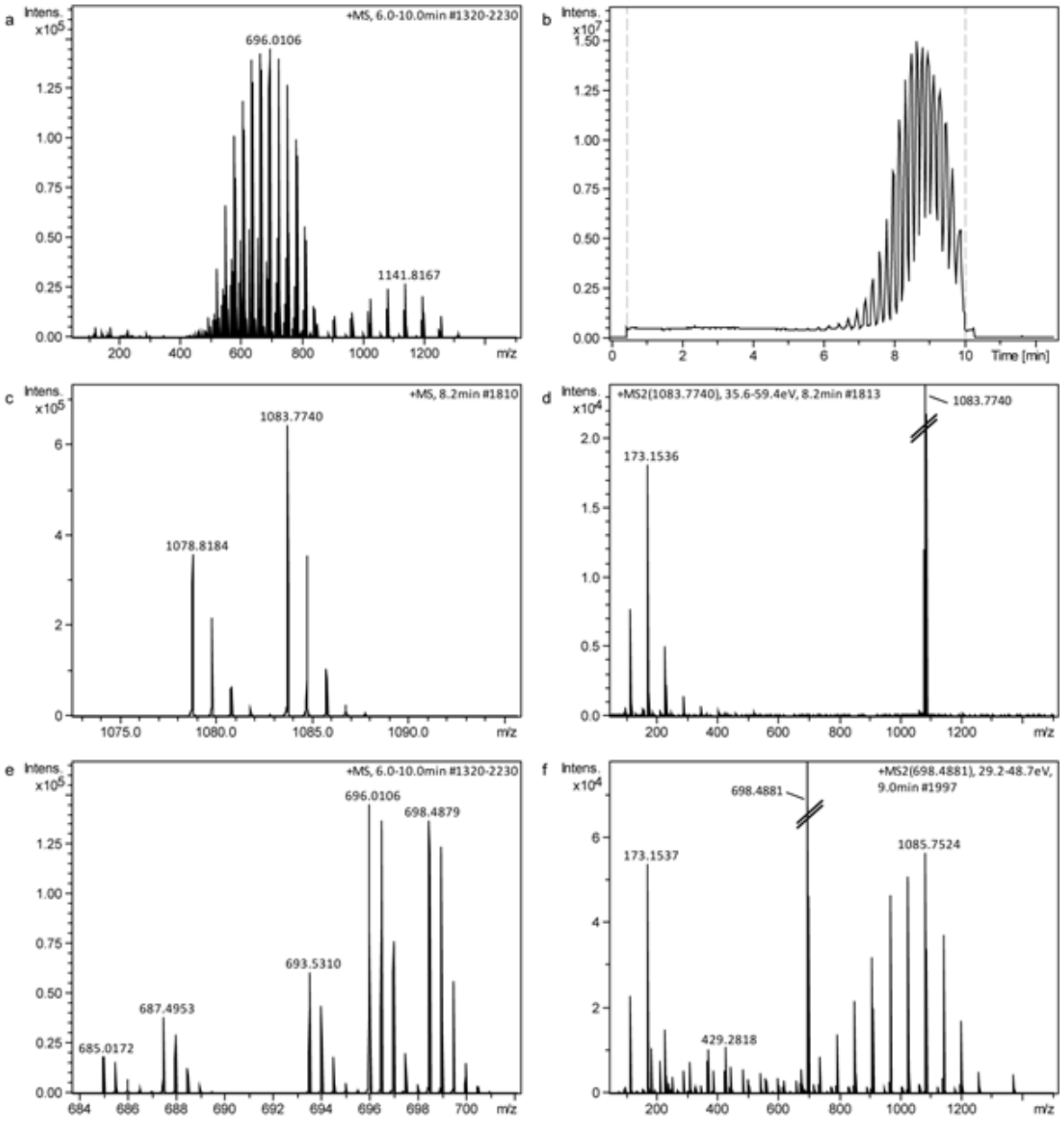


Figure 2.S8. (**a**) Average MS^1 spectrum (6.0 - 10.0 min) of the right axillary sample of subject BF1637. (**b**) Total Ion Chromatogram of the right axillary sample of subject BF1637. Note, dotted lines indicate times at which column eluent was split to waste. (**c**) Illustrative ion in MS^1 spectrum of a singly-charged PPG oligomer and (**d**) the complementary MS^2 spectrum obtained from *m/z* 1083.7740. (**e**) A zoomed in region of the average MS^1 spectrum in which a number of doubly-charged PPG oligomers are observed. (**f**) The MS^2 spectrum of *m/z* 698.4881 is displayed.

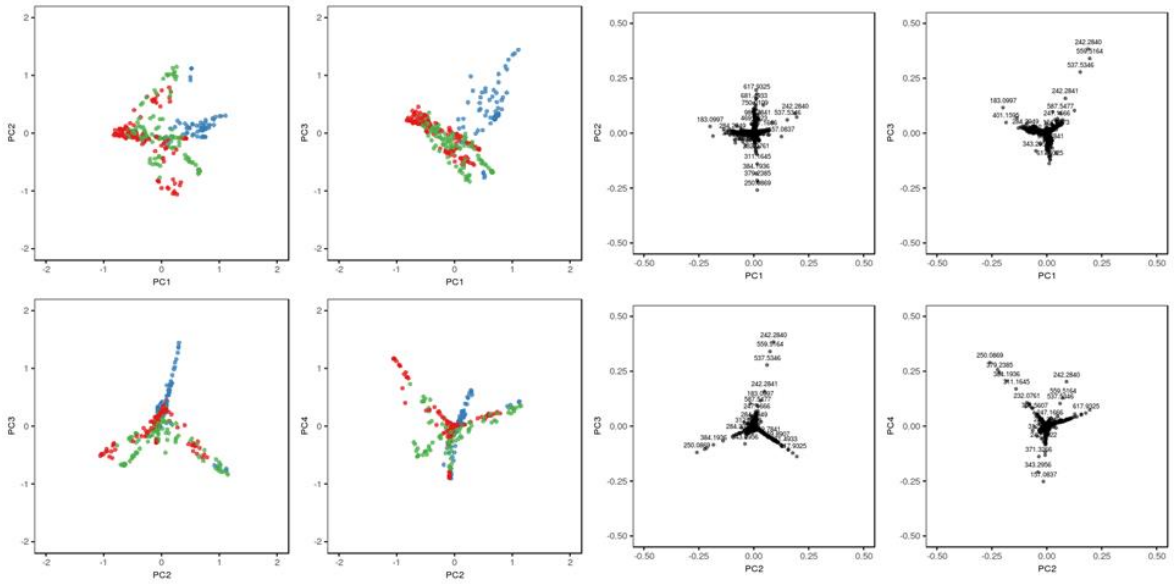


Figure 2.S9. PCA score and loading plots obtained from the skin swab samples of organ transplant recipients, 15 subjects (axillary, blue; face, green; and hands, red). KMF was performed for PPG. KMF parameters: KMD = 0.01, RT = 0.8 min, and NOS = 2.

Table 2.S1. PCA variances by principal component for original skin metabolomics data

Principal Component	Variance (%)	Cumulative Variance (%)
1	8.576668	8.576668
2	4.766919	13.343587
3	4.035778	17.379366
4	3.656365	21.035731
5	3.088650	24.124381

Table 2.S2. PCA variances by principal component for KMF skin metabolomics data

Principal Component	Variance (%)	Cumulative Variance (%)
1	5.026975	5.026975
2	3.812778	8.839753
3	3.772599	12.612352
4	3.170089	15.782441
5	3.071120	18.853560

H. References

1. Kendrick E; A Mass Scale Based on CH_2= 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Anal. Chemistry* **35**, 2146–2154 (1963).

2. Hughey C A, Hendrickson CL, Rodgers RP, Marshall AG, Qian; Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem*. **73**, 4676–4681 (2001).

3. Kim S, Kramer RW, Hatcher PG; Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Anal. Chem.* **75**, 5336–5344 (2003).

4. Fouquet T & Sato H; Extension of the Kendrick Mass Defect Analysis of Homopolymers to Low Resolution and High Mass Range Mass Spectra Using Fractional Base Units. *Anal. Chem*. **89**, 2682–2686 (2017).

5. Sleno L; The use of mass defect in modern mass spectrometry. *J. Mass Spectrom* **47**, 226–236 (2012).

6. Zhang H, Zhang D, Ray K, Zhu M; Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry. *J. Mass Spectrom*. **44**, 999–1016 (2009).

7. McMillan A, Renaud JB, Gloor GB, Reid G, Sumarah MW; Post-acquisition filtering of salt cluster artefacts for LC-MS based human metabolomic studies. *J. Cheminform*. **8**, 44 (2016).

8. Paguigan ND, El-Elimat T, Kao D, Raja HA, Pearce CJ, Oberlies NH; Enhanced dereplication of fungal cultures via use of mass defect filtering. *J. Antibiot*. **70**, 553–561 (2017).

9. Simón-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, Mallard WG, Bearden DW, Schock TB, Tchekhovskoi DV, Blonder N, Yan X, Liang Y, Zheng Y, Wallace WE, Neta P, Phinney KW, Remaley AT, Stein SE; Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. *Anal. Chem*. **85**, 11725–11731 (2013).

10. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J,

Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P; A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol*. **30**, 918–920 (2012).

11. Pluskal T, Castillo S, Villar-Briones A, Oresic M; MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).

12. Stacklies W, Redestig H, Scholz M, Walther D, Selbig J; pcaMethods a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**, 1164–1167 (2007).

13. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N; Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol*. **34**, 828–837 (2016).

Chapter III

Protocol for Community-created Public MS/MS Reference Library Within the GNPS Infrastructure

A. Abstract

1. Rationale

A major hurdle in identifying chemicals in mass spectrometry experiments is the availability of MS/MS reference spectra in public databases. Currently, scientists purchase databases or use public databases such as Global Natural Product Social Molecular Networking (GNPS). The MSMS-Chooser workflow empowers the creation of MS/MS reference spectra directly in the GNPS infrastructure.

2. Methods

The MSMS-Chooser Sample Template was used to programmatically generate the MSMS-Chooser Submission File and Sequence Table for data acquisition. Standards from the Mass Spectrometry Metabolite Library (MSMLS) suspended in a methanol-water (1:1) solution were analyzed. Flow injection on a LC-MS/MS system was used to generate negative and positive mode data using data-dependent acquisition. The MS/MS spectra and Submission File were uploaded to MSMS-Chooser workflow in GNPS for automatic selection of MS/MS spectra.

3. Results

Data acquisition and processing required ~2 hours and ~2 min respectively per 96-well plate using MSMS-Chooser. Analysis of the MSMLS, over 600 small molecules, using MSMS-Chooser added 889 spectra (including multiple adducts) to the public library in GNPS.

Manual validation of one plate indicated accurate selection of MS/MS scans (99.2%). The MSMS-Chooser output includes a table formatted for inclusion in the GNPS library as well as the ability to directly launch searches via MASST.

4. Conclusions

MSMS-Chooser enables rapid data acquisition, data analysis (selection of MS/MS spectra), and a formatted table for inspection and upload to GNPS. Open file-format data (.mzML or .mzXML) from most mass spectrometry platforms containing MS/MS can be processed using MSMS-Chooser. MSMS-Chooser democratizes the creation of MS/MS reference spectra in GNPS which will improve annotation and strengthen the tools which use the annotation information.

B. Introduction

Identifying the chemicals detected in untargeted metabolomics experiments is challenging. Mass spectrometry (MS) is one method by which untargeted metabolomics can be performed. MS detects the mass of ions, *i.e.* mass-to-charge (*m/z*), as well as their abundance. Tandem MS (*aka* MS/MS), provides information about the molecular structure of chemicals. MS/MS, specifically product ion scans, contain interpretable patterns of product ions from which structural information is derived. Manual interpretation of MS/MS spectra is impracticable given the large amount of data generated in untargeted metabolomics experiments.

One solution to the identification challenge is to match MS/MS spectra with that of MS/MS reference spectra to provide an annotation. MS/MS matches to reference library spectra are considered level 2 or level 3 annotations based on the Metabolomics Standards Initiative (MSI) (1). Reference libraries are time-consuming and costly to generate. Laboratories develop in-house databases, but their use to the broader community is often difficult due to differences in file formats, metadata, etc. Curatr, a web-based application for library generation, is one solution that improves upon the manual interpretation of data by providing a graphical user interface (2). The Global Natural Products Social Molecular Networking (GNPS) platform (3) offers a community-built MS/MS reference library which is part of the suite of libraries available for use, *e.g.* MassBank (Japan, European Union, and North America) (4,5), ReSpect (6), MIADB, CASMI (7), PNNL lipids (8), Sirenas/Gates, and EMBL MCF (2). While the option to contribute MS/MS spectra to the GNPS spectral library has existed since its inception, it remains a time-consuming process to select reference MS/MS and provide the required information for association with the entry (*e.g.* InChI, SMILES, CAS number, adduct, charge, and instrument).

To address the limitations in building public MS/MS spectral libraries, we have created a workflow incorporated within GNPS (dubbed MSMS-Chooser). MSMS-Chooser enables users to quickly generate MS/MS reference spectra (specifically product ion spectra) from chemical standards using flow injection. Guidelines for sample preparation and MSMS-Chooser use (data formatting, the MSMS-Chooser submission file, etc) is detailed in online documentation. To illustrate the utility of MSMS-Chooser, we generated and made public 889

MS/MS spectra from chemical standards in the Sigma-Aldrich Mass Spectrometry Metabolite Library (MSMLS) kit quickly, accurately, and with minimal manual data processing.

C. Materials and Methods

1. Chemical Standards

Chemical standards were purchased from Sigma-Aldrich (SKU: MSMLS-1EA) to test the MSMS-Chooser workflow (**Figure 3.1a**). The Mass Spectrometry Metabolite Library (MSMLS) kit contained over 600 unique small molecules in 96-well plates (5 µg per well). The library contained standards from the following classes: carboxylic acids, amino acids, biogenic amines, polyamines, nucleotides, coenzymes, vitamins, monosaccharides, disaccharides, fatty acids, lipids, steroids, and hormones. A spreadsheet with information such as compound name, molecular formula, and InChI (used to represent the chemical structure) was included with the reference library and used to complete the manually entry of the MSMS-Chooser Sample Template.

2. MSMS-Chooser Sample Template

The MSMS-Chooser Google Sheet can be accessed via a hyperlink on the online documentation page (https://ccms-ucsd.github.io/GNPSDocumentation/msmschooser/). Users are required to make a copy of the MSMS-Chooser Google Sheet for each 96-well plate within Google Sheets. The MSMS-Chooser Google Sheet (**Figure 3.1a**) has multiple tabs, including README, Sample Template, Submission File, and Sequence Table. The MSMS-Chooser

Sample Template tab was manually filled using the information provided with the chemical standards that were purchased. The information input in the MSMS-Chooser Sample Template tab is used to automatically fill the remaining tabs which are used in the MSMS-Chooser workflow.

3. Sample Preparation and Data Acquisition

Chemical standards were resuspended to a final concentration of 1 µM using a resuspension solvent of methanol-water (1:1) and transferred to a new 96-well plate (**Figure 3.1a**). An ultra-high performance liquid chromatography (UHPLC) system (Vanquish, Thermo) coupled to an orbitrap mass spectrometer (QExactive, Thermo) was used (**Figure 3.1b**). A sequence table ("QE_Sequence_Generator" tab) compatible with the Vanquish-QExactive system used was programmatically generated from the MSMS-Chooser Sample Template. The data acquisition time was set to 0.3 min for each injection. Separate positive and negative ionization mode methods were used (*i.e.* one injection in positive mode and one injection in negative mode); the major differences in the methods are the ionization source parameters. 5 µL was injected from each well into a flow of solvent. The flow rate was set to 0.500 mL min^{-1} with equal parts 0.1% formic acid in water and 0.1% formic acid in methanol. More detailed information on the instrument method can be found in the MSMS-Chooser documentation (https://ccms-ucsd.github.io/GNPSDocumentation/msmschooser/) and the QExactive tune files and instrument methods for positive and negative mode can be found on MassIVE via the following accession number: MSV000084286

4. Data Processing and Upload

Data collected in Thermo's proprietary MS file format (.raw) were converted to the open-source format (.mzXML) using the ProteoWizard tool MSConvert (9). All MS files (.raw and .mzXML files) and the MSMS-Chooser Submission File (downloaded as a .tsv from Google Sheets) were uploaded to MassIVE. To aid in dataset organization and utility it is recommended that all MSMS-Chooser datasets uploaded to MassIVE be prepended with the following text, "GNPS - MSMS-Chooser -", and made publicly available. Data generated for this study can be accessed via the following MassIVE accession numbers: MSV000084286, MSV000084072, MSV000084494, MSV000084492, MSV000084493, MSV000084479, MSV000084495, MSV000084496. Classyfire was used to generate chemical class information for the chemical standards within dataset GNPS – MSMS-Chooser- GNPS00001 by querying the SMILES (10).

5. MSMS-Chooser

The MSMS-Chooser (v1.0) workflow, **Figure 3.1c**, is accessed on GNPS via the following link: https://gnps.ucsd.edu/ProteoSAFe/index.jsp?params=%7B%22workflow%22:%22MSMS-CHOOSER%22%7D (a free GNPS account is required). The source code is available at https://github.com/CCMS-UCSD/GNPS_Workflows. The .mzXML/.mzML files (positive and negative mode) and the MSMS-Chooser Submission file were selected for each plate and processed using the MSMS-Chooser workflow (detailed instructions are available in the documentation https://ccms-ucsd.github.io/GNPSDocumentation/msmschooser/#msms-

chooser-v10). The MSMS-Chooser workflow uses the information in the MSMS-Chooser Submission file to calculate the monoisotopic mass [M] and subsequently calculate the monoisotopic *m/z* of adducts using the SMILES or InChi. The current set of designated adducts considered are $[M+H]^+$, $[2M+H]^+$, $[M+Na]^+$, $[2M+Na]^+$, $[M-H_2O+H]^+$, $[M-2(H_2O)+H]^+$, and $[M-3(H_2O)+H]^+$, $[M-H]^-$, $[2M-H]^-$, $[2M-2H+Na]^-$. MSMS-Chooser searches for the calculated *m/z* of all possible adducts within a 10 ppm mass error tolerance and outputs the associated MS/MS scan (if detected). The results of MSMS-Chooser include the filename, chemical (*i.e.* compound_name), molecular mass, scan number (*i.e.* extract scan), SMILES, adduct, charge, ionization mode (*i.e.* ion mode), and exact mass. An illustrative example of the MSMS-Chooser GNPS result page is available via the following link https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=8454490b1ecc49ab85e1cece2f2f944c. The resulting output table (*i.e.* "Output batch table") of MSMS-Chooser was checked manually for quality via visual inspection by selecting the "view spectrum" button, **Figure 3.1d**. Poor quality or incorrect spectra were deleted (row-wise) from the results table upon download. Note, the quality rating of the spectra in GNPS is based on the sample source which is input by the user in the MSMS-Chooser Template. In this instance, a gold rating ("1") was used as the data were obtained from chemical standards. The MSMS-Chooser output ("batch table") was processed using an in-house workflow that uploads the spectra to the GNPS MS/MS spectral library. Users of MSMS-Chooser will not have access to the in-house workflow for security reasons. Users are instructed to email their MSMS-Chooser result files to a dedicated team of administrators. Detailed instructions can be found in the MSMS-Chooser documentation (https://ccms-ucsd.github.io/GNPSDocumentation/msmschooser/). The resulting list of chemicals added

from this example dataset (GNPS – MSMS-Chooser- GNPS00001) is tabulated in **Table 3.S1** and the library can be accessed on GNPS via the following link: https://gnps.ucsd.edu/ProteoSAFe/gnpslibrary.jsp?library=GNPS-MSMLS.

D. Results and Discussion

Positive and negative mode MS/MS spectra were collected using two separate injections into a flow of solvent. The method was 0.3 min in length, providing enough time for the chemical standard to be detected and the system purged of any remaining material, as depicted in the chromatograms of two illustrative chemicals in positive and negative mode (**Figure 3.2a** and **Figure 3.2d**, respectively). The flow injection method was used for higher-throughput data acquisition (~2 hours for a 96-well plate), opposed to chromatographic separation.

The data obtained were processed using the MSMS-Chooser workflow in GNPS (~2 minutes per 96-well plate) to identify MS/MS spectra associated with different adducts of the chemical standards within a mass error of 10 ppm. Multiple adducts were detected for some of the chemical standards which are formed during the electrospray ionization process under the experimental conditions (*e.g.* salinity, pH, and concentration). For example, the analysis of dethiobiotin in the positive ionization mode resulted in the detection of the protonated ($[M+H]^+$), sodiated ($[M+Na]^+$), and singly dehydrated species ($[M-H_2O+H]^+$), **Figure 3.2b**, and their corresponding MS/MS spectra selected by MSMS-Chooser (**Figure 3.2c**). Similarly, *N*-acetylleucine was detected as a deprotonated ($[M-H]^-$) ion and as a $[2M-2H+Na]^-$ ion in the negative ionization mode (**Figure 3.2d**), and MSMS-Chooser selected the corresponding MS/MS spectra (**Figure 3.2f**). Additional examples of spectra are displayed in **Figure 3.S2**. All

spectra resulting from the analysis of standards in the MSMLS are available on GNPS (https://gnps.ucsd.edu/ProteoSAFe/gnpslibrary.jsp?library=GNPS-MSMLS).

We evaluated the performance of MSMS-Chooser on its ability to correctly select MS/MS spectra in a subset of the MSMLS (GNPS – MSMS-Chooser- GNPS00001, randomly selected). GNPS00001 contained 87 chemical standards of which only 63 were detected in our experiment. Multiple hours of manual data analysis found 73 MS/MS scans in positive mode and 53 MS/MS scans in negative mode (**Table 3.1**) corresponding to the 63 chemicals detected. MSMS-Chooser was able to correctly identify a total of 70 MS/MS scans in positive mode and 50 MS/MS scans in negative mode in minutes without user input. MSMS-Chooser correctly selected 99.2% of the MS/MS spectra, excluding five MS/MS spectra from this calculation as they exceeded the 10 ppm mass error tolerance. The adducts detected from this illustrative set of chemical standards (*i.e.* GNPS00001) were tabulated and plotted in **Figure 3.3a**. The predominant adducts in positive mode were $[M+H]^+$ and $[M+Na]^+$; and $[M-H]^-$ in negative mode. The high levels of sodiated adducts observed, **Figure 3.3B**, is likely related to the number of sugars (*i.e.* Organooxygen compounds) included in the test dataset.

The MSMS-Chooser output provides users the ability to query their MS/MS spectra against all publicly available tandem MS datasets in GNPS via the MASST tool. The MASST results indicated the datasets in which the query spectra had been previously detected. **Figure 3.S1** depicts the MASST result of biotin, one of the chemical standards run in the MSMS-Chooser proof of concept dataset (GNPS00001). Biotin was shown to be found in 16 GNPS datasets, comprised of microbial, humans, and food sample types. The illustrative MASST job

can be accessed using the following link: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=39cd886540c147e9a8a618b275e5f541.

E. Conclusions

MSMS-Chooser is a web-based data processing platform and instrument protocols for the rapid and systematic generation of MS/MS reference spectra. The aim of this workflow is high-throughput analysis with minimal user input, rather than a comprehensive coverage of all chemicals and all possible adducts. In our proof-of-concept results, chemicals from the MSMLS were analyzed resulting in 889 new MS/MS reference library spectra in GNPS. MS data was collected in approximately two hours per 96-well plate and processed using MSMS-Chooser in approximately two minutes per plate. The completion of the MSMS-Chooser Sample Template is the only manual step in the data processing, not including review of the MSMS-Chooser results. MSMS-Chooser was 99.2% accurate in selecting MS/MS spectra compared to manual evaluation (inspection of results is recommended). The highly automated selection of MSMS-Chooser, even with manual inspection, may erroneously introduce poor quality or incorrectly identified spectra (true for manually curated data as well). However, GNPS provides users with the ability to modify the information included with all reference spectra to address this inevitability.

One strength of the workflow is the consideration of multiple adducts without user input. Chemical have the varying ionization efficiency and preferentially form different adducts (*e.g.* protonation or sodiation). MSMS-Chooser assists in building MS/MS spectral libraries by considering multiple adducts that may be encountered using different electrospray ionization

conditions. We demonstrate results that compounds of particular chemical classes are observed to preferably form adducts, such as sugars (*i.e.* Organooxygen compounds) forming sodiated adducts in the positive ionization mode.

We have provided the method files used to generate data on our Vanquish-QExactive LC-MS platform as a matter of transparency as well as a reference for others to generate flow injection methods using their own MS instrumentation. In principal, MS/MS data generated using any MS instrumentation via direct injection, flow injection, or chromatographic elution are all compatible with the MSMS-Chooser workflow. Caution should be used when using this approach with anything but flow injection as it was not its original intent. Further, MS/MS data which results from any method of ion activation (collision-induced dissociation, higher-energy collision-induced dissociation, surface-induced dissociation, electron capture dissociation, electron transfer dissociation, infrared multiphoton dissociation, ultraviolet photodissociation, etc) is accepted in the GNPS spectral library. Currently, MSMS-Chooser does not capture collision energy but it is possible to do so; however, such information is not currently utilized in GNPS and therefore has limited utility. We strongly believe the variability in MS/MS spectra is best captured via the contribution of spectra of the same chemical collected on multiple instruments and multiple conditions. Contribution of all MS/MS spectra is highly encouraged.

MSMS-Chooser will aid the community of scientists who use GNPS (or the associated GNPS knowledge base) by increasing the number of MS/MS reference spectra in the public domain, thereby improving annotation rates in untargeted metabolomics experiments. Any additional MS/MS spectra uploaded to the GNPS MS/MS spectral library will be included in the periodic living data analysis in GNPS. Data submitters will be emailed when new chemical

annotations are found in their MassIVE deposited data. We have integrated MASST in the results of MSMS-Chooser allowing users to query the GNPS knowledgebase and determine where a particular MS/MS spectrum has been previously observed, illustrated by our search of the MS/MS spectrum obtained from biotin (11). Further, additional MS/MS library spectra will extend the depth of annotation in ReDU (12), a recently developed web-enabled tool to reuse public MS/MS data. Lastly, the providence of the original data and MSMS-Chooser Sample Template is retained in MassIVE and available for re-analysis in a systematic and automated fashion.

F. Acknowledgements

The authors would like to thank Julia M. Gauglitz, Ph.D. and Daniel Petras Ph.D. for insightful discussion. AKJ thanks the American Society for Mass Spectrometry for the Postdoctoral Career Development Award. The authors acknowledge funding by the Office of Naval Research Multidisciplinary University Research Initiative Award, (No. N00014-15-1-2809) and the Center for Microbiome Innovation. MC and PCD were supported by NSF grant IOS-1656475.

Chapter 3, in full, has been submitted for publication of the material as it may appear in Rapid Communications in Mass Spectrometry, 2019, Vargas, F; Weldon, K. C; Sikora, N; Wang, M; Zhang, Z; Gentry, E. C; Panitchpakdi, M. W; Caraballo, M; Dorrestein, P. C; Jarmusch, A. K.; John Wiley & Sons, 2019. The dissertation author was the primary investigator and author of this paper.

G. Figures and Tables

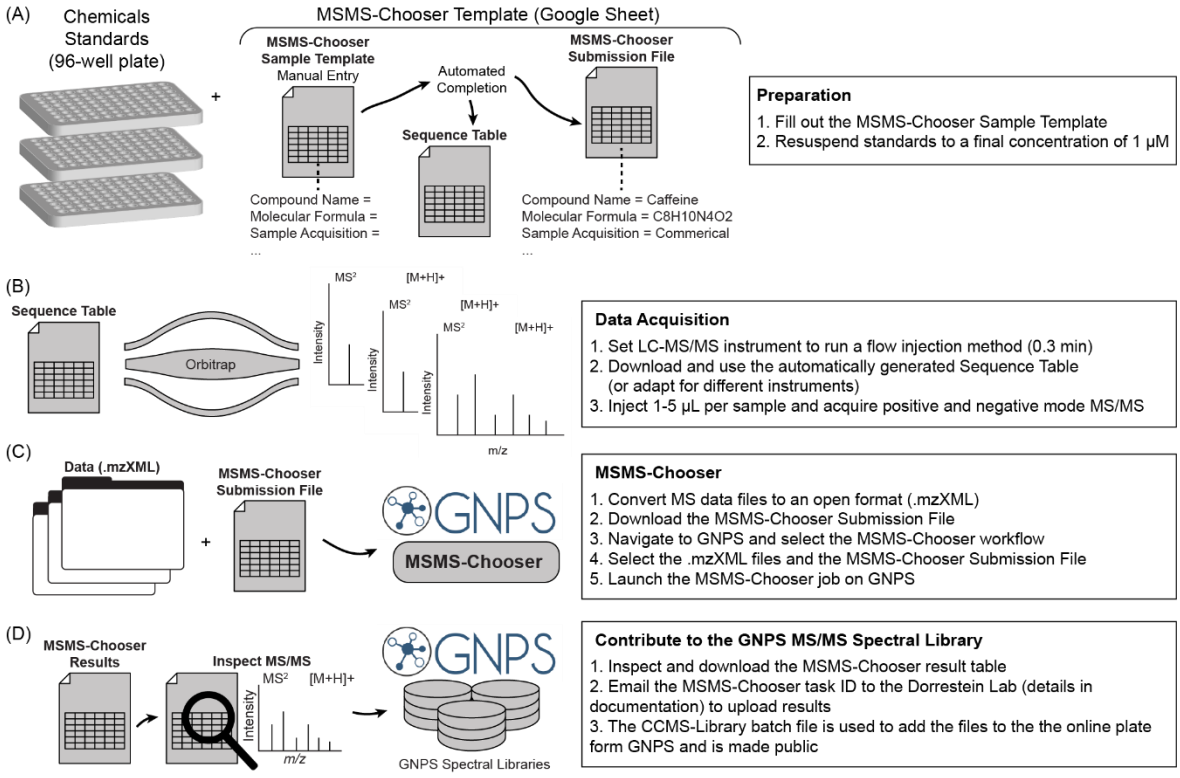


Figure 3.1. Overview of the MSMS-Chooser workflow and protocol.

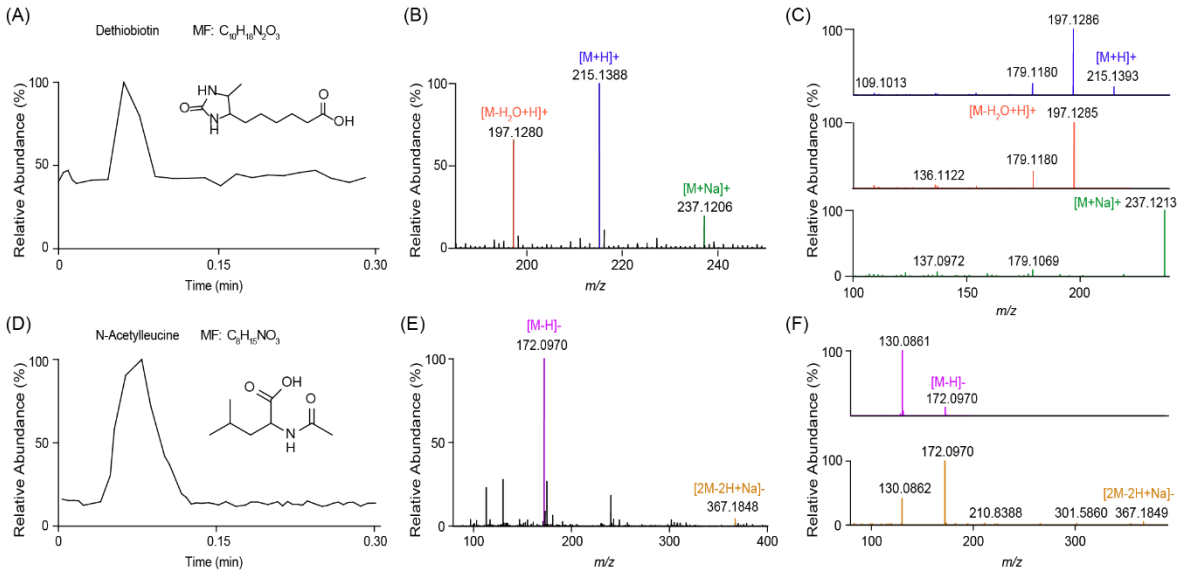


Figure 3.2. (**a**) Illustrative base peak chromatogram, (**b**) average MS spectrum, and (**c**) MS/MS spectra selected by MSMS -Chooser for dethiobiotin in the positive ionization mode. Peaks in the MS are colored corresponding to their MS/MS spectra. (**d**) Illustrative base peak chromatogram, (**e**) average MS spectrum, and (**f**) MS/MS spectra selected by MSMS-Chooser for *N*-acetylleucine in the negative ionization mode. Peaks in the MS are colored corresponding to their MS/MS spectra.

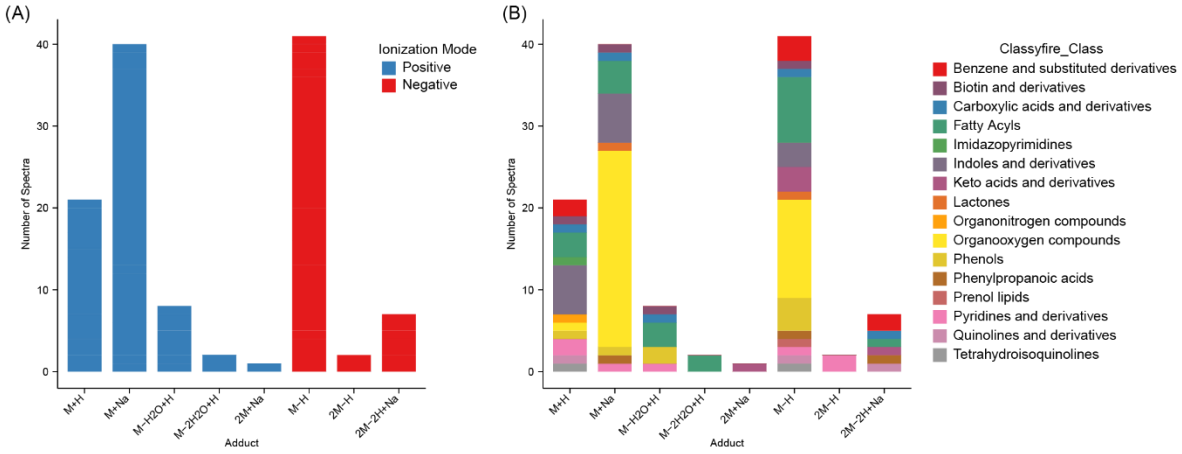


Figure 3.3. Adducts versus the number of MS/MS spectra selected by MSMS-Chooser in the proof-of-concept set of chemical standards, (**a**) colored by ionization mode and (**b**) colored by chemical classification, class level, obtained via Classyfire.

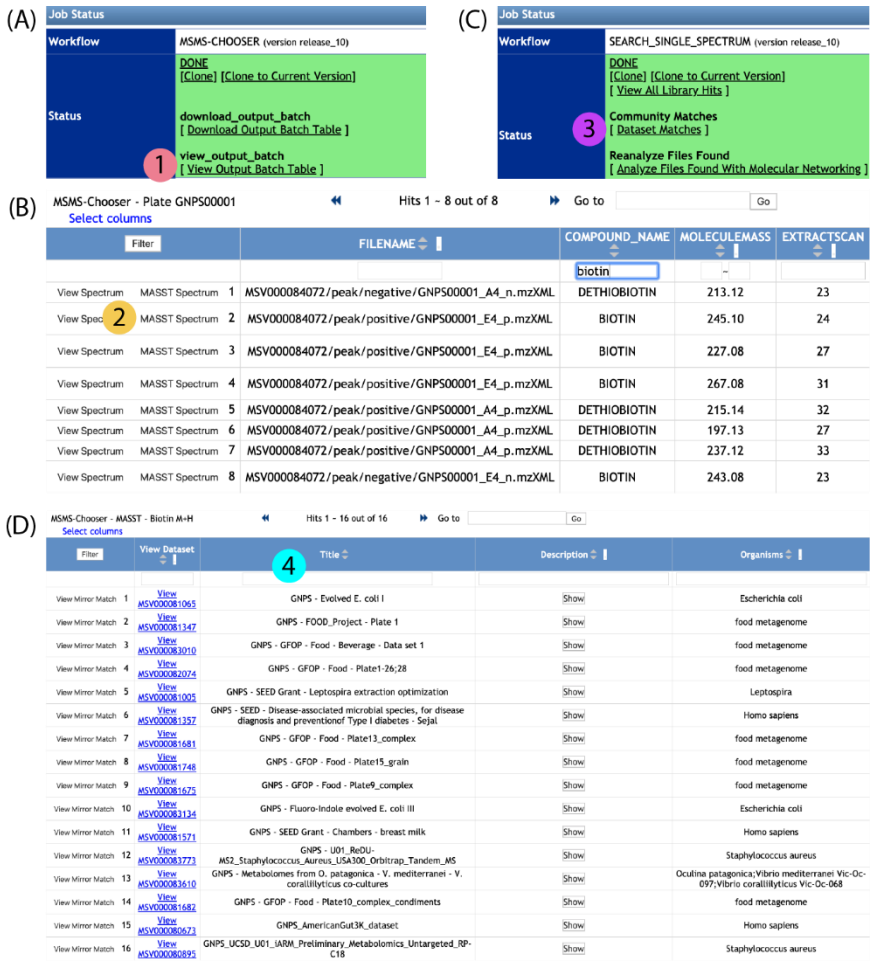


Figure 3.S1. Overview of launching a MASST search from the MSMS-Chooser workflow results page. (**a**) MSMS-Chooser results page, (**1**) click on the "View Output Batch Table". (**b**) View output batch page, (**2**) clicking on the "MASST Spectrum" button will launch a MASST search (*e.g.* biotin). (**c**) MASST search results page (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=39cd886540c147e9a8a618b275e5f541) , (**3**) click on the "Dataset Matches". (**d**) Datasets in which biotin (an example) was detected using MASST, (**4**) the title and dataset accession link can be used to further investigate the data.

Table 3.1. Tabulated results comparing MSMS-Chooser and manual inspection of data included in MSV000084286.

Ionization Mode	Number of MS/MS spectra, Manual Inspection	Number of MS/MS spectra, MS-MS Chooser[a]	MS-MS Chooser Accuracy (%)	MS-MS Chooser Accuracy (%)[b]
Positive	73	70	95.9	98.6
Negative	53	50	94.3	100
Total	126	120	95.2	99.2

[a] MSMS-Chooser correctly identified 70 MS/MS spectra and incorrectly identified one MS/MS spectra in positive mode upon manual inspection which as removed for subsequent calculations.

[b] Manual inspection found 2 positive mode and 3 negative mode spectra that were not chosen by MSMS-Chooser. These spectra were excluded in the accuracy calculation as their mass error was greater than the 10 ppm mass error limit (an adjustable parameter) used in MSMS-Chooser.

Table 3.S1. Chemicals analyzed and the resulting number of MS/MS spectra added via the MSMS-Chooser workflow. See attached spreadsheet

H. References

1. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI); Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211–221 (2007).

2. Palmer A, Phapale P, Fay D, Alexandrov T; Curatr: a web application for creating, curating and sharing a mass spectral library. *Bioinformatics* **34**, 1436–1438 (2018).

3. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N; Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).

4. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T; MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45**, 703–714 (2010).

5. The MassBank consortium https://github.com/MassBank (accessed Nov 10, 2019).

6. Sawada Y, Nakabayashi R, Yamada Y, Suzuki M, Sato M, Sakata A, Akiyama K, Sakurai T, Matsuda F, Aoki T, Hirai MY, Saito K; RIKEN tandem mass spectral

database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry* **82**, 38–45 (2012).

7. Schymanski EL, Neumann S; The Critical Assessment of Small Molecule Identification (CASMI): Challenges and Solutions. *Metabolites* **3**, 517–538 (2013).

8. Kyle JE, Crowell KL, Casey CP, Fujimoto GM, Kim S, Dautel SE, Smith RD, Payne SH, Metz TO; LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data. *Bioinformatics* **33**, 1744–1746 (2017).

9. Kessner D, Chambers M, Burke R, Agus D, Mallick P; ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536 (2008).

10. Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, Fahy E, Steinbeck C, Subramanian S, Bolton E, Greiner R, Wishart DS; ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **8**, 61 (2016).

11. Wang M, Jarmusch AK, Vargas F, Aksenov AA, Gauglitz JM, Weldon K, Petras D, da Silva R, Quinn R, Melnik AV, van der Hooft JJJ, Rodríguez AMC, Nothias LF, Aceves CM, Panitchpakdi M, Brown E, Di Ottavio F, Sikora N, Elijah EO, Labarta-Bajo L, Gentry EC, Shalapour S, Kyle KE, Puckett SP, Watrous JD, Carpenter CS, Bouslimani A, Ernst M, Swafford AD, Zúñiga EI, Balunas MJ, Klassen JL, Loomba R, Knight R, Bandeira N, Dorrestein PC; MASST: A Web-based Basic Mass Spectrometry Search Tool for Molecules to Search Public Data. doi:10.1101/591016

12. Jarmusch AK, Wang M, Aceves CM, Advani RS, Aguire S, Aksenov AA, Aleti G, Aron AT, Bauermeister A, Bolleddu S, Bouslimani A, Rodríguez AMC, Chaar R, Coras R, Elijah EO, Ernst M, Gauglitz JM, Gentry EC, Husband M, Jarmusch SA, Jones II KL, Kamenik Z, Gouellec AL, Lu A, McCall LI, McPhail KL, Meehan MJ, Melnik AV, Menezes RC, Giraldo YAM, Nguyen NH, Nothias LF, Nothias-Esposito M, Panitchpakdi M, Petras D, Quinn R, Sikora N, van der Hooft JJJ, Vargas F, Vrbanac A, Weldon K, Knight R, Bandeira N, Dorrestein PC; Repository-Scale Co- and Re-Analysis of Tandem Mass Spectrometry Data. doi:10.1101/750471.

Chapter IV

Paroxetine administration in mice affects bile acid

A. Abstract

Recent interest in the role of microbiota in health and disease has implicated gut microbiota dysbiosis in psychiatric disorders including major depressive disorder. Several antidepressant drugs that belong to the class of Selective Serotonin Reuptake Inhibitors have been found to display antimicrobial activities. In fact, one of the first antidepressants discovered serendipitously in the 1950s, the monoamine-oxidase inhibitor Iproniazid, was a drug used for the treatment of tuberculosis. In the current study we chronically treated DBA/2J mice for two weeks with Paroxetine, a Selective Serotonin Reuptake Inhibitor, and collected fecal pellets as a proxy for the gut microbiota from the animals after 7 and 14 days. Behavioral testing with the forced swim test revealed significant differences between Paroxetine and vehicle treated mice. Untargeted mass spectrometry and 16S rRNA profiling of fecal pellet extracts showed several primary and secondary bile acid level, and alpha diversity differences between Paroxetine and vehicle treated mice. In addition to their lipid absorbing activities bile acids have important signaling activities and have been associated with gastrointestinal diseases and colorectal cancer. Antidepressant drugs like Paroxetine should therefore be used with caution to prevent undesirable side effects.

B. Introduction

Selective Serotonin Reuptake Inhibitors (SSRIs) are commonly used drugs for the treatment of depression, post-traumatic stress disorder, generalized anxiety disorder (GAD) and other psychiatric disorders. However, a favorable patient response to SSRIs is not guaranteed and currently there are no biomarkers that can predict a positive treatment response; thereby

preventing its strategic utilization to treat psychiatric disorders (1). Furthermore, SSRI treatment can have side effects such as diarrhea, headaches, insomnia, nausea, and weight gain, preventing it from being routinely used (2). A more thorough investigation into the mechanistic behavior of SSRI drugs is critical for optimizing drug efficacy to improve patient outcomes and minimize side effects.

The gut-brain-axis (GBA) is a bidirectional mode of communication between the central nervous system (CNS) and the enteric nervous system (ENS) which uses the vagus nerve to coordinate the nervous system, the endocrine system, and the immune system. Recent studies have shown that the GBA is mediated by the production of microbially secreted molecular constituents that impact central nervous system function and behavior relevant to psychiatric disorders (3-9). Because SSRI drugs are administered orally, it is reasonable to assume that they can alter the gut microbiota and their functions, thereby impacting other body organs including the brain (10). In this regard it has been shown that medications can affect gut physiology by altering the intestinal microbiota which can result in unfavorable side effects that include constipation and tissue toxicity (11).

Previous studies have shown that Paroxetine (PARO)-treated DBA/2J mice show less floating in the forced swim test (FST) and increased time spent in the lit compartment of the dark light box, both indicating lower depression-like behavior (12-16). In order to assess the chronic effects of SSRI treatment, we examined fecal pellets from DBA/2J mice that were treated for two weeks with PARO. Each fecal pellet was then subjected to 16S rRNA sequencing and liquid chromatography tandem mass spectrometry analysis (LC-MS/MS) to identify bacterial taxa and metabolite profiles, respectively that differ between PARO- and

vehicle (VEH)-treated mice. Our results suggest that no significant alterations of the fecal taxa occur in response to PARO treatment. However, we observed several fecal bile acid level differences between PARO- and VEH-treated mice.

In mice, bile acids are synthesized from cholesterol and after conjugation to taurine are secreted into the bile and then to the small intestine. In the intestine bile acids are deconjugated by microbes (17) and further modified by different reactions (17) resulting in a variety of different secondary bile acids, whose numbers are still growing. Most of the bile acids are reabsorbed in the ileum and recycled with only a small amount (~2%) found in the colon and feces (17).

Bile acids are critical for lipid digestion and absorption and drug absorption, impacting drug pharmacokinetics (18). In addition, bile acids represent signaling molecules for several nuclear receptors with functions in lipid and glucose metabolism gene expression (19). Secondary bile acids such as deoxycholic acid and lithocholic acid produced by the gut microbiota have functions in host metabolic processes, drug metabolism, and immune response (20-22). Depending on the drug dose and length of treatment this could be responsible for some of the observed side effects associated with SSRI treatment. In the present study we have found that PARO treatment affects bile acid levels in mice including two recently identified novel secondary bile acids. This suggests that microbiota functions are altered by the drug.

C. Materials and Methods

1. Animal Housing and Husbandry

The experiments were carried out with male DBA/2J mice (Charles River Laboratories, Chatillon-sur-Chalaronne, France) with ages ranging between 8-10 weeks old. Prior to the beginning of the experiments, all animals were single-housed for at least one week. The mice were held under normal light and temperature conditions (12 h light: 12 h dark light cycle, lights on at 7 pm, temperature at 22 ± 2 °C, and humidity at 55 ± 5%) with standard bedding and nesting material, in polycarbonate cages (21 x 15 x 14 cm). Water and standard housing food pellets (ssniff-Spezialdiäten GmbH, Soest, Germany) were provided *ad libitum*. All procedures were carried out in accordance with the European Communities Council Directive 2010/63/EU and approved by the local animal welfare authority (G 15-1-041, Landesuntersuchungsamt Rheinland-Pfalz, Koblenz, Germany).

2. Drug Administration

To mimic clinical conditions as close as possible, animals received customized palatable pills (23), with a concentration of 5 mg/kg body weight paroxetine hydrochloride (Sigma Aldrich, Germany) or vehicle. We randomized the mice to treatment or control groups. Three days before the start of treatment, all mice were habituated to voluntary intake of vehicle pills given twice daily. Subsequently, mice received PARO or vehicle pills twice daily. Compliance was monitored twice daily by screening the cages for remnants of pills. Mice that did not eat two or more pills over the 14 days treatment schedule were excluded from analyses.

3. Behavioral Analyses

Forced Swim Test (FST). We introduced the mice into a 2-L glass beaker (diameter 13cm, height 24cm) filled with tap water (21±1 °C) to a height of 15cm. We videotaped the mice for 5 min and floating and active coping behavior was scored by an experienced, treatment-blinded observer.

4. Stool collection and DNA and metabolite extraction

Fecal pellets were collected from the same mice following 7 and 14 days (prior to behavioral analysis) of PARO- or VEH-treatment, snap-frozen on dry ice and stored at -80 °C prior to 16S and metabolomics extraction. Fecal samples were dried via a centrifugal low-pressure SpeedVac Plus system (Savant, Hyannis, MA) and their dry weights in microcentrifuge tubes recorded. 150 µL of sterile DEPC water and a chemically cleaned stainless-steel bead were added aseptically to each sample and followed by tissue homogenization using a TissueLyzer II (Qiagen, Hilden, Germany) for 5 min at 25 Hz. Samples were centrifuged at 14,000 RPM for 10 min and the resulting supernatant aspirated into a DNA/RNA free tube and stored at −80 °C for 16S sequence analysis. One ml 50% methanol was added to the remaining fecal homogenate and sonicated again for 5 min at 25 Hz. Following centrifugation at 14,000 RPM for 10 min, the supernatant was aspirated into a clean microcentrifuge tube, dried in a SpeedVac and stored at -80 °C until mass spectrometry analysis. Samples were resuspended with 150 µL 50:50 MeOH:H_2O spiked with 2 µM sulfadimethoxine and 2 µM sulfamethazineprior for mass spectrometry analysis.

5. 16S rRNA and bioinformatic analyses

Samples were processed following the Earth Microbiome Project 16S protocol (PMID: 27822518). In short, DNA from samples was extracted via the MoBio PowerSoil kit, then extracts were amplified using PCR (515F/806R primers: 5'-GTGCCAGCMGCCGCGGTAA-3' and 5'- GGACTACHVGGGTWTCTAAT -3', respectively), which targets the V4 region of the 16S ribosomal subunit. Amplicons were sequenced using the Illumina MiSeq platform at the Institute for Genomic Medicine of the University of California, San Diego. Metadata and microbiome data was uploaded and processed using Qiita (24). Raw sequences were demultiplexed and quality controlled as defined in the default processing within Qiita and the resulting sequences were collapsed into Amplicon Sequence Variants (ASV) using Deblur 1.1.0 (PMID: 28289731). Further down analysis where performed using QIIME2 [PMID: 31341288] as implemented in Qiita. The ASV table was rarefied at 5,000 sequences per sample for further down analysis. Alpha and beta-diversity was compared between treatment and control groups using Faith's Phylogenetic Diversity (PMID: 19455206), and the weighted and unweighted UniFrac distances (25). Principal co-ordinate analysis plots were used to visualize the beta-diversity among the samples and PERMANOVA testing was used to identify statistically different clusters.

6. Metabolomics analysis

Mouse fecal pellet extracts were analyzed using an ultra-high pressure liquid chromatography system (Vanquish, Thermo Scientific, Waltham, MA) coupled to an Orbitrap mass spectrometer (Q Exactive, Thermo Scientific) fitted with a heated electrospray ionization (HESI-II, Thermo Scientific) probe. Chromatographic separation was accomplished using a

Kinetex C18 1.7 µm, 100 Å, 2.1 mm x 50 mm column fitted with a C18 guard cartridge (Phenomenex, Torrance, CA) with a flow rate of 0.5 mL/min. Five µL of extract was injected per sample/QC. The column compartment and autosampler were held at 40 °C and 4 °C, respectively throughout all runs. Mobile phase composition was: A, LC-MS grade water with 0.1 % formic acid (v/v) and B, LC-MS grade acetonitrile with 0.1 % formic acid (v/v). The chromatographic elution gradient was: 0.0 - 1.0 min, 5% B; 1.0 - 9.0 min, 100% B; 9.0 - 11.0 min, 100% B; 11.0 - 11.5 min, 5% B; and 11.5 - 12.5 min, 5% B. Heated electrospray ionization parameters were: spray voltage, 3.5 kV; capillary temperature, 380.0 °C; sheath gas flow rate, 60.0 (arb. units); auxiliary gas flow rate, 20.0 (arb. units); auxiliary gas heater temperature, 300.0 °C; and S-lens RF, 60 (arb. units). Mass spectrometry data was acquired in positive mode using a data dependent method with a resolution of 35,000 in MS1 and a resolution of 17,000 in MS2. An MS1 scan from 100-1500 *m/z* was followed by an MS2 scan, using collision induced dissociation, of the five most abundant ions from the prior MS1 scan.

7. Metabolomics data pre-processing and statistical analysis

The acquired mass spectrometry data was converted to the open mzXML-format using MSconvert from ProteoWizard (26). Files were then further processed with the implemented ADAP-modules (27) in MZmine II (v. 2.35) (28) to generate MS1 features with associated peak area and MS2 scans (parameters are listed in **Table 4.S1**). The “export to GNPS” module on MZmine II was used to generate MS1 (quant.csv) and MS2 (.mgf) files for use in the feature based molecular networking workflow (https://ccms-ucsd.github.io/GNPSDocumentation/featurebasedmolecularnetworking/) on the GNPS website (http://gnps.ucsd.edu). The data was

filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and MS/MS fragment ion tolerance of 0.02 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Edges between two nodes were kept in the network only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks.

Metabolic features were annotated by searching the GNPS (https://gnps.ucsd.edu) network and matching MS2 spectra. After GNPS networking we searched specifically for MS2 spectra that matched to bile acids. In order to remove redundancy and increase specificity we manually extracted single peaks of individual bile acid compounds and adducts of molecular ions when possible.

For statistical analysis the Perseus software (v1.5.5.3) (29) was used. Raw peak intensities of metabolic features were log2-transformed, missing values were estimated by replacing from normal distribution (width = 0.3, downshift = 1.8). Two sided T-test was performed. For correlation analysis of raw peak intensities and experimental meta data ("correlogram") the corrplot package (v0.84) (https://cran.r-

project.org/web/packages/corrplot/index.html) in R with RStudio (v1.1.456) (https://www.rstudio.com/) was used.

If not stated otherwise, graphs and tests were performed in Microsoft Excel.

D. Results and Discussion

Previous studies (23) have shown that DBA/2J mice respond to the SSRI PARO with behavioral alterations that reflect a reduced depression-like phenotype. In line with previous experiments, we fed mice with either VEH- or PARO-containing pills (2 x 5 mg/kg/day) for 14 days. Compared to direct injection, this method of drug administration is less stressful for the animal and mitigates the effects of confounding behavioral assays used to assess depression-like behavior. In addition, oral administration is the preferred method considering the aim of the study to analyze the effects of the drug on gut microbiota. PARO pill consumption was carefully monitored to ensure that every animal was subjected to the same amount of the drug.

Confirming results from previous experiments (12-16), mice of the PARO-treated group showed significantly higher active coping time ($p = 1.5E-7$) in the FST compared to VEH-treated animals indicating the antidepressant-like and anxiolytic effects of PARO in DBA/2J mice (**Figure 4.1**).

In order to investigate the effects of PARO on the gut microbiota we used fecal pellets from drug- and VEH-treated animals as a proxy. Fecal pellets were collected from the same mice following one week and two weeks of daily treatments. Fecal pellet microbiota were collected for 16S bacterial rRNA sequence analysis and metabolites extracted and prepared for LC-MS/MS as outlined in Methods.

No differences were found for beta-diversity of the microbiome in the fecal pellets after 7 and 14 days of PARO treatment (**Figure 4.S1**) with either the weighted or unweighted UniFrac distances. However, alpha diversity showed a reduced and significant ($p < 0.04$) Faith's Phylogenetic diversity in the 14 days PARO treatment. Furthermore, untargeted mass spectrometry analyses showed increased levels for 13 (3 $p < 0.05$) and 11 (4 $p < 0.05$, 2 $p < 0.01$) primary and secondary bile acids after one and two weeks of treatment, respectively (**Figure 4.2a,b**). Four sulfated bile acids with sum formulas $C_{24}H_{38}O_8S$ and $C_{24}H_{40}O_8S$, each detected with different chromatographic retention times, showed higher levels in PARO-treated mice.

Correlation analysis (**Figure 4.2c**, and **Figure 4.S2**) revealed negative correlations between body weight gain (BWgain) and PARO and three bile acid levels (DCA, $C_{24}H_{40}O_5$_1, $C_{24}H_{40}O_8S$_2) following 2 weeks of treatment. FST active coping correlates negatively with PARO metabolites after 1 week and 2 weeks, with several bile acid levels ($C_{24}H_{38}O_5$_1, $C_{24}H_{40}O_5$_3, $C_{24}H_{38}O_8S$_2) after 1 week. In addition, a strong correlation of FST active coping was found with GDCA after 2 weeks. Interestingly, the four sulfated bile acids negatively correlated with FST behavior and PARO metabolite III levels ($p < 0.1$) (**Figure 4.2c**). Bile acid sulfation is a known mechanism for elimination and detoxification (30,31) by increasing their solubility and decreasing intestinal absorption which results in higher fecal content. Greater sulfated bile acid production may thus be caused by PARO reflecting a mechanism to dispose of high drug levels from the organism.

Of particular interest are the correlations we found with behavioral measures of the recently characterized secondary bile acids phenylalanocholic acid and

phenylalanodeoxycholic acid (32) implicating that microbiota are affected in PARO-treated mice (**Figure 4.2a,b** and **Figure 4.S3**). Strong positive correlations were found for PARO and for several bile acids after one week. After two weeks phenylalanocholic acid showed a weak negative correlation with sulfated bile acids (**Figure 4.2c**).

Although we were unable to detect any alterations in taxa it is conceivable that gut microbiota can still have an impact on bile acid levels in response to PARO treatment. This is supported by previous studies where antibiotic treatment results in primary and secondary bile acid changes (33). In this regard several studies have shown that antidepressant drugs including Sertraline, Fluoxetine and PARO have antimicrobial activities that impact Gram positive bacteria (34). Of particular significance are recent findings that implicate bile acids in several gastrointestinal disorders (35-38) and intestinal tumorigenesis through their interaction with the nuclear farnesoid X receptor (FXR) (39,40).

Bile acids are generated in the liver from cholesterol by a multi-step mechanism involving 17 enzymes. Subsequently, they are metabolized in the intestine by the gut microbiota (41). Subsequent deconjugation takes place in gut residing bacteria with the help of bile salt hydrolase activity. Following deconjugation a small fraction of primary bile acids that do not undergo resorption enter the colon, where they are processed into secondary bile acids by *Clostridium* and *Eubacterium* (42-45). They can be either absorbed by passive processes or excreted in the feces. Secondary bile acids synthesized by the microbiota only make up a very small fraction of the total bile acid pool. Which steps of primary and secondary bile acid synthesis are affected by PARO treatment is unknown and awaits further studies on the mechanism of action of the drug in bacteria. A change in bile acid composition and levels have

been associated with several diseases including intestinal dysmotility, inflammatory bowel disease phenotypes, nonalcoholic fatty liver disease, and progression of colon cancer (46,47). Microbiota that impact secondary bile acids are an additional factor that when altered can contribute to health and disease by affecting body weight, lipid metabolism, intestinal mucosal function and cardiovascular function (47-49).

Our results suggest that antidepressant drugs like PARO should be used with caution to prevent undesirable side effects. Whether bile acid levels can serve as biomarkers for monitoring the SSRI treatment response remains to be tested with a greater number of animals that allow stratification in drug responders and non-responders.

E. Acknowledgements

This work was supported by the Max Planck Society and funded by Office of Naval Research Multidisciplinary University Research Initiative (MURI) Award, Award No. N00014-15-1-2809. We thank Gregory Humphrey, Karenina Sanders and Tara Schwartz for microbiome sample and library preparation.

Chapter 4, in full, has been submitted for publication of the material as it may appear in Frontiers in Psychiatry, 2019, Vargas, F; Dethloff, F; Emmanuel, E; Quinn, R; Park, D. I; Herzog, D. P, Müller, M. B; Gentry, E. C; Knight, R; Gonzalez, A; Dorrestein, P. C; Turck, C. W.; Frontiers Media, 2019. The dissertation author was the primary investigator and author of this paper.

F. Figures and Tables

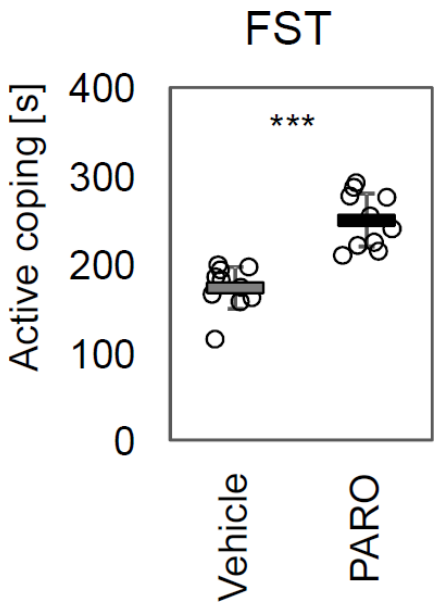


Figure 4.1. Paroxetine treatment revealed antidepressant-like and anxiolytic effects in DBA/2J mice. Chronic paroxetine (PARO) treatment for 14 days increased time of active coping (p =1.5 E-7) in the Forced Swim Test (FST) (*n* =10). Bars represent mean ± SD.

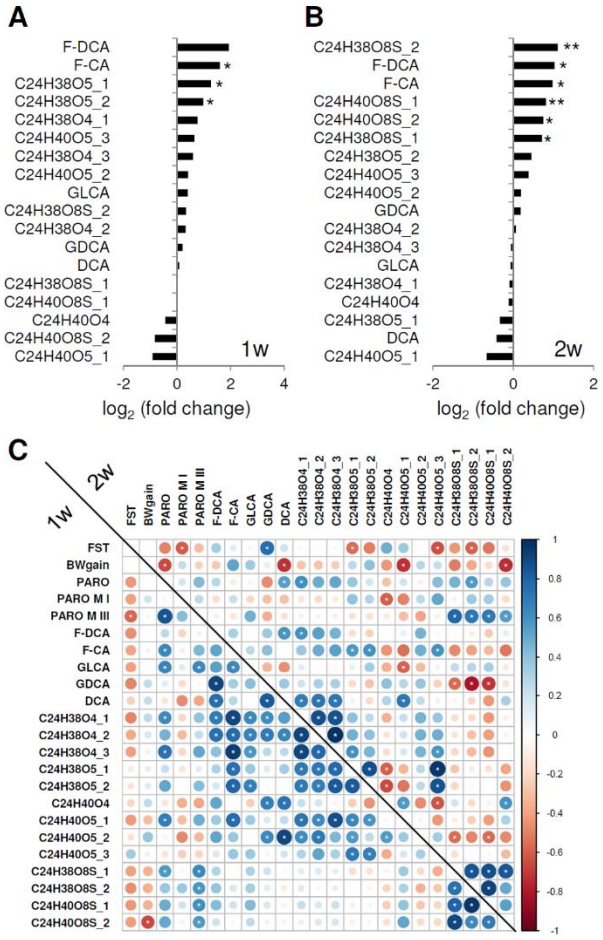


Figure 4.2. (**a**) Fecal pellet bile acid level ratios following 1 week (1w) and (**b**) two weeks of paroxetine (PARO) treatment. Bars represent mean log2 fold changes PARO *Vehicle-1. Significance was tested using Student's T-test; significance is indicated ($p < 0.05$ = *; $p < 0.01$= **). Several fecal pellet bile acid levels are increased following PARO treatment. (**c**) Several fecal bile acid levels correlate significantly with behavior, body weight gain, PARO metabolites, and other bile acid levels (indicated with white asterisk, $p < 0.1$). Correlogram displays the Pearson correlation coefficient values ($n = 10$) of treatment with PARO. Lower left represents the correlations after one week (1w) of treatment und upper right the correlations after two weeks (2w) of treatment. Size and color intensity reflect the absolute value as indicated by the color bar. Abbreviations in order of appearance: body weight gain (BWgain), forced swim test (FST), PARO (Paroxetine), PARO M I (Paroxetine metabolite I/II), PARO M III (Paroxetine metabolite III), phenylalanodeoxycholic acid (F-DCA), phenylalanocholic acid (F-CA), glycolithocholic acid (GLCA), glycodeoxycholic acid (GDCA), deoxycholic acid (DCA).

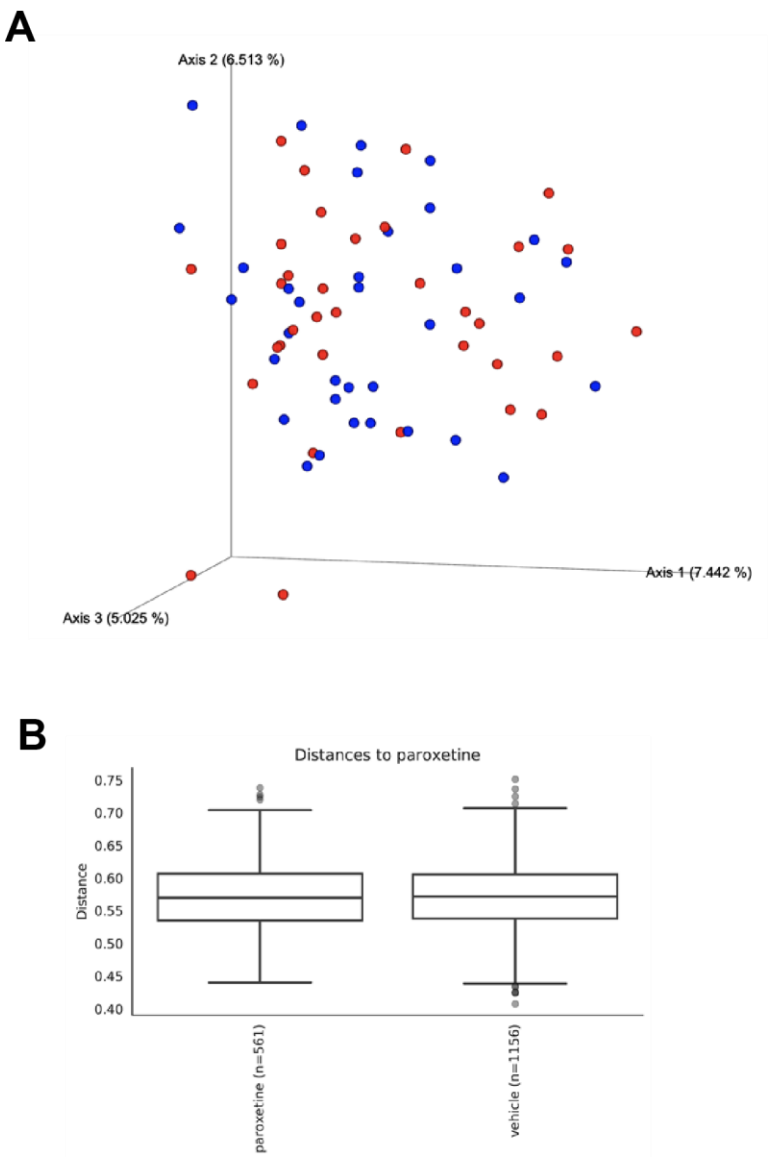


Figure 4.S1. (**a**) Unweighted UniFrac microbiome distances projected using principle coordinate analysis for analysis of beta-diversity of the microbiome data in this study. Blue spheres are fecal samples from vehicle-treated mice, red spheres are samples from paroxetine-treated mice. (**b**) Boxplots of unweighted UniFrac distances of samples in PCoA plot above used to calculate PERMANOVA testing. p-value from the PERMANOVA test is 0.982.

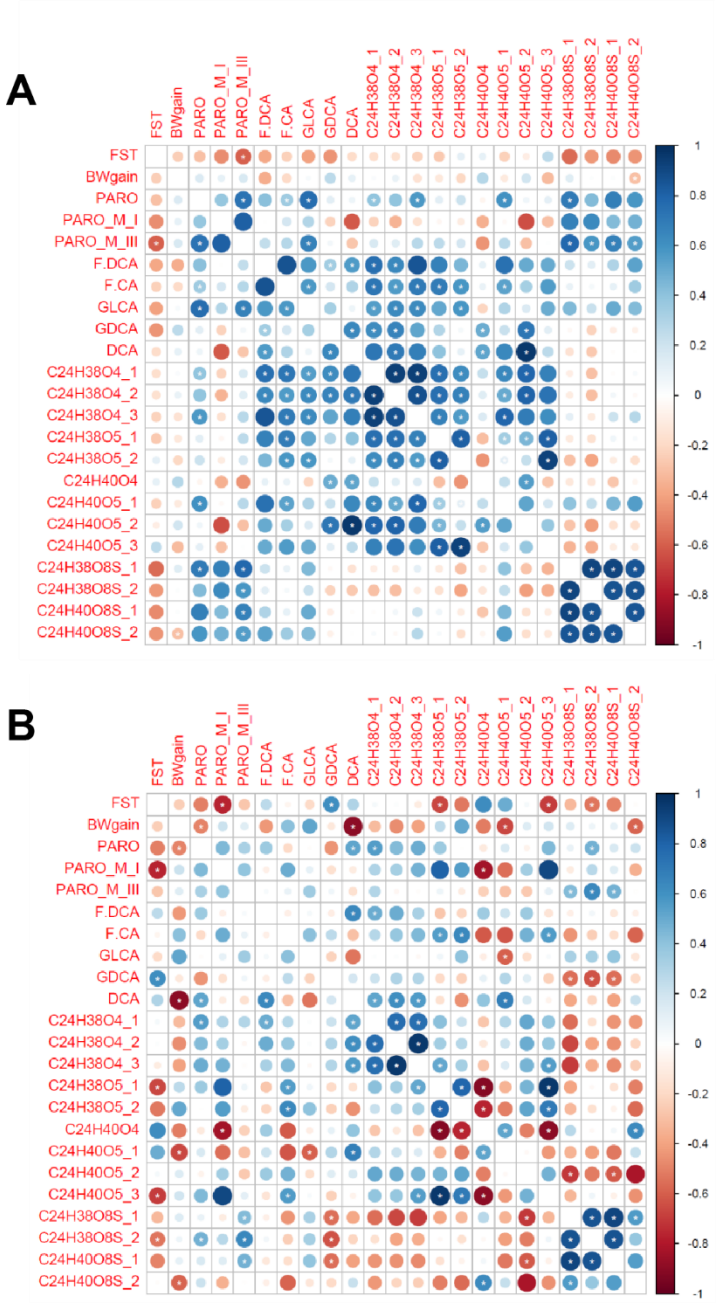


Figure 4.S2. Fecal pellet bile acid levels can be associated with body weight gain, behavior, and other bile acid levels. Correlation matrix is based on a Spearman correlation. Significance is indicated by white asterisk ($p < 0.1$). (**a**) One-week treatment and (**b**) Two-week treatment. Correlations are based on the PARO-treated mouse group ($n = 10$).

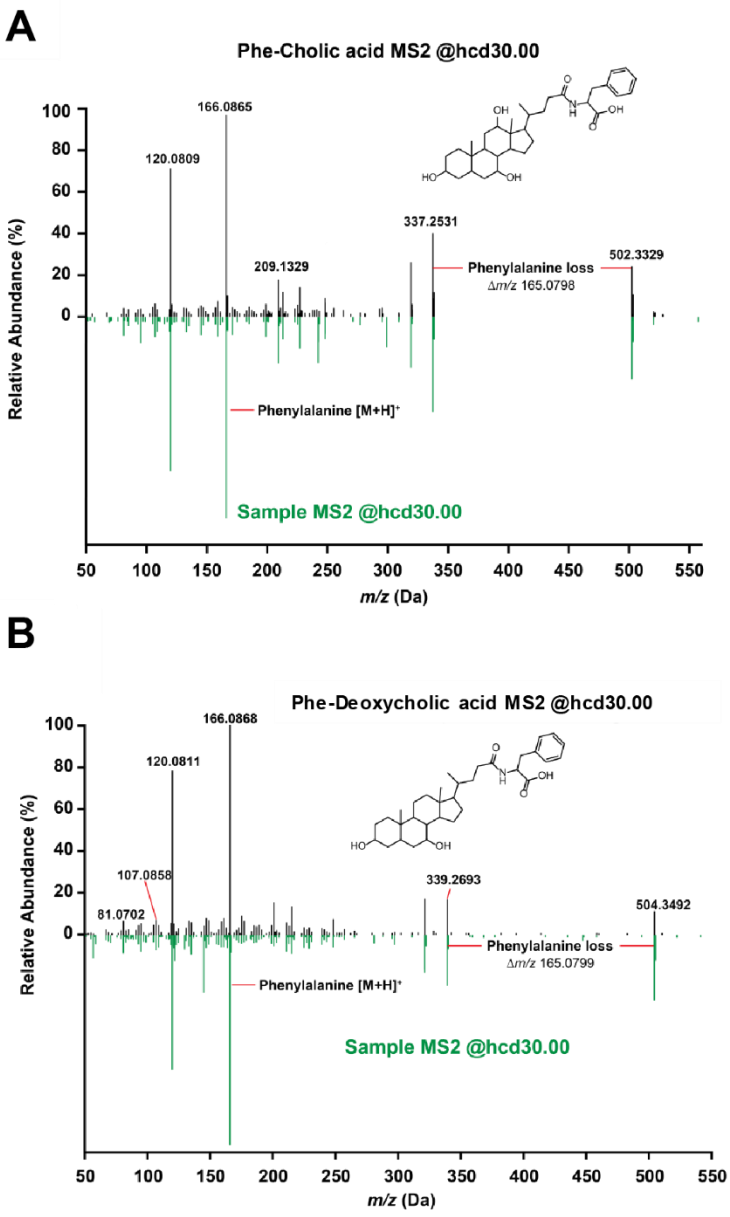


Figure 4.S3. (**a**) MS^2 spectrum of the newly found Phe-Cholic acid molecule in the sample mirrored against the Phe-Cholic acid molecule found in our GNPS database. (**b**) MS^2 spectrum of the newly found Phe-Deoxycholic acid molecule in the sample mirrored against the Phe-Deoxycholic acid molecule found in our GNPS database. Second best hit was to Phe-Chenodeoxycholic acid spectrum.

Table 4.S1. Tabulated Parameters used for MZmine feature detection

Crop filtering
- Retention time: 0 – 5.6 min
- m/z: 50 – 5990

Mass detection
- MS1 mass detector: 2.0 E5
- MS2 mass detector: 1.0 E2

ADAP Chromatogram builder
- Min group size in # scans 5
- Group intensity threshold 2.0E5
- Min time span (min): 0.05
- Min highest intensity: 6.0 E5
- m/z tolerance: 0.005 m/z or 10 ppm

Chromatogram deconvolution
- Local min search
  - Chromatogram threshold: 0.01%
  - Search minimum in RT range (min): 0.40
  - Min relative height: 0.01%
  - Min absolute height: 6.0 E5
  - Min ration of peak top/edge: 3
  - Peak duration range (min): 0.05 – 0.40
- m/z range for MS2 scan pairing (Da): 0.05
- RT range for MS2 scan pairing (min): 0.2

Isotope peak grouper:
- m/z tolerance: 0.01 m/z or 10 ppm
- Retention time tolerance (min): 0.1
- Maximum charge: 4

Join Alignment
- m/z tolerance: 0.005 m/z or 10 ppm
- Weight for m/z: 90
- Retention time tolerance (min): 0.3
- Weight for RT: 10

Peak list row filter
- Minimum peaks in a row: 2
- Reset the peak number ID

Gap fill
- Intensity tolerance: 20%
- m/z tolerance: 0.01 m/z or 10 ppm
- Retention time tolerance (min): 0.2

G. References

1. Fredman SJ, Fava M, Kienke AS, White CN, Nierenberg AA, Rosenbaum JF. Partial response, nonresponse, and relapse with selective serotonin reuptake inhibitors in major depression: a survey of current "next-step" practices. *J Clin Psychiatry*. 2000 Jun;**61**(6):403–408. PMID: 10901336.

2. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGrath PJ, Shores-Wilson K, Biggs MM, Balasubramani GK, Fava M, STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006 Jan;**163**(1):28–40. PMID: 16390886.

3. Jiang H, Ling Z, Zhang Y, Mao H, Ma Z, Yin Y, Wang W, Tang W, Tan Z, Shi J, Li L, Ruan B. Altered fecal microbiota composition in patients with major depressive disorder. *Brain Behav Immun*. 2015 Aug;**48**:186–194. PMID: 25882912.

4. Wong M-L, Inserra A, Lewis MD, Mastronardi CA, Leong L, Choo J, Kentish S, Xie P, Morrison M, Wesselingh SL, Rogers GB, Licinio J. Inflammasome signaling affects anxiety- and depressive-like behavior and gut microbiome composition. *Mol Psychiatry*. 2016 Jun;**21**(6):797–805. PMCID: PMC4879188.

5. Zheng P, Zeng B, Zhou C, Liu M, Fang Z, Xu X, Zeng L, Chen J, Fan S, Du X, Zhang X, Yang D, Yang Y, Meng H, Li W, Melgiri ND, Licinio J, Wei H, Xie P. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism [Internet]. *Molecular Psychiatry*. 2016;**21**:786–796.

6. Dinan TG, Cryan JF. Mood by microbe: towards clinical translation. *Genome Med*. 2016 Apr 6;**8**(1):36. PMCID: PMC4822287.

7. Yang C, Fujita Y, Ren Q, Ma M, Dong C, Hashimoto K. Bifidobacterium in the gut microbiota confer resilience to chronic social defeat stress in mice. *Sci Rep*. 2017 Apr 3;**7**:45942. PMCID: PMC5377462

8. Ho P, Ross DA. More Than a Gut Feeling: The Implications of the Gut Microbiota in Psychiatry [Internet]. *Biological Psychiatry*. 2017;**81**:e35–e37.

9. Sharon G, Cruz NJ, Kang D-W, Gandal MJ, Wang B, Kim Y-M, Zink EM, Casey CP, Taylor BC, Lane CJ, Bramer LM, Isern NG, Hoyt DW, Noecker C, Sweredoski MJ, Moradian A, Borenstein E, Jansson JK, Knight R, Metz TO, Lois C, Geschwind DH, Krajmalnik-Brown R, Mazmanian SK. Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. *Cell*. 2019 May 30;**177**(6):1600–1618.e17. PMID: 31150625.

10. Macedo D, Adriano José Maia, de Sousa CNS, Quevedo J, Barichello T, Júnior HVN, de Lucena DF. Antidepressants, antimicrobials or both? Gut microbiota dysbiosis in depression and possible implications of the antimicrobial effects of antidepressant drugs for antidepressant effectiveness [Internet]. *Journal of Affective Disorders*. 2017;**208**: 22–32.

11. Philpott HL, Nandurkar S, Lubel J, Gibson PR. Drug-induced gastrointestinal disorders. *Frontline Gastroenterol*. 2014 Jan;**5**(1):49–57. PMCID: PMC5369702.

12. Sillaber I, Panhuysen M, Henniger MSH, Ohl F, Kühne C, Pütz B, Pohl T, Deussing JM, Paez-Pereda M, Holsboer F. Profiling of behavioral changes and hippocampal gene expression in mice chronically treated with the SSRI paroxetine. *Psychopharmacology*. 2008 Nov;**200**(4):557–572. PMID: 18629477.

13. Webhofer C, Gormanns P, Tolstikov V, Zieglgänsberger W, Sillaber I, Holsboer F, Turck CW. Metabolite profiling of antidepressant drug action reveals novel drug targets beyond monoamine elevation. *Transl Psychiatry*. 2011 Dec 13;**1**:e58. PMCID: PMC3309495.

14. Webhofer C, Gormanns P, Reckow S, Lebar M, Maccarrone G, Ludwig T, Pütz B, Asara JM, Holsboer F, Sillaber I, Zieglgänsberger W, Turck CW. Proteomic and metabolomic profiling reveals time-dependent changes in hippocampal metabolism upon paroxetine treatment and biomarker candidates. *J Psychiatr Res*. 2013 Mar;**47**(3):289–298. PMID: 23207114.

15. Park DI, Dournes C, Sillaber I, Uhr M, Asara JM, Gassen NC, Rein T, Ising M, Webhofer C, Filiou MD, Müller MB, Turck CW. Purine and pyrimidine metabolism: Convergent evidence on chronic antidepressant treatment response in mice and humans. *Sci Rep*. 2016 Oct 12;**6**:35317. PMCID: PMC5059694.

16. Park DI, Dournes C, Sillaber I, Ising M, Asara JM, Webhofer C, Filiou MD, Müller MB, Turck CW. Delineation of molecular pathway activities of the chronic antidepressant treatment response suggests important roles for glutamatergic and ubiquitin-proteasome systems. *Transl Psychiatry*. 2017 Apr 4;**7**(4):e1078. PMCID: PMC5416684.

17. Ridlon JM, Kang DJ, Hylemon PB, Bajaj JS. Bile acids and the gut microbiome [Internet]. *Current Opinion in Gastroenterology*. 2014;**30**:332–338.

18. Cussotto S, Clarke G, Dinan TG, Cryan JF. Psychotropics and the Microbiome: a Chamber of Secrets…. *Psychopharmacology* . 2019 May;**236**(5):1411–1432. PMCID: PMC6598948.

19. Hylemon PB, Zhou H, Pandak WM, Ren S, Gil G, Dent P. Bile acids as regulatory molecules [Internet]. *Journal of Lipid Research*. 2009;**50**:1509–1520.

20. Allen K, Jaeschke H, Copple BL. Bile acids induce inflammatory genes in hepatocytes: a novel mechanism of inflammation during obstructive cholestasis. *Am J Pathol*. 2011 Jan;**178**(1):175–186. PMCID: PMC3070591.

21. Nguyen A, Bouscarel B. Bile acids and signal transduction: Role in glucose homeostasis [Internet]. *Cellular Signalling*. 2008;**20**:2180–2197.

22. Thomas C, Pellicciari R, Pruzanski M, Auwerx J, Schoonjans K. Targeting bile-acid signalling for metabolic diseases. *Nat Rev Drug Discov*. 2008 Aug;**7**(8):678–693. PMID: 18670431.

23. Carrillo-Roa T, Labermaier C, Weber P, Herzog DP, Lareau C, Santarelli S, Wagner KV, Rex-Haffner M, Harbich D, Scharf SH, Nemeroff CB, Dunlop BW, Craighead WE, Mayberg HS, Schmidt MV, Uhr M, Holsboer F, Sillaber I, Binder EB, Müller MB. Common genes associated with antidepressant response in mouse and man identify key role of glucocorticoid receptor sensitivity. *PLoS Biol*. 2017 Dec;**15**(12):e2002690. PMCID: PMC5746203.

24. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat Methods*. 2018 Oct;**15**(10):796–798. PMCID: PMC6235622.

25. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005 Dec;**71**(12):8228–8235. PMCID: PMC1317376.

26. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak M-Y, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*. 2012 Oct;**30**(10):918–920. PMCID: PMC3471674.

27. Myers OD, Sumner SJ, Li S, Barnes S, Du X. One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms

and Detecting Chromatographic Peaks. *Anal Chem*. 2017 Sep 5;**89**(17):8696–8703. PMID: 28752754.

28. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*. 2010 Jul 23;**11**:395. PMCID: PMC2918584.

29. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016 Sep;**13**(9):731–740. PMID: 27348712.

30. Alnouti Y. Bile Acid sulfation: a pathway of bile acid elimination and detoxification. *Toxicol Sci*. 2009 Apr;**108**(2):225–246. PMID: 19131563.

31. Hagey LR, Krasowski MD. Microbial Biotransformations of Bile Acids as Detected by Electrospray Mass Spectrometry [Internet]. *Advances in Nutrition*. 2013;**4**:29–35.

32. Quinn RA, Vrbanac A, Melnik AV, Patras KA, Christy M, Nelson AT, Aksenov A, Tripathi A, Humphrey G, da Silva R, Bussell R, Thron T, Wang M, Vargas F, Gauglitz JM, Meehan MJ, Poulsen O, Boland BS, Chang JT, Sandborn WJ, Lim M, Garg N, Lumeng J, Kazmierczak BI, Jain R, Egan M, Rhee KE, Haddad GG, Siegel D, Mazmanian S, Nizet V, Knight R, Dorrestein PC. Chemical Impacts of the Microbiome Across Scales Reveal Novel Conjugated Bile Acids [Internet]. Available from: http://dx.doi.org/10.1101/654756.

33. Behr C, Slopianka M, Haake V, Strauss V, Sperber S, Kamp H, Walk T, Beekmann K, Rietjens IMCM, van Ravenzwaay B. Analysis of metabolome changes in the bile acid pool in feces and plasma of antibiotic-treated rats. *Toxicol Appl Pharmacol*. 2019 Jan 15;**363**:79–87. PMID: 30502395.

34. Munoz-Bellido JL, Munoz-Criado S, Garcìa-Rodrìguez JA. Antimicrobial activity of psychotropic drugs [Internet]. *International Journal of Antimicrobial Agents*. 2000;**14**:177–180.

35. Dior M, Delagrèverie H, Duboc H, Jouet P, Coffin B, Brot L, Humbert L, Trugnan G, Seksik P, Sokol H, Rainteau D, Sabate J-M. Interplay between bile acid metabolism and microbiota in irritable bowel syndrome. *Neurogastroenterol Motil*. 2016 Sep;**28**(9):1330–1340. PMID: 27060367.

36. Duboc H, Rajca S, Rainteau D, Benarous D, Maubert M-A, Quervain E, Thomas G, Barbu V, Humbert L, Despras G, Bridonneau C, Dumetz F, Grill J-P, Masliah J, Beaugerie L, Cosnes J, Chazouillères O, Poupon R, Wolf C, Mallet J-M, Langella P, Trugnan G, Sokol H, Seksik P. Connecting dysbiosis, bile-acid dysmetabolism and gut

inflammation in inflammatory bowel diseases. *Gut*. 2013 Apr;**62**(4):531–539. PMID: 22993202.

37. Pereira-Fantini PM, Bines JE, Lapthorne S, Fouhy F, Scurr M, Cotter PD, Gahan CG, Joyce SA. Short bowel syndrome (SBS)-associated alterations within the gut-liver axis evolve early and persist long-term in the piglet model of short bowel syndrome. *J Gastroenterol Hepatol*. 2016 Dec;**31**(12):1946–1955. PMID: 27037739.

38. Theriot CM, Bowman AA, Young VB. Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for Clostridium difficile Spore Germination and Outgrowth in the Large Intestine. *mSphere*. 2016 Jan;**1**(1).

39. Anakk S, Watanabe M, Ochsner SA, McKenna NJ, Finegold MJ, Moore DD. Combined deletion of Fxr and Shp in mice induces Cyp17a1 and results in juvenile onset cholestasis. *J Clin Invest*. 2011 Jan;**121**(1):86–95. PMCID: PMC3007143.

40. Fu T, Coulter S, Yoshihara E, Oh TG, Fang S, Cayabyab F, Zhu Q, Zhang T, Leblanc M, Liu S, He M, Waizenegger W, Gasser E, Schnabl B, Atkins AR, Yu RT, Knight R, Liddle C, Downes M, Evans RM. FXR Regulates Intestinal Cancer Stem Cell Proliferation. *Cell*. 2019 Feb 21;**176**(5):1098–1112.e18. PMCID: PMC6701863.

41. de Aguiar Vallim TQ, Tarling EJ, Edwards PA. Pleiotropic roles of bile acids in metabolism. *Cell Metab*. 2013 May 7;**17**(5):657–669. PMCID: PMC3654004.

42. Russell DW. The Enzymes, Regulation, and Genetics of Bile Acid Synthesis [Internet]. *Annual Review of Biochemistry*. 2003;**72**:137–174.

43. Kitahara M, Takamine F, Imamura T, Benno Y. Assignment of Eubacterium sp. VPI 12708 and related strains with high bile acid 7alpha-dehydroxylating activity to Clostridium scindens and proposal of Clostridium hylemonae sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol*. 2000 May;**50** Pt 3:971–978. PMID: 10843034.

44. Kitahara M, Benno Y, Takamine F, Imamura T. Clostridium hiranonis sp. nov., a human intestinal bacterium with bile acid 7alpha-dehydroxylating activity [Internet]. *International Journal of Systematic and Evolutionary Microbiology*. 2001;**51**:39–44.

45. Ridlon JM, Bajaj JS. The human gut sterolbiome: bile acid-microbiome endocrine aspects and therapeutics. *Acta Pharm Sin B*. 2015 Mar;**5**(2):99–105. PMCID: PMC4629220.

46. Camilleri M. Advances in understanding of bile acid diarrhea. *Expert Rev Gastroenterol Hepatol*. 2014 Jan;**8**(1):49–61. PMCID: PMC4211077.

47. Jones ML, Martoni CJ, Ganopolsky JG, Labbé A, Prakash S. The human microbiome and bile acid metabolism: dysbiosis, dysmetabolism, disease and intervention. *Expert Opin Biol Ther*. 2014 Apr;**14**(4):467–482. PMID: 24479734.

48. Nieuwdorp M, Gilijamse PW, Pai N, Kaplan LM. Role of the Microbiome in Energy Regulation and Metabolism [Internet]. *Gastroenterology*. 2014;**146**:1525–1533.

49. Swann JR, Want EJ, Geier FM, Spagou K, Wilson ID, Sidaway JE, Nicholson JK, Holmes E. Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proc Natl Acad Sci U S A*. 2011 Mar 15;**108** Suppl 1:4523–4530. PMCID: PMC3063584.

Chapter V

The fecal metabolome is altered by stress and dietary prebiotics: Relationships between sleep, gut metabolites, and the gut microbiome

A. Abstract

Dietary prebiotics produce favorable changes in the commensal gut microbiome and reduce host vulnerability to stress induced disruptions in complex behaviors such as sleep. The mechanisms for how prebiotics modulate stress physiology remain unclear; however, emerging evidence suggests that gut microbes and their metabolites may play a role. This study tested if stress and/or dietary prebiotics (Test diet) alter the fecal metabolome; and explored if these changes were related to sleep and/or gut microbial alpha diversity. Male F344 rats on either Test or Control diet were instrumented for electroencephalography biotelemetry measures of sleep/wake. After 5 weeks on diet, rats were either stressed or remained in their home cages. Based on untargeted liquid chromatography coupled to mass spectrometry and 16S rRNA gene sequencing, both stress and Test diet altered the fecal metabolome/microbiome. In addition, Test diet prevented the stress-induced reduction in microbial alpha diversity based on PD_Whole_Tree. Network propagation analysis revealed that stress increased members of a neuroactive steroidal molecular family; and that Test diet reduced this effect. We also discovered links between sleep, alpha diversity, and pyrimidine, secondary bile acid, and neuroactive glucocorticoid/pregnanolone-type steroidal metabolites. These results reveal novel microbial-dependent metabolites that may modulate stress physiology and sleep.

B. Introduction

Evidence in both the human and animal literatures suggest that stressor exposure can negatively impact sleep (1-4) and the gut microbiota (5-7). We have previously reported, for

example, that male rats exposed to an acute stressor (90 min, intermittent tail shocks) experience disruptions in sleep architecture, have flattened diurnal rhythmicity of core body temperature, learning deficits, express depressive-like behaviors, and altered gut microbial alpha diversity (4,8-10). *Ad libitum* consumption of prebiotic diets, compared to nutrient/calorically matched control diets, prior to stressor exposure attenuates many of these effects (7,11). In addition, ingestion of prebiotic diet also improves non-rapid eye movement (NREM) sleep, promotes increases in rapid eye movement (REM) sleep after stressor exposure (REM rebound), and prevents stress-induced reductions in gut microbial alpha diversity (7).

The mechanisms for how prebiotic diet-induced changes in the gut microbiota impact stress physiology remain unclear; however, there is emerging evidence that gut microbial metabolites likely play a role (12,13). Bacterial dependent metabolites, such as short chain fatty acids and secondary bile acids, can signal the brain through the blood and/or vagal afferents (14-16). It is feasible, therefore, that gut microbial modulatory diets produce changes in the gut metabolome that signal the brain and impact complex brain functions. Fecal samples were collected both before and after stressor exposure and impacts on overall gut metabolomic structure were determined using untargeted mass spectrometry chromatography coupled to a hybrid quadrupole-Orbitrap mass spectrometer fitted with a HESI probe (LC-MS/MS). Those features driving the changes in metabolomic structure produced by diet and/or stress were then targeted for additional metabolomics standards initiative (MSI) identification and state-of-the-art network propagation analyses. Additional analyses explored if any of the identified gut metabolite changes were related to our previously reported effects on sleep and alpha diversity measured in fecal samples in the same rats (7). Discovering and identifying gut metabolites that

are modulated by diet and/or stress and relate to sleep and microbial alpha diversity adds to our understanding of microbiota-gut-brain signaling and could hasten the development of health promoting microbiome therapeutics.

C. Materials and Methods

1. Animals

Adult male F344 rats (*n* = 52, Harlan Laboratories) were housed with controlled temperature and humidity and all procedures were approved by the University of Colorado Animal Care and Use as previously described in detail (7). Briefly, animals weighed 40 - 50 g upon arrival at post-natal day (PND) 24 and were maintained on a 12:12 h light/dark cycle. All rats were housed in Nalgene Plexiglas cages and were placed on control or Test diet (*ad libitum*) upon arrival at PND 24. Rats were allowed to acclimate to housing conditions for 1 week prior to study initiation. No differences in body weight for food consumption were found and there were 14 - 15 animals per group (diet) at PND day 70 data and 7 - 8 animals per group (diet x stress) for the PND 91 data (**Figure 5.1a**).

2. Experimental design

The experimental design is depicted in **Figure 5.1** and adapted from Thompson *et al.*, 2017. Fecal samples used for gut metabolomics analysis were collected at the same time points as those used for gut microbiome analyses previously reported (7). Animals consumed diet for 7 weeks prior to the first fecal sample collection at PND 70 and then for 11 weeks. Animals

were then exposed to inescapable tail shock stress or not, and four days following inescapable tail shock stress another fecal sample was taken at PND 91 (**Figure 5.1a**).

3. Control and Test diet Availability

Test diet contained the following gut microbial modulatory nutrients, which are absent from control diet: galactooligosaccharides (GOS 7.00 g (active)/Kg ; FrieslandCampina, Zwolle, Netherlands), polydextrose (PDX, 6.58 g (active)/Kg; Danisco, Terre Haute, IN, USA), lactoferrin (LAC, 1.86 g/Kg; Tatua Cooperative Dairy Company, Morrinsville, New Zealand), and whey protein concentrate milk fat globular membrane protein-10 (MFGM, 15.9 g/Kg; Arla Food Ingredients, Aarhus, Denmark). GOS is a non-absorbable complex carbohydrate derived from the enzymatic breakdown of lactose; and PDX is a processed polymer derived from glucose and classified as a soluble fiber by the US Food and Drug Administration. Both GOS and PDX are classified as prebiotic substrates because they are (1) not hydrolyzed or absorbed in the stomach or small intestine; (2) selective substrates for beneficial commensal bacteria in the colon, such as *Lactobacillus* spp.; and (3) induce beneficial luminal/systemic effects within the host (42-46). LAC impacts the gut microbiota through microbiostatic and antimicrobial activity (47,48), and MFGM alters antimicrobial activity (49) as well as the microbiota (50,51). Animals were fed either control or Test diet and experimenters were blind to diet type. The diets were formulated by Mead Johnson Nutrition (MJN, Evansville IN, USA) based on AIN-93G specifications and were isocaloric with similar carbohydrate, protein, fat, vitamin, and mineral levels.

4. Fecal and sample collection

Stool samples were collected as previously described (7). Briefly, rats were placed into a sterile cage until defecation (< 10 min) where samples were collected and rats were immediately returned to the home cage. Samples were then transferred and stored in a -80 °C freezer for untargeted metabolomics analysis.

5. Stress protocol

Inescapable stress is a well-characterized laboratory stressor that produces depression/anxiety-like behavior, increases corticosterone, and disrupts diurnal physiology, sleep and the microbiome (3,4,7,852,53). In brief, rats were placed in Plexiglas restraining tubes (23.4 cm long and 7.0 cm in diameter) and exposed to 100, 5 sec, 1.5 mA inescapable tail shocks (Stress). Shocks were delivered at random with an average interval of 60 sec between shocks and occurred during the inactive (light) cycle between ~0800 and 1100 h. After exposure to inescapable tail shock rats were immediately returned to their home cages. Animals that were not exposed to the stressor (No Stress), remained undisturbed in their home cages.

6. Sleep measures

Sleep was measured using *in vivo* biotelemetry, as previously described (4,8,54). The complete sleep results from these rats have been previously reported in (7). In brief, the F40-EET biotelemetry transmitters (Data Sciences International, St. Paul Minnesota) were implanted into animals and the electroencephalographic (EEG) insulated leads were passed subcutaneously to the base of the skull, where they were attached to stainless steel screws

(Plastics One Inc.) and served as EEG recording electrodes. Biotelemetry recordings of EEG were acquired using Dataquest ART 4.3 Gold Acquisition/Analysis Software (Data Sciences International, St. Paul, MN) and sleep/wake cycles were scored using the automated Neuroscore 2.1.0 software (Data Sciences International, St. Paul, MN). After sleep recordings were autoscored, they were corrected for accuracy by an observer blind to the experimental treatment of each animal.

The current study used two measures of sleep, NREM bout duration and % REM. NREM and REM measures were derived from the trace EEG signal after fast Fourier Transformation (FFT), yielding spectra between 0.5 and 30 Hz in 0.5 Hz frequency bins. NREM sleep was identified by increased absolute EEG amplitude with integrated values for the delta frequency band (0.5 - 4.5 Hz) being greater than those for the theta frequency band (6.0 - 9.0 Hz). REM sleep was characterized by low amplitude EEG with integrated values for the delta frequency band less than those for the theta frequency band. Time spent in REM was calculated as a percentage (%) of time spent in a specific behavioral state per hour. Average bout durations per hour of NREM were also calculated. Bout durations were defined by any change in sleep/wake state for 10 seconds (i.e., an NREM bout was defined based on the appearance of 10 sec epoch or longer of NREM and the end of that epoch was defined as the appearance of any 10 sec epoch of either REM or WAKE).

7. 16S rRNA Gene sequencing and microbial alpha diversity analyses

Fecal samples were previously collected at PND 70 and PND 91 and sequenced as described (7). These same fecal samples were used for metabolomics analysis. In brief, after

purification and precipitation to remove polymerase chain reaction (PCR) artifacts, samples were exposed to multiple sequencing on an Illumina Genome Analyzer IIx. Operational taxonomic units (OTUs) were picked using a 'closed reference' approach (55). GreenGenes May 2013 version was the reference database used (56), and all sequence processing was done with QIIME v 1.8.0 (57) using the UCLUST algorithm (58). Taxonomy and phylogeny were taken from the GreenGenes reference collection. The current experiment generated 14,207,155 sequences, where 11,016,354 were discarded because of uncorrectable barcode errors, of which 6,481 were too short to read (based on default parameters set in QIIME script 'split_libraries_fastq.py') and the remaining 3,190,801 sequences were used. The resulting OTU table was rarefied at 7400 sequences/sample to correct for uneven sequencing depth due to amplification differences between samples. Alpha diversity for this manuscript was measured using species richness (PD_Whole_Tree). This measure captures phylogenetic diversity for a given sample (59).

8. Metabolomics

Sample information – LC-MS/MS

A subset of frozen fecal samples was transferred via dry ice to the University of California, San Diego for metabolomic analysis. Fecal samples were stored in 1.5 mL centrifuge tubes at -80 °C prior to extractions. Sample ID's were manually uploaded into an electronic spreadsheet and subsequently used to assign filenames during LC-MS/MS data acquisition. All solvents used for the metabolomic analysis were of LC-MS grade.

Fecal pellet extraction – LC – MS/MS

This method was adapted from a previously published protocol (60). Fecal pellets were weighed to 50.0 +/- 2 mg wet weight and transferred to 2.0 mL round bottom microcentrifuge tubes (Qiagen Catalog# 990381) for metabolite extractions. A clean stainless-steel bead (Qiagen Catalog# 69989) and 1.5 mL chilled extraction solvent (50% MeOH) was added to each sample. The samples were then homogenized for 5 min at 25 Hz using a TissueLyser II system (Qiagen Catalog# 85300) and allowed to incubate for 20 min at -20 °C. The fecal homogenates were then centrifuged at 14000 rpm for 15 min at 4 °C. 1.2 mL aliquots were then transferred into Nunc 2.0 mL DeepWell plate (Thermo Catalog# 278743) and frozen at -80 °C prior to lyophilization using a FreeZone 4.5 L Benchtop Freeze Dryer with Centrivap Concentrator (Labconco). Wells were resuspended with 200 µL of resuspension solvent (50% MeOH spiked with 2.0 µM sulfadimethoxine), vortexed for 30 secs, and centrifuged at 2000 rpm for 15 min at 4 °C. 150 µL of the supernatant was transferred into a 96-well plate and maintained at 4 °C prior to LC-MS analysis. A resuspension solvent QC and a six standard mix QC (50% MeOH spiked with 1.0 µM Sulfamethazine, 1.0 µM Sulfamethizole, 1.0 µM Sulfachloropyridazine, 1.0 µM Amitrypline, and 1.0 µM Coumarin 314) was run every 12^{th} sample to asses sample background, carry over, chromatography behavior, peak picking and plate effects.

LC-MS/MS parameters

Fecal extracts were analyzed using an ultra-high performance liquid chromatography system (Vanquish, Thermo) coupled to a hybrid quadrupole-Orbitrap mass spectrometer (Q-

Exactive, Thermo) fitted with a HESI probe. Reverse phase chromatographic separation was achieved using a Kinetex C18 1.7 µm, 100 Å, 50 x 2.1 mm column (Phenomenex) held at 40 °C with a flow rate of 0.5 mL/min. 5.0 µL aliquots were injected per sample/QC. The mobile phase was (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. The elution gradient was: 5% B for 1 min, increased to 100% B in the next 8 min, held at 100% B for two min, returned to 5.0% B in 0.5 min, equilibrated at 5.0% B for 2 min. Positive electrospray ionization parameters were: sheath gas flow rate of 52 (arb. units), aux gas flow rate of 14 (arb. units), sweep gas flow rate of 3 (arb. units), spray voltage of 3.5 kV, capillary temperature of 270 °C, S-Lens RF level of 50 (arb. units), and aux gas heater temperature of 435 °C. Negative electrospray ionization parameters were: sheath gas flow rate of 52 (arb. units), aux gas flow rate of 14 (arb. units), sweep gas flow rate of 3 (arb. units), spray voltage of 2.5 kV, capillary temperature of 270 °C, S-Lens RF level of 50 (arb. units), and aux gas heater temperature of 435 °C. MS data was acquired using a data dependent acquisition method with a resolution of 35,000 in MS^1 and 17,000 in MS^2. An MS^1 scan from 100-1500 *m/z* was followed by an MS^2 scan, produced by collision induced disassociation, of the five most abundant ions from the prior MS^1 scan.

Standards run for Bile Acids

Primary, secondary, conjugated and unconjugated bile acids were purchased and used for level 1 identification of some of our unknown molecules. Standards were solubilized to a final concentration of 10 µM in 50% MeOH prior to LC-MS/MS injection.

9. Data Processing and Availability

LC-MS/MS parameters

All mass spectrometry data (.raw, .mzXML, and mgf files), mzMine, and Sirius files can be found in the online mass spectrometry repository Massive (http://massive.ucsd.edu) using the following accession numbers: MSV000079329 and MSV000079339.

Data Processing and Analysis

The orbitrap files (.raw) were exported to mzXML files using MSConvert (61). Feature detection of the MS^1 data was performed using MZmine2 (62), parameters can be found in **Table 5.S1**, which generated a data matrix of MS^1 features (*m/z* and retention time) and peak area. Each feature in a given sample was normalized against the spiked internal standard, to remove any spray variation across runs, followed by normalization of each sample by the row sum of its features. The twice-normalized data matrix was used for all subsequent statistical analysis. The SIRIUS export module found within MZmine was used to create .mgf files for molecular formula annotation and molecular structure prediction in the SIRIUS desktop software.

Global Natural Products Social Molecular Networking (GNPS) job parameters

Molecular networking was ran using networking parameters that yielded a false-discovery rate (FDR) of annotation, using Passatuto, for spectral matches against reference libraries of 1% (**Figure 5.S5**). A molecular network was created using the online workflow at GNPS (63). The data was then clustered with MS-Cluster with a parent mass tolerance of 0.05

Da and a MS^2 fragment ion tolerance of 0.05 Da to create consensus spectra. Further, consensus spectra that contained less than 2 spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.6 and more than 6 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. All matches kept between network spectra and library spectra were required to have a score above 0.52 and at least 6 matched peaks (64). The molecular networking job can be accessed using the following GNPS positive mode link: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=92166ef840924b078fed960323cbd558.

Metabolite annotation

Metabolites were annotated following the guidelines established by the 2007 metabolomics Standards initiative (MSI). Accurate mass with retention time alignment and MS^2 fragmentation pattern between a metabolite of interest and a chemical reference standard was used for all MSI level 1 annotations. Accurate mass and MS^2 fragmentation spectral alignment between a metabolite of interest and a reference library, via GNPS, was used for all MSI level 2 annotations. SIRIUS 4.0.1 was used for molecular formula annotation (MSI level 4) and CSI:FingerID, via SIRIUS 4.0.1, was used for molecular structure predication (MSI level 3) (65,66).

Statistical Analysis

Metabolomics data were visualized and analyzed using metaboanalyst open source www.metaboanalyst.ca built on R statistical software (67-71) and all other analyses were performed in SPSS version 25. All features were log transformed prior to further analysis. Principal components analysis was used to reduce the high dimensionality of the untargeted fecal metabolomics dataset. Unsupervised heat maps were generated using the Euclidean distance matrix with Ward clustering algorithm. For clarity and ease of interpretation only the top 50 unidentified features are displayed on the heat maps. For PND 70, a volcano plot (two group data) was used to identify significant differences between control and Test diet groups ($p < 0.05$; FDR $p < 0.01$). For PND 91, one-way ANOVA (four group data) was used ($p < 0.05$; FDR $p < 0.01$). In a final step, we examined potential relationships between host physiology, the gut microbiome (phylum level), and the identified gut metabolites using stepwise multiple regression analyses. These analyses were run on the normalized microbiome/metabolomics data and examined relationships between sleep (NREM and REM) and alpha diversity data that have been previously published (7). Differences were considered significant when $p < 0.05$, unless otherwise noted.

D. Results and Discussion

1. Test diet alters the fecal gut metabolome

At PND 70, the principal component (PC) plot demonstrates separation due to diet on the scores plot along component 1 and component 2 (**Figure 5.1a,b**). The top 50 unidentified features and targeted metabolites on the heat map show clustering by diet (**Figure 5.2**). Four days after stressor exposure at PND 91, diet impacted the fecal metabolome displayed on the

PC plot mostly along component 1 (**Figure 5.1a,c**). The top 50 unidentified features and targeted metabolites on the heat map (**Figure 5.3**) show main clustering by diet and some clustering by stress within the control diet group.

2. Test diet and stress significantly alter specific ions and metabolites

Volcano plot analysis (**Figure 5.S1**) of the fecal metabolome on PND 70 revealed that 21 features were significantly impacted by diet. All features, except one, were significantly higher in the Test diet when compared to the control diet. We identified 10 of these significant metabolites using CSIfinger ID achieving metabolomics standards initiative (MSI) level 3 annotation (**Figure 5.4a-j; Table 5.1**).

Analyses of the fecal metabolome on PND 91 revealed 36 unidentified features as significantly different between groups (one-way ANOVA $p < 0.05$, FDR $p < 0.01$; **Figure 5.2**). There were 28 of 36 features that were higher in the Test diet when compared to control diet, with no effect of stress. The remaining 8 of 36 features were altered by both diet and/or stress. We identified 14 of these features using CSIfinger ID achieving metabolomics standards initiative (MSI) level 3 annotation (**Figure 5.5a-n**; **Table 5.1**). See **Figure 5.5** for specific effects and post hoc analyses on individual metabolites.

3. Network analysis of neuroactive steroids

Based on the GNPS database, the stress-responsive allopregnanolone precursor in **Figure 5.5j** is a spectral match to the neuroactive steroid 5.alpha.-Pregnane-3.alpha.,21-diol-11,20-dione. The calculated mass is within 5 ppm and the MS/MS spectrum is a nearly identical

match (**Figure 5.S3**). Many sterols give rise to the same parent matches and similar MS^2 spectra. The spectral match to 5.alpha.-Pregnane-3.alpha.,21-diol-11,20-dione was at level 3, according to MSI (17) . A spectral match at MSI Level 3 confirms that the feature matches a neuroactive steroidal molecular family. In order to further elucidate this metabolite class in our fecal samples, we ran state-of-the-art network propagation analysis (18) in order to classify the allopregnanolone precursor (5.alpha.-Pregnane-3.alpha.,21-diol-11,20-dione) with familial molecules of similar structure.

The resulting network propagation analysis is depicted in **Figure 5.6** (see also Supplementary **Figure 5.4**). Interestingly, 5.alpha.-Pregnane-3.alpha.,21-diol-11,20-dione (yellow in **Figure 5.6**) was related to several other molecules. Specifically, 3.beta.-Allotetrahydrocortisol, 21-hydroxyallopregnanolone, 5.alpha.-Pregnane-3.alpha.,11.beta.,21-triol-20-one, and 5.alpha.-Pregnane-3.alpha.,21-diol-20-one (blue in **Figure 5.6**). The metabolite 5.alpha.-Pregnane-3.alpha.,21-diol-11,20-dione (yellow in **Figure 5.6**) belongs to a neuroactive steroidal molecular family and molecular network for glucocorticoid and pregnanolone-type steroids (MSI Level 3 spectral match), adding confidence that the Allopregnanolone precursor depicted in **Figure 5.5j** is correctly annotated as 5.alpha.-Pregnane-3.alpha.,21-diol-11,20-dione.

4. Relationships between identified metabolites, sleep architecture, and microbiome alpha diversity

We previously reported that rats fed Test diet had an increase in NREM sleep bout durations, when compared to those fed control diet [**Figure 5.4b,d** from (7)]. Potential

relationships between the observed NREM sleep changes at PND 71, 72 from our previous findings and the identified fecal metabolome at PND 70 (**Figure 5.4**) were explored using stepwise multiple regression analysis. This analysis revealed a significant linear relationship between Pyrimidine Nucleotide ($C_{16}H_{25}N_2O_{14}P$) depicted in **Figure 5.4i** and NREM sleep ($F_{(1, 28)} = 8.939$; $p = 0.006$; adj. $r^2 = 0.249$; $y = 10.013x + 415.225$; (**Figure 5.7a**)).

We also previously published that rats eating Test diet had increased REM sleep rebound following acute stress exposure [**Figure 5.5** from (7)]. It may be possible that a relationship exists between the increased REM sleep rebound after stress exposure and the gut metabolome. Indeed, stepwise multiple regression revealed that two identified features at PND 91 were related to REM sleep after stress at PND 87 ($F_{(2, 27)} = 5.736$; $p = 0.009$; adj. $r^2 = 0.260$; $y = 0.907x_1 + 0.467x_2 + 26.021$; **Figure 5.7b**). The first feature x_1 is the stress responsive Ketone Steroid (**Figure 5.5k**) and the second x_2 is Ethanebis(thioate) Derivative (**Figure 5.5b**).

Test diet attenuated the stress induced increase in alpha diversity of the gut microbiome [**Figure 5.7C** (PD_Whole_Tree) from (7)] and thus we examined potential relationships between the identified gut metabolites and alpha diversity. There was a significant relationship between two identified metabolites at PND 91 and alpha diversity ($F_{(2, 27)} = 27.092$; $p < 0.001$; adj. $r^2 = 0.659$; $y = 12.895x_1 – 0.727x_2 + 36.027$; **Figure 5.7c**). The two metabolites are x_1 = Hyodeoxycholic acid and x_2 = Alloprenanolone Precursor (yellow in **Figure 5.6**; **Figure 5.5j**).

E. Conclusions

Ingestion of a prebiotic diet (Test diet) improves undisturbed non-rapid eye movement (NREM) sleep, promotes REM sleep rebound after stress exposure, and prevents stress-induced

reductions in gut microbial alpha diversity (7). The results from the current report demonstrate that Test diet also modulates the fecal metabolome community, increases specific metabolites (fatty acids, sugars, steroids, nucleotides), and that prebiotic-induced sleep improvements are related to several gut metabolites. This report along with prior studies (7,9,11,19) suggests that gut metabolites are an important effector arm of the microbiota-gut-brain axis and adds to emerging evidence linking metabolomics and sleep physiology (20).

Stressor exposure affected the gut metabolome differently in the control vs. the Test diet. Test diet attenuated the stress-induced increase in the Allopregnanolone Precursor, the Ketone Steroid, as well as two other unidentified metabolite features. These small gut metabolites belong to a family of endogenous metabolites of corticosterone/progesterone. Network propagation analysis (18) revealed several other potential metabolites in the family. One such metabolite is 5.alpha.-Pregnane-3.alpha.,11.beta.,21-triol-20-one, which has been reported to block voltage dependent calcium channels (21) and another is 5.alpha.-Pregnane-3.alpha.,21-diol-20-one, better known as allotetrahydrodeoxycorticosterone, which is also a neuroactive steroid that potentiates GABAergic inhibition (22). This metabolite has also recently been linked with reduced sleep quality during pregnancy (23) and is involved in the acute stress response in rodents (24-26). These data suggest that stress may affect neuroactive steroid signaling in the gut, which is attenuated by a prebiotic diet. It could be that the negative consequences of stress exposure, in part, are mitigated through gut microbial modulatory substrates such as dietary prebiotics. This idea is congruent with a role for dysregulated neuroactive steroid signaling in stress-related psychiatric disorders (27).

We discovered several novel gut metabolites that were related to measures of sleep suggesting a potential link between the gut metabolome and sleep physiology. Specifically, improved NREM sleep was related to a Pyrimidine Nucleotide at PND 70. This relationship is consistent with prior literature examining a role for pyrimidine metabolism in sleep (28). CSIfinger ID gives the molecular formula for this Pyrimidine Nucleotide as $C_{16}H_{25}N_2O_{14}P$, however no known reference standard is yet available for this byproduct of pyrimidine metabolism, however based on the proposed molecular formula it may be involved in uracil metabolism, a metabolite known to increase NREM sleep (29,30). Similarly, the previously reported enhanced REM sleep rebound following stress (7) was significantly related to the stress-responsive Ketone Steroid ($C_{21}H_{32}O_2$) + Ethanebis(thioate) Derivative ($C_{10}H_{16}N_2O_4S_2$). Based on the molecular formula for the Ketone Steroid this molecule is a steroid derivative with a pregnane skeleton, however we verified using a standard that this molecule is not pregnenolone, per se, rather likely related to pregnenolone. Given that this class of molecules has been implicated in the regulation of sleep physiology (31); it is possible that the Ketone Steroid detected in this study is a novel sleep modulatory gut metabolite. The second factor in the relationship with REM sleep was an Ethanebis(thioate) Derivative. Based on the molecular formula this gut metabolite is still unannotated and represents a novel potential molecule potentially involved in sleep. These data, although clearly not conclusive, suggest for the first time a potential relationship between gut pyrimidine metabolism, stress responsive gut neuroactive steroids, and sleep physiology. Previous studies have linked both plasma/urine metabolites with sleep physiology (32-35), but to the best of our knowledge this is the first

study to relate changes in *fecal* gut metabolome produced after ingestion of gut microbial modulatory substrates with sleep physiology.

The gut metabolome and gut microbiome are linked through gut microbial metabolism. Our findings help clarify this relationship, such that microbiome alpha diversity was related to a fecal hyodeoxycholic acid and the Allopregnanolone Precursor. Hyodeoxycholic acid is a secondary bile acid that is dependent on the gut microbiota and impacts health and disease (36-38). It is possible, therefore, that some of the health benefits often associated with elevated gut microbial alpha diversity (7,39,40) are due to modulation of secondary bile acids. This idea is further supported by our data demonstrating that a Fatty Acid Derivative is modulated by both stress and Test diet. The second metabolite in the regression is the Allopregnanolone Precursor, which belongs to a molecular class of neuroactive steroids acting at GABA receptors (24). The relationship reported here between alpha diversity, secondary bile acids, and a neuroactive gut metabolite strengthens the idea that the gut metabolome is linked through gut microbial metabolism and taken together with other results suggests that these play a role in modulating stress and sleep physiology. This work represents an important step towards uncovering the potential mechanisms underlying health promoting gut microbial modulating substrates. There remain, however, many currently unidentifiable features that were modulated by prebiotic and/or stress. With continued improvements in available reference samples (41), future work will expand the identification of potential gut derived biomolecular pathways and could facilitate the discovery of additional novel molecules capable of impacting physiology and complex behavior.

F. Acknowledgements

We would like to acknowledge dietary formulation information from Tina Herfel at Envigo.

Chapter 5, in full, has been submitted for publication of the material as it may appear in Scientific Reports, 2019, Thompson, R. S; Vargas, F; Dorrestein, P. C; Chichlowski, M; Berg, B. M; Fleshner, M.; Nature Publishing Group, 2019. The dissertation author was a primary investigator and author of this paper.

G. Figures and Tables

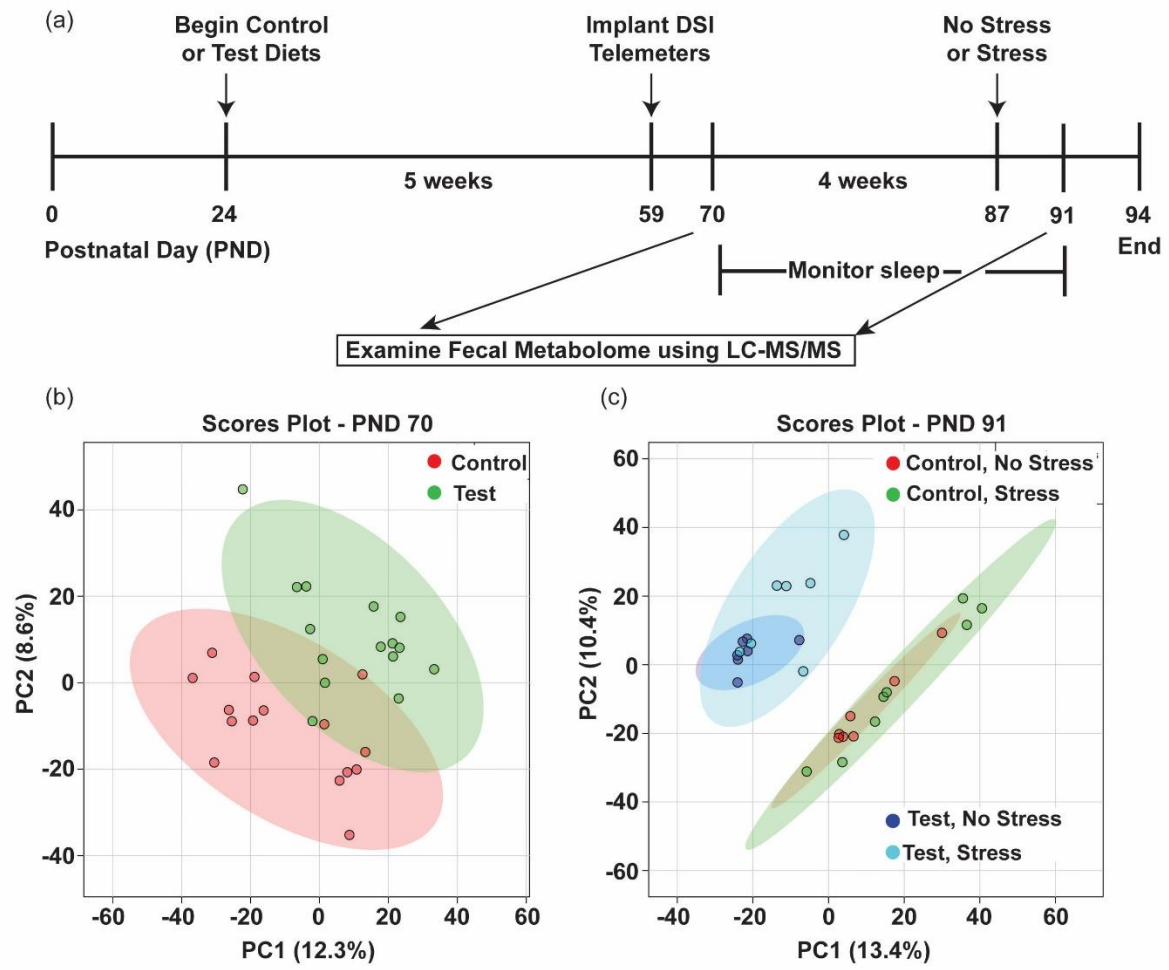


Figure 5.1. Adapted (**a**) timeline from Thompson *et al*. (2017), showing fecal samples were collected at PND 70 and at PND 91 for untargeted metabolomics analysis. (**b**) PC scores plot demonstrates separation by diet of the untargeted metabolome along component 1 and 2 at PND 70. (**c**) PC scores plot demonstrates large separation by diet of the untargeted metabolome along component 1 and 2 at PND 91. Shaded areas represent 95% confidence intervals. Abbreviations: Data Sciences International (DSI); Liquid chromatography-mass spectrometry (LC-MS/MS); Post-natal day (PND), Principal components (PC).

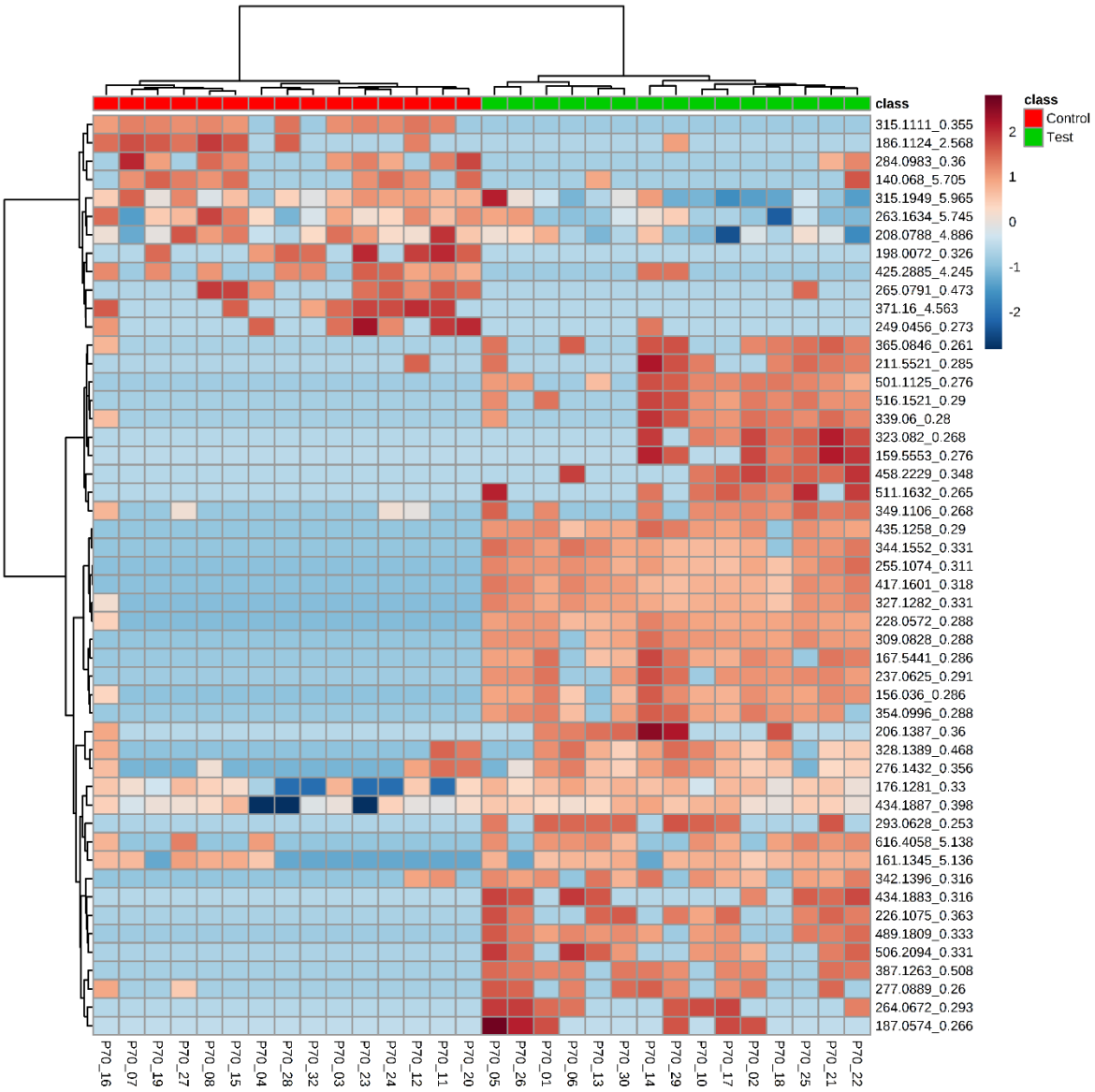


Figure 5.2. Heat map of the top 50 features that cluster (unsupervised) by control and Test diet measured in fecal samples collected on PND 70. Individual subjects are along the bottom, while features [mass/charge (*m/z*) retention times (min)] are listed to the right of the heat map. Approximately 12 features appear higher in the control compared to the Test diet while 38 appear higher in the Test diet.

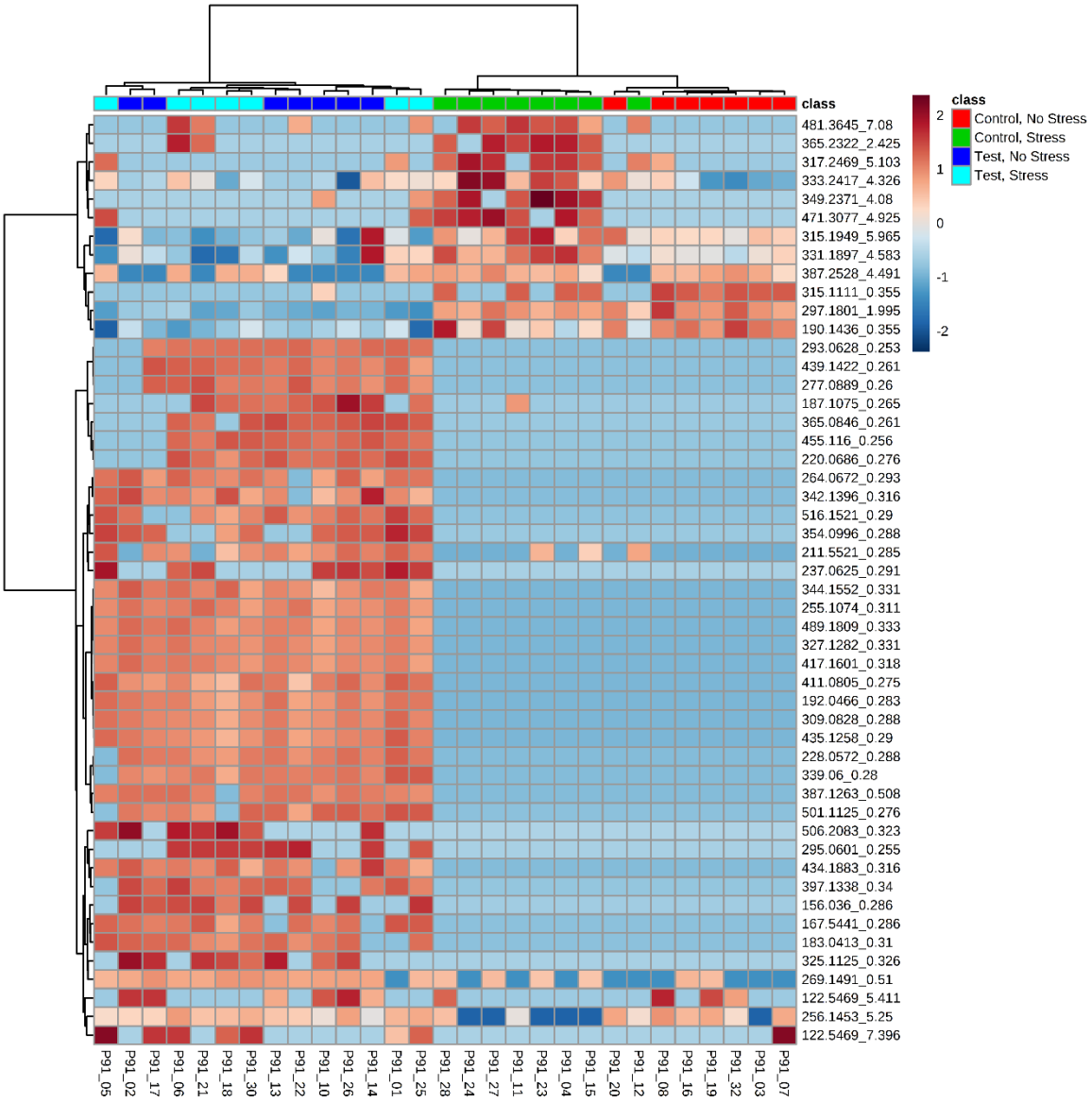


Figure 5.3. Heat map of the top 50 features that cluster (unsupervised) by diet and stressor exposure measured in fecal samples collected on PND 91. The features within the control diet clustered mostly by whether they were exposed to stress or not, suggesting a potential effect of stress exposure (upper right half of heat map). This effect of clustering by stress was absent in the Test diet group (upper left half of heat map). Overall, the features mostly clustered by diet, rather than stress, which is consistent with the PC plot. Individual subjects are along the bottom and features [mass/charge (*m/z*) retention times (min)] are listed to the right of the heat map.

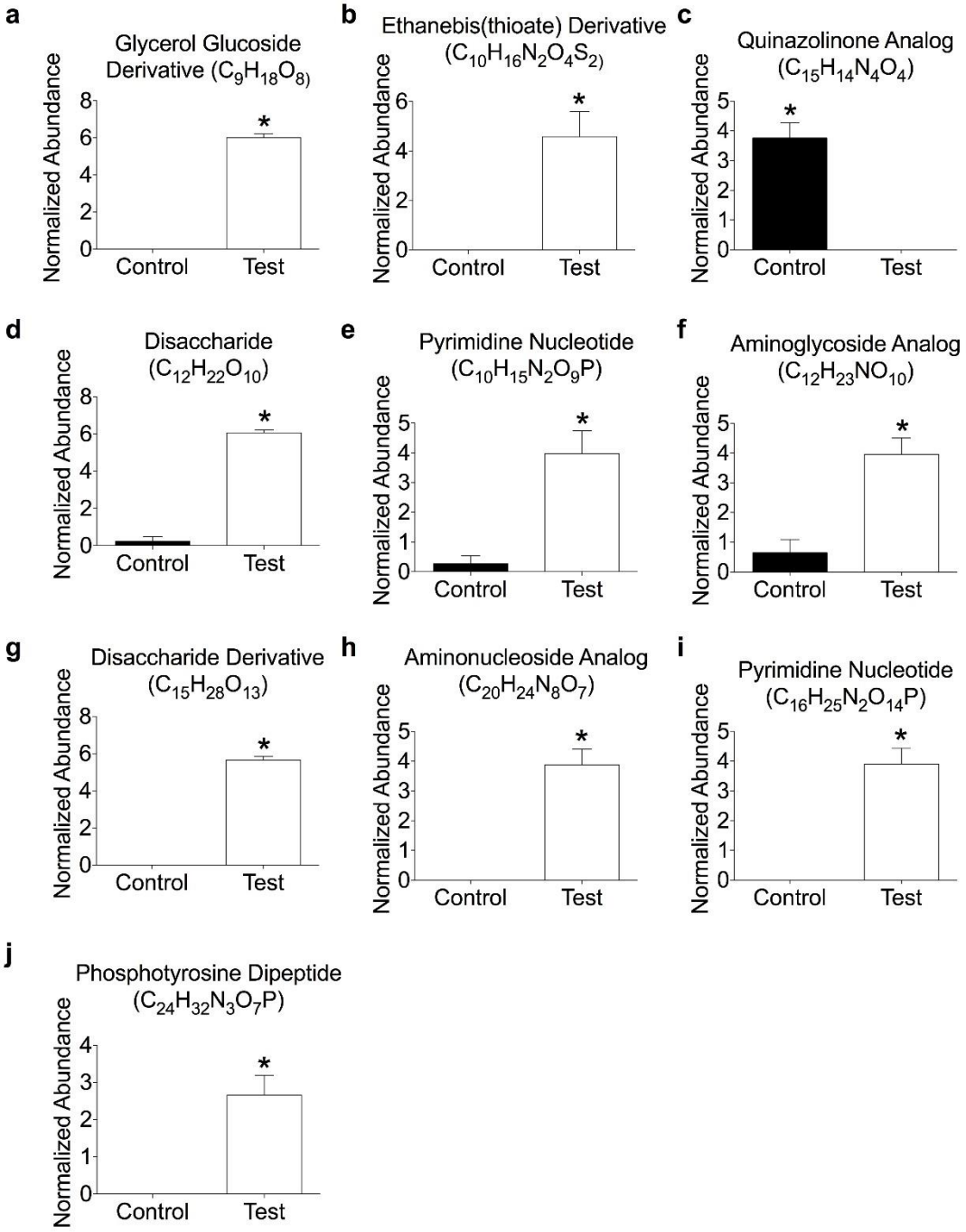


Figure 5.4. (**a**,**b**, **d**-**j**) Nine of the 10 identified metabolites were significantly higher in the Test diet when compared to the control diet at PND 70. (c) Alternatively, the quinazolinone analog was higher in the control diet when compared to the Test diet. (* $p < 0.05$ vs. control/test diet).

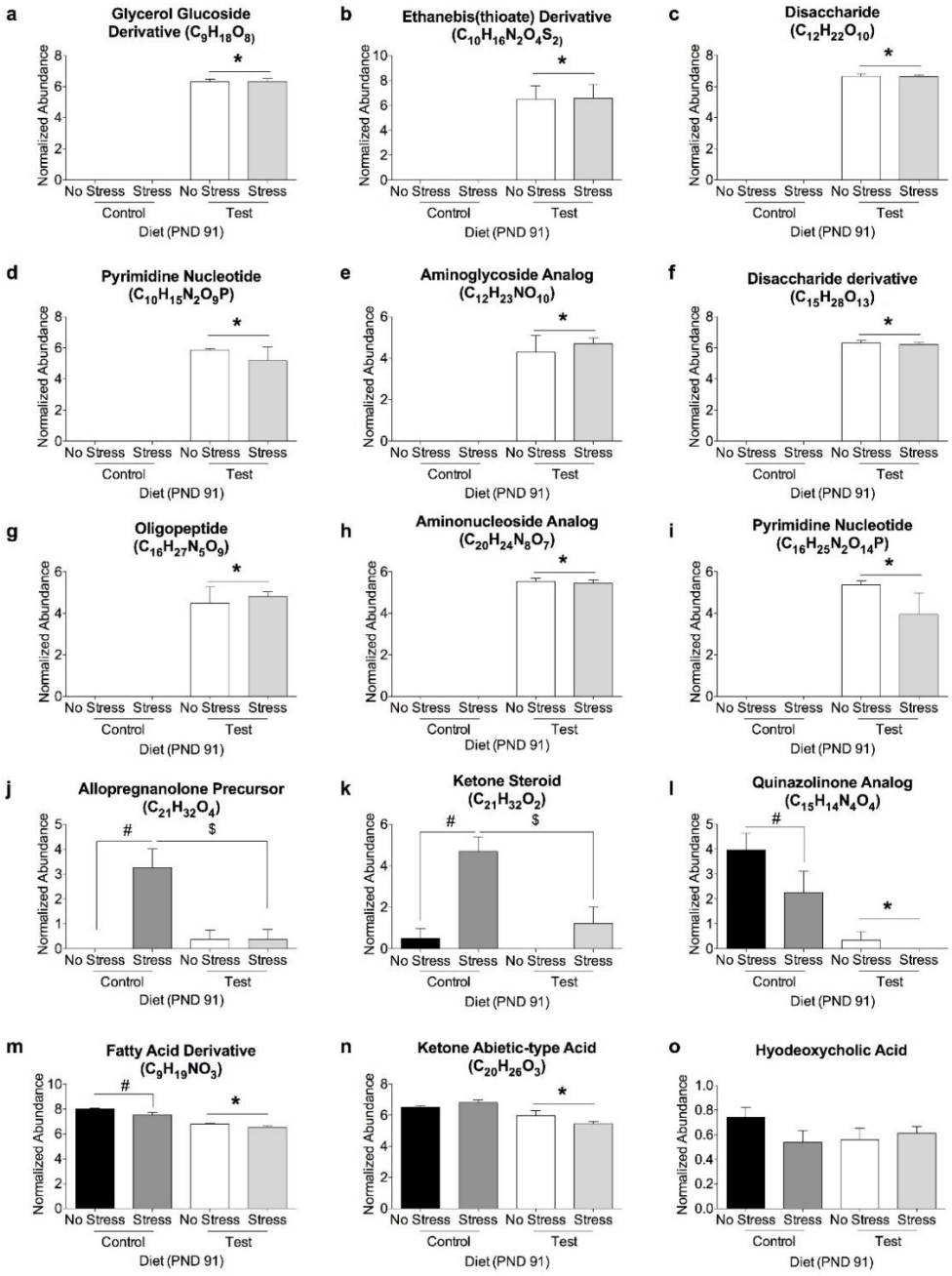


Figure 5.5. The metabolites in (**a-i**) were significantly higher in Test diet compared to control diet. The stress-induced increase in the (**j**) allopregnanolone precursor and (**k**) ketone steroid was attenuated by Test diet. (**l**) Quinazolinone analog, which was higher in the control diet, was decreased due to stress exposure. Both (**m**) the Fatty Acid Derivative and (**n**) the Ketone Abietic-type Acid were lower in the test diet, where stress decreased the Fatty Acid Derivative in the control diet group only. Finally, there was no significant effect of either diet or stress on (**o**) the secondary bile acid hyodeoxycholic acid. (* $p < 0.05$ vs. control diet; # $p < 0.05$ as compared to no stress; $ $p < 0.05$ control stress vs. Test stress).

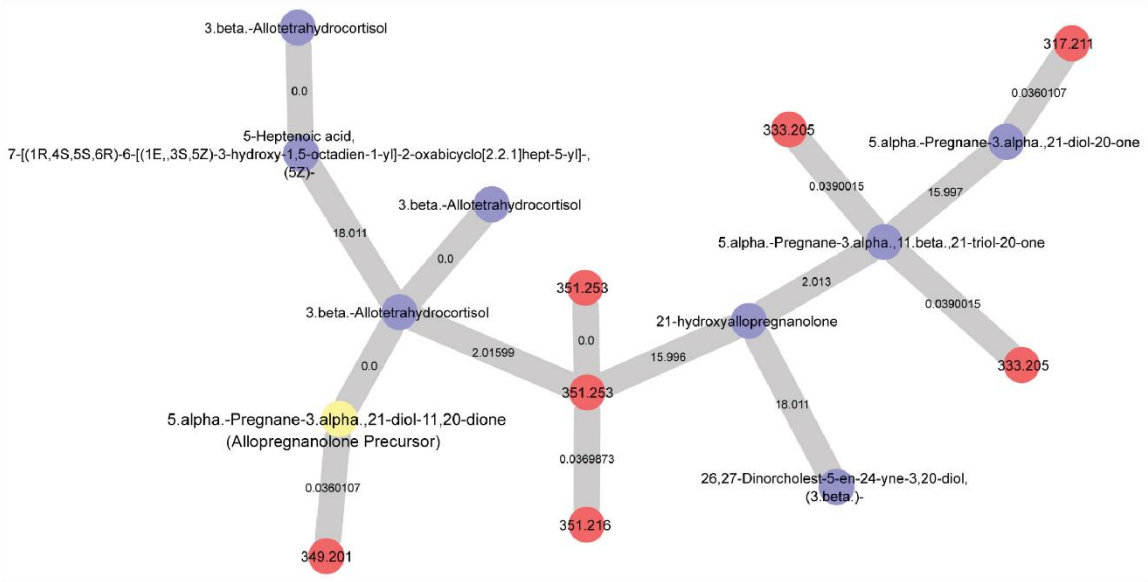


Figure 5.6. Molecular network analysis reveals several metabolites that are structurally related to the Allopregnanolone precursor (yellow node). Nodes are labeled according to their precursor mass or *m/z* (red nodes) or their GNPS spectral library match (blue nodes). Nodes are connected via edges where they are labeled based on the *m/z* difference between nodes.

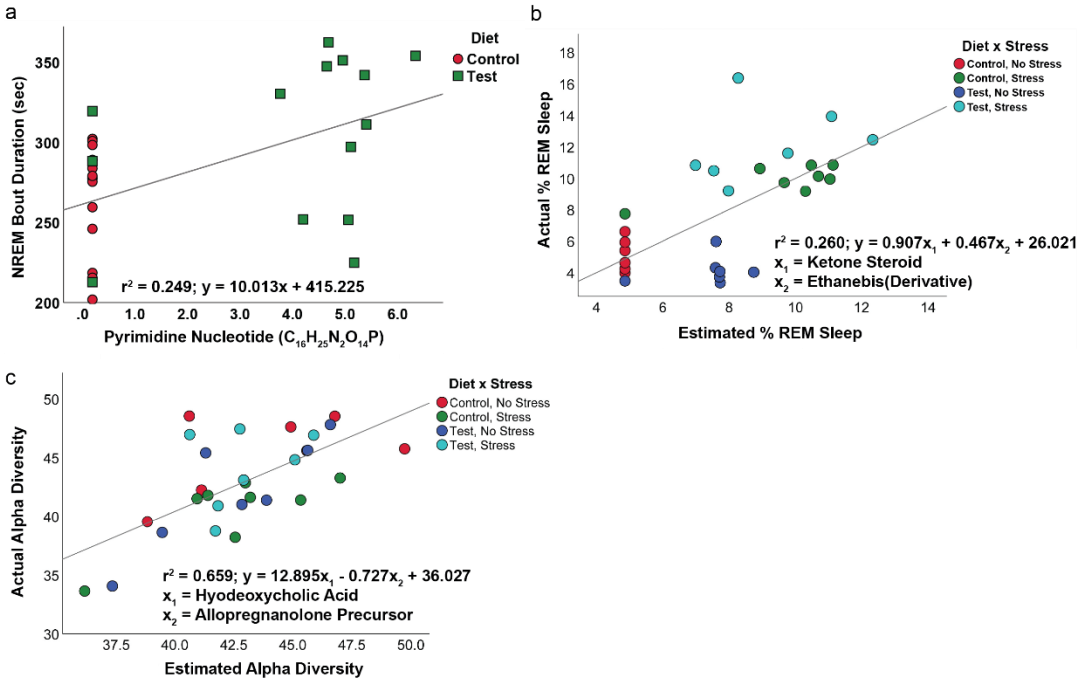


Figure 5.7. (**a**) Regression analysis demonstrating a significant, linear relationship between the Pyrimidine Nucleotide and NREM sleep bout duration at PND 70. (**b**) Regression demonstrating a significant relationship between the stress responsive Ketone Steroid + the Ethanebis(thioate) Derivative and REM sleep rebound after stress exposure. (**c**) Regression demonstrating a significant relationship between alpha diversity (PD_Whole_Tree) and Hyodeoxycholic acid + the Allopregnanolone Precursor.

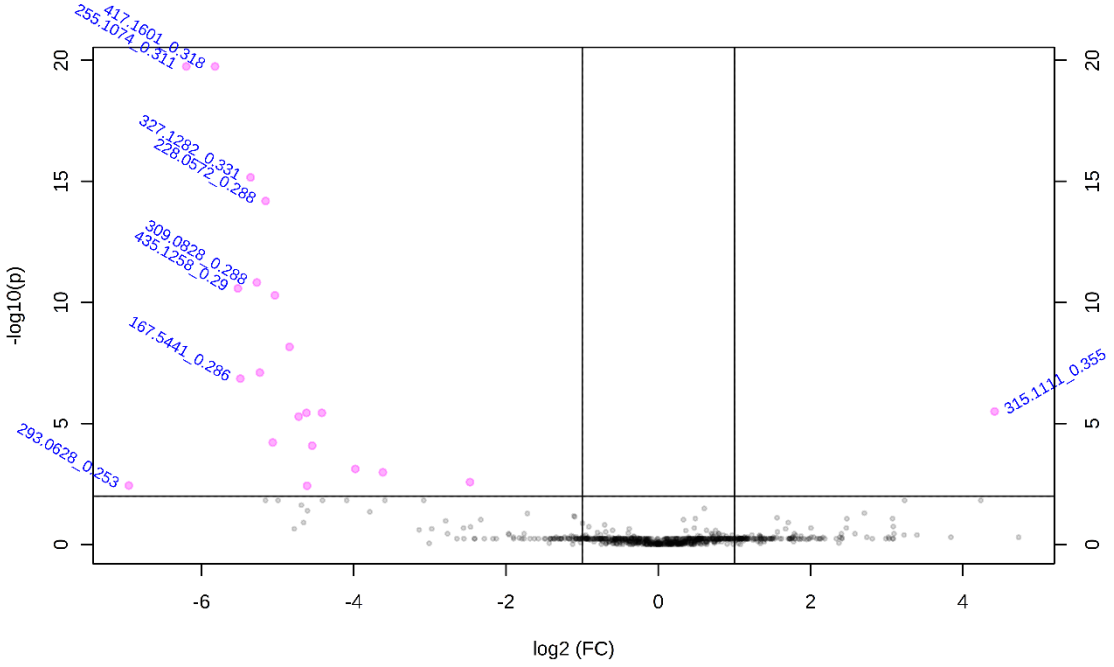


Figure 5.S1. Volcano plot demonstrating the features that are significantly different in control vs. Test diet groups measured in fecal samples collected on PND 70 ($p < 0.05$; FDR $p < 0.01$).

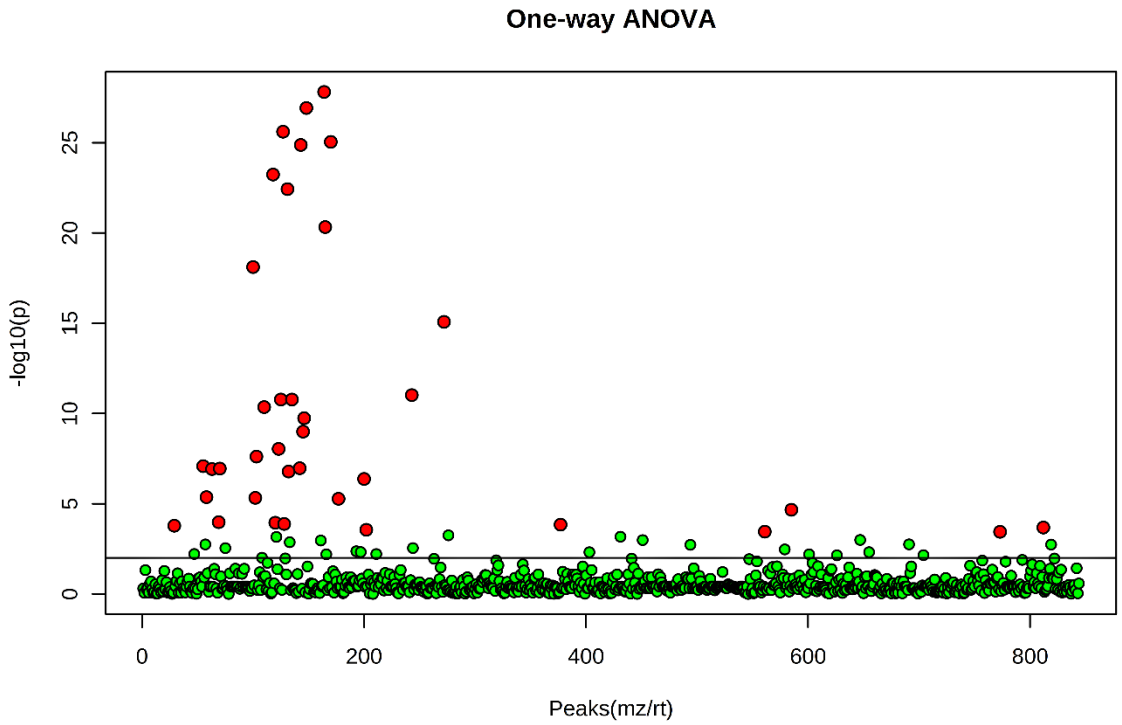


Figure 5.S2. One-way ANOVA demonstrating that 36 features were significantly different (red vs. green) between groups at PND 91 ($p < 0.05$; FDR < 0.01).

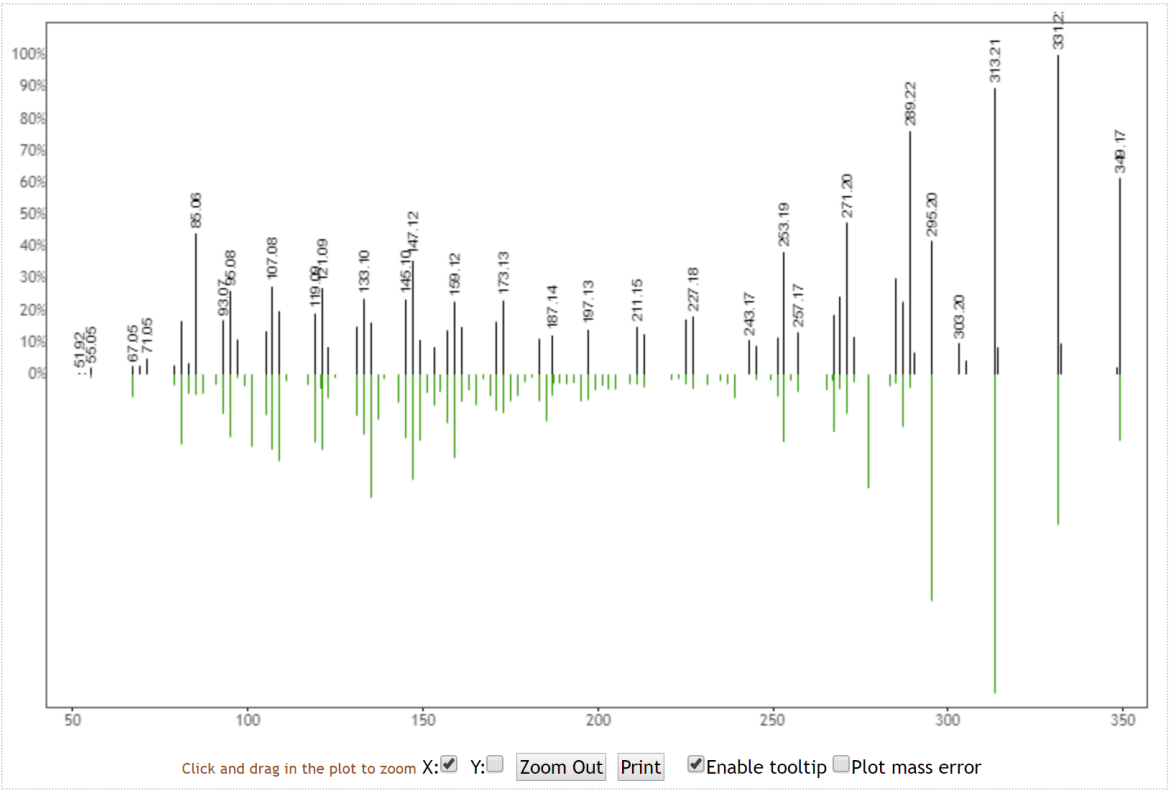


Figure 5.S3. Direct MS/MS comparison between the Allopregnanolone Precursor (*m/z* 349.2371 rt 4.08) vs. the GNPS spectral library reference of 5.alpha.-Pregnane-3.alpha.,21-diol-11,20-dione. The precursor ppm error between our feature and the reference standard was below 2 ppm. The tandem MS^2 spectra of our unidentified feature and reference library shows an overlap of 46 peaks.

Table 5.1. Metabolites that were identified using SIRIUS and CSIfinger ID. See additional classification specifications for these features in Supplemental Table 1.

CSIfinger ID predicted family	Figure	row ID	row m/z	retention time	MF(SIRIUS)	Tree Score	ISOTOPE	Peaks	MSI
Fatty Acid Derivative	5m	69	190.1436725	0.354737288	$C_9H_{19}NO_3$	109.0061428	3.215039989	35	3
Glycerol Glucoside Derivative	4a, 5a	110	255.1077526	0.329195152	$C_9H_{18}O_8$	176.2721696	5.50958751	46	3
Ethanebis(thioate) Derivative	4b, 5b, 6c	20	293.0628507	0.255751818	$C_{10}H_{16}N_2O_4S_2$	151.43	0	34	3
Quinazolinone Analog	4c, 5l	1678	315.1109357	0.354012147	$C_{15}H_{14}N_4O_4$	149.8741117	0	44	3
Ketone Abietic-type Acid	5n	38	315.1948988	5.966663559	$C_{20}H_{26}O_3$	47.63533148	0	25	3
Ketone Steroid	5k, 6c	175	317.2469707	5.065049718	$C_{21}H_{32}O_2$	125.0395243	3.410851867	51	3
Disaccharide	4d, 5c	114	327.1282495	0.333138636	$C_{12}H_{22}O_{10}$	182.4499773	3.476393104	54	3
Pyrimidine Nucleotide	4e, 5d	58	339.060153	0.280809322	$C_{10}H_{15}N_2O_9P$	77.00574093	2.785892451	40	3
Aminoglycoside Analog	4f, 5e	244	342.1397425	0.305590805	$C_{12}H_{23}NO_{10}$	151.636548	0	55	3
Allopregnanolone Precursor	5j, 6d	379	349.2369875	4.085015819	$C_{21}H_{32}O_4$	189.93	3.8	55	3
Disaccharide Derivative	4g, 5f	160	417.1602149	0.319994737	$C_{15}H_{28}O_{13}$	191.2993464	3.629846586	52	3
Oligopeptide	5g	107	434.1887824	0.399667232	$C_{16}H_{27}N_5O_9$	53.33577828	3.458464457	36	3
Aminonucleoside Analog	4h, 5h	222	489.1809597	0.334691441	$C_{20}H_{24}N_8O_7$	171.5941177	2.036780536	53	3
Pyrimidine Nucleotide	4i, 5i, 6b	128	501.1125744	0.27855787	$C_{16}H_{25}N_2O_{14}P$	74.72	0	48	3
Phosphotyrosine Dipeptide	4j	328	506.2087682	0.332613333	$C_{24}H_{32}N_3O_7P$	151.2632419	0	51	3

Table 5.S1. Details of LC-MS/MS parameters used for mzMine.

Mzmine Parameters
Mass Detection
MS level: 1
Noise level: 2E5
Chromatogram builder
Min time span (min): 0.05
Min height: 6E5
m/z tolerance: 0.0005 m/z or 10 ppm
Chromatogram deconvolution
Local min search
Chromatogram threshold: 0.01%
Search minimum in RT (min): 0.20
Minimum relative height: 0.01%
Minimum absolute height: 6E5
Min ratio of peak top/edge: 3
Peak duration range (min): 0.05-0.30
Isotope peaks grouper
m/z tolerance: 0.005 m/z or 10 ppm
Retention time tolerance (min): 0.01
Maximum charge: 4
Join aligner
m/z tolerance: 0.0005 m/z or 10 ppm
Weight for m/z: 90
Retention time tolerance (min): 0.3
Weight for RT: 10
Peak list row filter
Minimum peaks in a row: 2
Minimum peaks in isotope pattern: 2
Reset the peak ID: on

H. References

1. Palagini L, Moretto U, Novi M, Masci I, Caruso D, Drake CL, Riemann D. Lack of Resilience Is Related to Stress-Related Sleep Reactivity, Hyperarousal, and Emotion Dysregulation in Insomnia Disorder. *J Clin Sleep Med*. 2018 May 15;**14**(5):759–66.

2. Palagini L, Petri E, Novi M, Caruso D, Moretto U, Riemann D. Adult insecure attachment plays a role in hyperarousal and emotion dysregulation in Insomnia Disorder [Internet]. **Vol. 262**, *Psychiatry Research*. 2018. p. 162–7. Available from: http://dx.doi.org/10.1016/j.psychres.2018.01.017.

3. Greenwood BN, Thompson RS, Opp MR, Fleshner M. Repeated Exposure to Conditioned Fear Stress Increases Anxiety and Delays Sleep Recovery Following Exposure to an Acute Traumatic Stressor [Internet]. **Vol. 5**, *Frontiers in Psychiatry*. 2014. Available from: http://dx.doi.org/10.3389/fpsyt.2014.00146.

4. Thompson RS, Roller R, Greenwood BN, Fleshner M. Wheel running improves REM sleep and attenuates stress-induced flattening of diurnal rhythms in F344 rats. *Stress*. 2016 May;**19**(3):312–24.

5. Maslanik T, Tannura K, Mahaffey L, Loughridge AB, Benninson L, Ursell L, Greenwood BN, Knight R, Fleshner M. Correction: Commensal Bacteria and MAMPs Are Necessary for Stress-Induced Increases in IL-1β and IL-18 but Not IL-6, IL-10 or MCP-1 [Internet]. **Vol. 8**, *PLoS ONE*. 2013. Available from: http://dx.doi.org/10.1371/annotation/c7a5bcbd-3464-4e9e-bf74-e1aafe3f685b.

6. Tetel MJ, de Vries GJ, Melcangi RC, Panzica G, O'Mahony SM. Steroids, stress and the gut microbiome-brain axis. *J Neuroendocrinol* [Internet]. 2018 Feb;**30**(2). Available from: http://dx.doi.org/10.1111/jne.12548.

7. Thompson RS, Roller R, Mika A, Greenwood BN, Knight R, Chichlowski M, Berg BM, Fleshner M. Dietary Prebiotics and Bioactive Milk Fractions Improve NREM Sleep, Enhance REM Sleep Rebound and Attenuate the Stress-Induced Decrease in Diurnal Temperature and Gut Microbial Alpha Diversity. *Front Behav Neurosci*. 2016;**10**:240.

8. Thompson RS, Christianson JP, Maslanik TM, Maier SF, Greenwood BN, Fleshner M. Effects of stressor controllability on diurnal physiological rhythms. *Physiol Behav*. 2013 Mar 15;**112-113**:32–9.

9. Mika A, Fleshner M. Early-life exercise may promote lasting brain and metabolic health through gut bacterial metabolites. *Immunol Cell Biol*. 2016 Feb;**94**(2):151–7.

10. Greenwood BN, Foley TE, Day HEW, Campisi J, Hammack SH, Campeau S, Maier SF, Fleshner M. Freewheel running prevents learned helplessness/behavioral

depression: role of dorsal raphe serotonergic neurons. *J Neurosci*. 2003 Apr 1;**23**(7):2889–98.

11. Mika A, Day HEW, Martinez A, Rumian NL, Greenwood BN, Chichlowski M, Berg BM, Fleshner M. Early life diets with prebiotics and bioactive milk fractions attenuate the impact of stress on learned helplessness behaviours and alter gene expression within neural circuits important for stress resistance. *Eur J Neurosci*. 2017 Feb;**45**(3):342–57.

12. Bonaz B, Bazin T, Pellissier S. The Vagus Nerve at the Interface of the Microbiota-Gut-Brain Axis. **Front Neurosci**. 2018 Feb 7;**12**:49.

13. Raybould HE. Gut chemosensing: interactions between gut endocrine cells and visceral afferents. *Auton Neurosci*. 2010 Feb 16;**153**(1-2):41–6.

14. Hoyles L, Snelling T, Umlai U-K, Nicholson JK, Carding SR, Glen RC, McArthur S. Microbiome-host systems interactions: protective effects of propionate upon the blood-brain barrier. *Microbiome*. 2018 Mar 21;**6**(1):55.

15. Bellono NW, Bayrer JR, Leitch DB, Castro J, Zhang C, O'Donnell TA, Brierley SM, Ingraham HA, Julius D. Enterochromaffin Cells Are Gut Chemosensors that Couple to Sensory Neural Pathways. *Cell*. 2017 Jun 29;**170**(1):185–98.e16.

16. Mertens KL, Kalsbeek A, Soeters MR, Eggink HM. Bile Acid Signaling Pathways from the Enterohepatic Circulation to the Central Nervous System. *Front Neurosci*. 2017 Nov 7;**11**:617.

17. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, -M. Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. Proposed minimum reporting standards for chemical analysis [Internet]. **Vol. 3**, *Metabolomics*. 2007. p. 211–21. Available from: http://dx.doi.org/10.1007/s11306-007-0082-2.

18. da Silva RR, Wang M, Nothias L-F, van der Hooft JJJ, Caraballo-Rodríguez AM, Fox E, Balunas MJ, Klassen JL, Lopes NP, Dorrestein PC. Propagating annotations of molecular networks using in silico fragmentation. *PLoS Comput Biol*. 2018 Apr;**14**(4):e1006089.

19. Mika A, Gaffney M, Roller R, Hills A, Bouchet CA, Hulen KA, Thompson RS, Chichlowski M, Berg BM, Fleshner M. Feeding the developing brain: Juvenile rats fed diet rich in prebiotics and bioactive milk fractions exhibit reduced anxiety-related behavior and modified gene expression in emotion circuits [Internet]. **Vol. 677**, *Neuroscience Letters*. 2018. p. 103–9. Available from: http://dx.doi.org/10.1016/j.neulet.2018.01.052.

20. Rhoades SD, Sengupta A, Weljie AM. Time is ripe: maturation of metabolomics in chronobiology. *Curr Opin Biotechnol*. 2017 Feb;**43**:70–6.

21. Kokate TG, Svensson BE, Rogawski MA. Anticonvulsant activity of neurosteroids: correlation with gamma-aminobutyric acid-evoked chloride current potentiation. *J Pharmacol Exp Ther*. 1994 Sep;**270**(3):1223–9.

22. Trachsel L, Dodt HU, Zieglgänsberger W. The intrinsic optical signal evoked by chiasm stimulation in the rat suprachiasmatic nuclei exhibits GABAergic day-night variation. *Eur J Neurosci*. 1996 Feb;**8**(2):319–28.

23. Crowley SK, O'Buckley TK, Schiller CE, Stuebe A, Morrow AL, Girdler SS. Blunted neuroactive steroid and HPA axis responses to stress are associated with reduced sleep quality and negative affect in pregnancy: a pilot study. *Psychopharmacology* . 2016 Apr;**233**(7):1299–310.

24. Morris KD, Moorefield CN, Amin J. Differential modulation of the gamma-aminobutyric acid type C receptor by neuroactive steroids. *Mol Pharmacol*. 1999 Oct;**56**(4):752–9.

25. Purdy RH, Morrow AL, Moore PH Jr, Paul SM. Stress-induced elevations of gamma-aminobutyric acid type A receptor-active steroids in the rat brain. *Proc Natl Acad Sci U S A*. 1991 May 15;**88**(10):4553–7.

26. Reddy DS, Rogawski MA. Stress-induced deoxycorticosterone-derived neurosteroids modulate GABA(A) receptor function and seizure susceptibility. *J Neurosci*. 2002 May 1;**22**(9):3795–805.

27. Maguire J. Neuroactive Steroids and GABAergic Involvement in the Neuroendocrine Dysfunction Associated With Major Depressive Disorder and Postpartum Depression. *Front Cell Neurosci*. 2019 Mar 8;**13**:83.

28. Vincenzetti S, Polzonetti V, Micozzi D, Pucciarelli S. Enzymology of Pyrimidine Metabolism and Neurodegeneration. *Curr Med Chem*. 2016;**23**(14):1408–31.

29. Borbély AA, Tobler I. Endogenous sleep-promoting substances and sleep regulation. *Physiol Rev*. 1989 Apr;**69**(2):605–70.

30. Honda K, Komoda Y, Nishida S, Nagasaki H, Higashi A, Uchizono K, Inoué S. Uridine as an active component of sleep-promoting substance: its effects on nocturnal sleep in rats [Internet]. **Vol. 1**, *Neuroscience Research*. 1984. p. 243–52. Available from: http://dx.doi.org/10.1016/s0168-0102(84)80003-6.

31. Teran-Perez G, Arana-Lechuga Y, Esqueda-Leon E, Santana-Miranda R, Rojas-Zamorano JA, Velazquez Moctezuma J. Steroid Hormones and Sleep Regulation [Internet]. **Vol. 12**, *Mini-Reviews in Medicinal Chemistry*. 2012. p. 1040–8. Available from: http://dx.doi.org/10.2174/138955712802762167.

32. Haack M, Lee E, Cohen DA, Mullington JM. Activation of the prostaglandin system in response to sleep loss in healthy humans: potential mediator of increased spontaneous pain. *Pain*. 2009 Sep;**145**(1-2):136–41.

33. Saksvik-Lehouillier I, Harrison SL, Marshall LM, Tranah GJ, Ensrud K, Ancoli-Israel S, Clemons A, Redline S, Stone KL, Schernhammer ES, Osteoporotic Fractures in Men (MrOS) Study Group. Association of Urinary 6-Sulfatoxymelatonin (aMT6s) Levels and Objective and Subjective Sleep Measures in Older Men: The MrOS Sleep Study. *J Gerontol A Biol Sci Med Sci*. 2015 Dec;**70**(12):1569–77.

34. Xiao Q, Derkach A, Moore SC, Zheng W, Shu X-O, Gu F, Caporaso NE, Sampson JN, Matthews CE. Habitual Sleep and human plasma metabolomics. *Metabolomics* [Internet]. 2017 May;**13**(5). Available from: http://dx.doi.org/10.1007/s11306-017-1205-z.

35. Weljie AM, Meerlo P, Goel N, Sengupta A, Kayser MS, Abel T, Birnbaum MJ, Dinges DF, Sehgal A. Oxalic acid and diacylglycerol 36:3 are cross-species markers of sleep debt. *Proc Natl Acad Sci U S A*. 2015 Feb 24;**112**(8):2569–74.

36. Taylor SA, Green RM. Bile Acids, Microbiota, and Metabolism. *Hepatology*. 2018 Oct;**68**(4):1229–31.

37. Mroz MS, Lajczak NK, Goggins BJ, Keely S, Keely SJ. The bile acids, deoxycholic acid and ursodeoxycholic acid, regulate colonic epithelial wound healing. *Am J Physiol Gastrointest Liver Physiol*. 2018 Mar 1;**314**(3):G378–87.

38. Charach G, Argov O, Geiger K, Charach L, Rogowski O, Grosskopf I. Diminished bile acids excretion is a risk factor for coronary artery disease: 20-year follow up and long-term outcome. *Therap Adv Gastroenterol*. 2018;**11**:1756283X17743420.

39. Menni C, Lin C, Cecelja M, Mangino M, Matey-Hernandez ML, Keehn L, Mohney RP, Steves CJ, Spector TD, Kuo C-F, Chowienczyk P, Valdes AM. Gut microbial diversity is associated with lower arterial stiffness in women. *Eur Heart J*. 2018 Jul 1;**39**(25):2390–7.

40. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun*. 2017 Dec 5;**8**(1):1784.

41. Hagiwara T, Saito S, Ujiie Y, Imai K, Kakuta M, Kadota K, Terada T, Sumikoshi K, Shimizu K, Nishi T. HPLC Retention time prediction for metabolome analysis [Internet]. **Vol. 5**, *Bioinformation*. 2010. p. 255–8. Available from: http://dx.doi.org/10.6026/97320630005255.

42. Cardelle-Cobas A, Corzo N, Olano A, Peláez C, Requena T, Ávila M. Galactooligosaccharides derived from lactose and lactulose: influence of structure on Lactobacillus, Streptococcus and Bifidobacterium growth. *Int J Food Microbiol*. 2011 Sep 1;**149**(1):81–7.

43. Herfel TM, Jacobi SK, Lin X, Fellner V, Walker DC, Jouni ZE, Odle J. Polydextrose enrichment of infant formula demonstrates prebiotic characteristics by altering intestinal microbiota, organic acid concentrations, and cytokine expression in suckling piglets. *J Nutr*. 2011 Dec;**141**(12):2139–45.

44. Saulnier DM, Ringel Y, Heyman MB, Foster JA, Bercik P, Shulman RJ, Versalovic J, Verdu EF, Dinan TG, Hecht G, Guarner F. The intestinal microbiome, probiotics and prebiotics in neurogastroenterology. *Gut Microbes*. 2013 Jan;**4**(1):17–27.

45. Schwab C, Gänzle M. Lactic acid bacteria fermentation of human milk oligosaccharide components, human milk oligosaccharides and galactooligosaccharides. *FEMS Microbiol Lett*. 2011 Feb;**315**(2):141–8.

46. Gibson GR, Hutkins R, Sanders ME, Prescott SL, Reimer RA, Salminen SJ, Scott K, Stanton C, Swanson KS, Cani PD, Verbeke K, Reid G. Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nat Rev Gastroenterol Hepatol*. 2017 Aug;**14**(8):491–502.

47. Alexander DB, Iigo M, Yamauchi K, Suzui M, Tsuda H. Lactoferrin: an alternative view of its role in human biological fluids. *Biochem Cell Biol*. 2012 Jun;**90**(3):279–306.

48. León-Sicairos N, Reyes-López M, Ordaz-Pichardo C, de la Garza M. Microbicidal action of lactoferrin and lactoferricin and their synergistic effect with metronidazole in Entamoeba histolytica. *Biochem Cell Biol*. 2006 Jun;**84**(3):327–36.

49. Clare DA, Zheng Z, Hassan HM, Swaisgood HE, Catignani GL. Antimicrobial properties of milkfat globule membrane fractions. *J Food Prot*. 2008 Jan;**71**(1):126–33.

50. Berding K, Wang M, Monaco MH, Alexander LS, Mudd AT, Chichlowski M, Waworuntu RV, Berg BM, Miller MJ, Dilger RN, Donovan SM. Prebiotics and Bioactive Milk Fractions Affect Gut Development, Microbiota, and Neurotransmitter Expression in Piglets. *J Pediatr Gastroenterol Nutr*. 2016 Dec;**63**(6):688–97.

51. Bhinder G, Allaire JM, Garcia C, Lau JT, Chan JM, Ryz NR, Bosman ES, Graef FA, Crowley SM, Celiberto LS, Berkmann JC, Dyer RA, Jacobson K, Surette MG, Innis SM, Vallance BA. Milk Fat Globule Membrane Supplementation in Formula Modulates the Neonatal Gut Microbiome and Normalizes Intestinal Development. *Sci Rep*. 2017 Mar 28;**7**:45274.

52. Greenwood BN, Foley TE, Burhans D, Maier SF, Fleshner M. The consequences of uncontrollable stress are sensitive to duration of prior wheel running. *Brain Res*. 2005 Feb 8;**1033**(2):164–78.

53. Speaker KJ, Cox SS, Paton MM, Serebrakian A, Maslanik T, Greenwood BN, Fleshner M. Six weeks of voluntary wheel running modulates inflammatory protein (MCP-1, IL-6, and IL-10) and DAMP (Hsp72) responses to acute stress in white adipose tissue of lean rats. *Brain Behav Immun*. 2014 Jul;**39**:87–98.

54. Thompson RS, Strong PV, Fleshner M. Physiological consequences of repeated exposures to conditioned fear. *Behav Sci* . 2012 Jun;**2**(2):57–78.

55. Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ, Huntley J, Ackermann GL, Berg-Lyons D, Holmes S, Caporaso JG, Knight R. Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol*. 2013;**531**:371–444.

56. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea [Internet]. **Vol. 6**, *The ISME Journal*. 2012. p. 610–8. Available from: http://dx.doi.org/10.1038/ismej.2011.139.

57. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010 May;**7**(5):335–6.

58. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010 Oct 1;**26**(19):2460–1.

59. Faith DP. Phylogenetic pattern and the quantification of organismal biodiversity. *Philos Trans R Soc Lond B Biol Sci*. 1994 Jul 29;**345**(1311):45–58.

60. Melnik AV, da Silva RR, Hyde ER, Aksenov AA, Vargas F, Bouslimani A, Protsyuk I, Jarmusch AK, Tripathi A, Alexandrov T, Knight R, Dorrestein PC. Coupling Targeted

and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples. *Anal Chem*. 2017 Jul 18;**89**(14):7549–59.

61. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak M-Y, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*. 2012 Oct;**30**(10):918–20.

62. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*. 2010 Jul 23;**11**:395.

63. Scheubert K, Hufsky F, Petras D, Wang M, Nothias L-F, Dührkop K, Bandeira N, Dorrestein PC, Böcker S. Significance estimation for large scale metabolomics annotations by spectral matching. *Nat Commun*. 2017 Nov 14;**8**(1):1494.

64. Wang J, Guo R, Yang Y, Jacobs B, Chen S, Iwuchukwu I, Gaines KJ, Chen Y, Simman R, Lv G, Wu K, Bihl JC. The Novel Methods for Analysis of Exosomes Released from Endothelial Cells and Endothelial Progenitor Cells [Internet]. **Vol. 2016**, *Stem Cells International*. 2016. p. 1–12. Available from: http://dx.doi.org/10.1155/2016/2639728.

65. Böcker S, Dührkop K. Fragmentation trees reloaded. *J Cheminform*. 2016 Feb 1;**8**:5.

66. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, Dorrestein PC, Rousu J, Böcker S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods*. 2019 Apr;**16**(4):299–302.

67. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res*. 2018 Jul 2;**46**(W1):W486–94.

68. Chong J, Xia J. MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data [Internet]. **Vol. 34**, **Bioinformatics**. 2018. p. 4313–4. Available from: http://dx.doi.org/10.1093/bioinformatics/bty528.

69. Xia J, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0—making metabolomics more meaningful [Internet]. **Vol. 43**, *Nucleic Acids Research*. 2015. p. W251–7. Available from: http://dx.doi.org/10.1093/nar/gkv380.

70. Xia J, Wishart DS. Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. *Curr Protoc Bioinformatics*. 2011 Jun;**Chapter 14**:Unit 14.10.

71. Xia J, Wishart DS. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat Protoc*. 2011 Jun;**6**(6):743–60.

Chapter VI

Repeated sleep disruption in mice leads to persistent shifts in the fecal microbiome and metabolome

A. Abstract

It has been established in recent years that the gut microbiota plays a role in health and disease, potentially via alterations in metabolites that influence host physiology. Although sleep disruption and gut dysbiosis have been associated with many of the same diseases, studies investigating the gut microbiome in the context of sleep disruption have yielded inconsistent results and have not assessed the fecal metabolome. We exposed mice to five days of sleep disruption followed by four days of *ad libitum* recovery sleep and assessed the fecal microbiome and fecal metabolome at multiple timepoints using 16S rRNA gene amplicons and untargeted LC-MS/MS mass spectrometry. We found global shifts in both the microbiome and metabolome in the sleep-disrupted group on the second day of recovery sleep, when most sleep parameters had recovered to baseline levels. We observed an increase in the *Firmicutes:Bacteroidetes* ratio, along with decreases in the genus *Lactobacillus*, phylum *Actinobacteria*, and genus *Bifidobacterium* in sleep-disrupted mice compared to control mice. The latter two taxa remained low at the fourth day post-sleep disruption. We also identified multiple classes of fecal metabolites that were differentially abundant in sleep-disrupted mice, some of which are physiologically relevant and commonly influenced by the microbiota. This included bile acids, and inference of microbial functional gene content suggested reduced levels of the microbial bile salt hydrolase gene in sleep-disrupted mice. Thus, repeated sleep disruption causes changes in key features of the fecal microbiome and metabolome, some of which last for days after recovery of objective sleep measures.

1. Significance Statement

Growing evidence suggests that the fecal metabolome acts as a functional readout of the gut microbiota, and the molecules produced or modified by the microbiota can influence host physiology. Repeated sleep disruption has been associated with a multitude of metabolic, psychiatric, and neurological pathologies. Here we report that five days of sleep disruption alters both the fecal microbiome and metabolome in mice. This included reductions in beneficial bacteria and in microbially modified metabolites like bile acids. Overall, this study adds to the evidence base linking disrupted sleep to the gut microbiota and expands it to the fecal metabolome, identifying sleep disruption-sensitive bacterial taxa and classes of metabolites that may serve as therapeutic targets to improve health after poor sleep.

B. Introduction

Inadequate sleep can lead to metabolic (1), immunologic (2,3), and cognitive deficits (4). Many of the pathological states that arise from sleep disruption also occur in conjunction with gut dysbiosis, defined as a disruption of the community structure of the gut microbiota. This includes metabolic disease (5-7) and cognitive impairment (8,9) as well as other proinflammatory and neuro-behavioral disorders such as multiple sclerosis (10), depression (11), anxiety (9,12), and posttraumatic stress disorder (13). This has led to the hypothesis that there is a relationship between inadequate sleep and the gut microbiota. Only a small number of studies have tested this hypothesis, using heterogeneous sleep disruption protocols (e.g., acute sleep restriction (14), chronic sleep fragmentation (15)), in humans (14,16), mice (15), and rats (16), and have yielded mixed results (16). More research is therefore required to explore

the relationship between sleep, the gut microbiota, and potential mediators of microbe-host interactions.

Despite the mounting evidence that supports an important role for the gut microbiota in normal physiology, the mechanisms by which commensal microorganisms influence the host are still unclear. Proposed mechanisms include direct interactions with the enteric nervous system (17), interactions with toll-like receptors in the intestinal epithelium (18), regulation of the immune system19, and signaling of microbially-modified metabolites including those originating from food sources and host bile acids (20-25). These metabolites serve as a functional measure of microbial activity, and the fecal metabolome closely reflects the composition of the fecal microbiome (26). Therefore, to understand the impacts of the microbiota on the host, it is crucial to study not only the microbiota, but also to examine the molecules that they produce and that are present in their microenvironment. However, there have been no studies to date examining the effects of sleep disruption on the fecal metabolome using untargeted metabolomics.

We thus investigated the impact of a sub-chronic, five-day sleep disruption protocol on the fecal microbiome and fecal metabolome in mice. Assessment of the fecal microbiome using 16S rRNA gene amplicons and of the fecal metabolome using untargeted LC-MS/MS mass spectrometry revealed a global shift in both the microbiome and metabolome after sleep disruption, and aspects of these changes persisted through the fourth day after returning to *ad libitum* sleep. Furthermore, microbial differential abundance testing and utilization of Global Natural Products Social Molecular Networking (27) (GNPS) allowed us to identify specific taxa of bacteria and families of metabolites that change in response to five days of sleep

disruption, many of which have known physiological relevance. These findings support the hypothesis that gut dysbiosis, and changes in the fecal metabolome, after sleep disruption may contribute to some of the health problems long known to be associated with inadequate sleep and that these changes may be present even after the sleep-wake state is normalized.

C. Materials and Methods

1. Animals

Seven-week old male C57BL/6N mice (Experiment 1, N = 7; Experiment 2, N = 20; Charles River Laboratories, USA) were used in these experiments. Mice were group-housed upon arrival for one week until surgery (Experiment 1) or until being placed into individual sleep disruption chambers (Experiment 2). Mice were maintained on a 12:12 L:D cycle at room temperature (23 °C + 2 °C) with food and water available *ad libitum* throughout the experiment. The light source was two 14 W fluorescent tubes (soft white, 3000 K), resulting in an average light intensity of ~500 lux inside the cylindrical sleep disruption cage. Zeitgeber Time (ZT) is defined as the number of hours after the onset of the light period (light onset = ZT0). All mice were housed and handled according to the Federal Animal Welfare guidelines, and all studies were approved in advance by the Institutional Animal Care and Use Committee at Northwestern University.

2. Experiment 1. Effects of a 5-day sleep disruption protocol on measures of sleep physiology and sleep recovery

In Experiment 1 we subjected adult male C57BL/6N mice to a five-day protocol whereby a slowly rotating horizontal metal bar (one pass every 7-8 seconds) at the bottom of the cage disrupted sleep for 20 h/day. The direction of rotation was programmed to reverse at a semi-random (unpredictable) interval from 0-20 seconds. The animals were allowed to sleep *ad libitum* from ZT2-ZT6 (Zeitgeber Time; ZT0 is the onset of light during the 12:12 light/dark cycle by convention), and control animals were left undisturbed in cages with a motionless metal bar. We recorded sleep in mice before, during, and after the sleep disruption protocol using implanted electroencephalogram (EEG)/electromyogram (EMG) (**Figure 6.1a**).

3. Experiment 2. Effects of a 5-day sleep disruption protocol on the fecal microbiome and metabolome

In Experiment 2, a second group of mice was subjected to an identical sleep disruption protocol, and fecal samples were collected before sleep disruption (Pre-SD), at day two post-sleep disruption (SD+2), and at day four post-sleep disruption (SD+4) (**Figure 6.1b**) to assess the fecal microbiome and fecal metabolome.

4. Sleep Recording and Analysis

One week after arrival, mice for Experiment 1 were implanted with electroencephalographic/electromyographic (EEG/EMG) sleep recording devices (Pinnacle Technologies, Lawrence, KS, USA). Surgical procedures were performed using a mouse stereotaxic apparatus with standard aseptic techniques in a ventilated, specially equipped surgical suite. Anesthesia was induced by IP injection of cocktail of ketamine HCl (98 mg/kg;

Vedco Inc, St. Joseph, MO, USA) and xylazine (10 mg/kg; Akorn Inc, Lake Forest, IL, USA) before surgical implantation of a headmount, which consisted of a plastic 6-pin connector connected to four EEG electrodes and two EMG electrodes. Four stainless steel screws serving as anchors for the EEG leads and grounds were screwed into the skull with one screw located 1 mm anterior to bregma and 2 mm lateral to the central suture, and the other at 1 mm anterior to lambda and 2.5 mm lateral to the central suture. The exposed ends of two stainless steel Teflon-coated wires (0.002 in. in diameter) serving as EMG leads were then inserted into the nuchal muscles using a pair of forceps. The headmount was then sealed by dental acrylic and a single suture at the front and back of the implant was given to close the skin. Subcutaneous injection of analgesic meloxicam (2 mg/kg; Norbrook Laboratories, Newry, Northern Ireland) was given to the animals immediately after the surgery while the animals were still under anesthesia and once more on the following day.

One week after implant surgery, mice were moved into cylindrical sleep recording cages (25 cm in diameter and 20 cm tall, Pinnacle Technologies) within individual acoustically isolated chambers and the headmount was connected to the transmission tether. Cages had corncob bedding and food/water available *ad libitum*. Two days were allowed for acclimation to the tether before baseline sleep was recorded. Sleep was recorded continuously throughout the remainder of the experiment. Data were collected using Pinnacle Acquisition software (Pinnacle Technologies), then scored as non-rapid eye movement sleep (NREM), rapid eye movement sleep (REM), or Wake in 10 second epochs using machine learning-assisted sleep scoring software developed in the Turek/Vitaterna laboratory (28). The initiation of a bout of NREM, REM, or Wake was defined by the occurrence of two consecutive epochs of NREM,

REM, or Wake (respectively). A bout was terminated when a bout of another state occurred. Sleep bouts were initiated by two consecutive epochs of a sleep state (NREM or REM) and were only terminated when a wake bout occurred. The delta power band was defined as 0.5-4 Hz. Relative power was calculated as the raw power (uV^2) in a particular band divided by the total power in all bands. Power was then reported as a percent of baseline to reduce inter-individual variability.

5. Sleep Disruption Protocol

The same sleep disruption protocol was used in Experiment 1 and Experiment 2. Prior to the sleep disruption protocol, all mice were transferred from their home cages into individual sleep disruption cylindrical cages. Cages had corncob bedding and food/water available *ad libitum*. Mice were allowed to acclimate to the chambers for 7 days before beginning the sleep disruption protocol. Sleep disruption was achieved using a commercially available system integrated into the chambers (Pinnacle Technology, Lawrence, KS, USA), which simulates the gentle handling technique via a rotating metal bar (22 cm in length) attached to a post at the center of the cage. For the sleep disruption period, the bar's rotation speed was set at seven rotations per minute with reversals of rotation direction (i.e., clockwise vs. counterclockwise) set to occur at semi-random intervals of 10 ± 10 seconds. The bar was programmed to rotate for 20 hours per day (ZT6-ZT2), and was stationary from ZT2-ZT6, for 5 days. Experimenters visually inspected mice at regular intervals during the sleep disruption windows to ensure that the bar mechanism was functioning properly and that the sleep-disrupted mice were awake. Control animals were placed in identical cages with bars that remained stationary throughout

the experiment. At ZT2 of the fifth sleep disruption day, the motorized bars were stopped, and mice were allowed to sleep ad libitum for the remainder of the experiment.

6. Fecal Sample Collection

In Experiment 2, fecal samples were collected at 3 different timepoints: 1) after mice were transferred to sleep disruption cages but before starting sleep disruption (Pre-SD); 2) on the second afternoon (~30 h) after the sleep disruption protocol was ended (SD+2); and 3) on the fourth afternoon after the sleep disruption protocol was ended (SD+4). At each collection, mice were placed into a clean sleep disruption chamber with fresh bedding and food and monitored closely until two fresh fecal pellets from each mouse were collected. Samples were placed into individual 1.5 mL microfuge tubes, and frozen at -80 ºC until microbiome/metabolome analysis. All fecal pellets were collected between ZT6 and ZT12.

7. Microbiome Analysis

Microbiome data were generally analyzed using the Quantitative Insights Into Microbial Ecology 2 (QIIME2, version 2018.2) bioinformatics software package (29,30). A total of 56 fecal samples (Pre-SD: n = 10/10 Con/SD; SD+2: n = 8/8 Con/SD; SD+4: n = 10/10 Con/SD) were processed for microbiome analyses. DNA was extracted from fecal samples and the V4 region of the 16S rRNA gene was amplified using the 515f/806rB primer pair with the barcode on the forward read (31) and sequenced as previously described (32) using an Illumina MiSeq. Sequence data were processed using Deblur v1.0.2 (33), trimming to 150 nucleotides to create sub-operational-taxonomic-units (sOTUs). These were then inserted into the Greengenes 13_8

(34) 99% reference tree using SATe-enabled Phylogenetic Placement (SEPP) (35). SEPP uses a simultaneous alignment and tree estimation strategy (36) to identify placements for sequence fragments within an existing phylogeny and alignment. Taxonomy was assigned using an implementation of the Ribosomal Database Project (RDP) classifier (37) as implemented in QIIME2 (29).

The OTU feature table was filtered to remove any features present in three or fewer samples (out of the 56 original samples), and alpha and beta diversity metrics were performed at a rarefied depth of 8431 reads, resulting in the removal of five samples from the dataset (final n for diversity metrics: Control/SD - Pre-SD: n = 8/8; SD+2: n = 8/8; SD+4: n = 10/9). Beta diversity was assessed using weighted UniFrac distance (38) matrices, which were used to generate PCoA plots and to perform PERMANOVA in QIIME2. Within-group distance was calculated from distance matrices by averaging the weighted UniFrac distance from an individual sample to all other samples in the same group (Con vs SD) at the same timepoint. Distance from baseline was calculated by averaging the distance from an individual sample at SD+2 or SD+4 to all samples in the same group at the Pre-SD timepoint. Alpha diversity metrics were calculated using scikit-bio 0.5.1 as implemented by QIIME2. To test for differentially abundant taxa between control and sleep-disrupted groups, samples with less than 8000 reads were removed (final n for differential abundance testing: Control/SD - Pre-SD: n = 8/8; SD+2: n = 8/8; SD+4: n = 10/9), and DESeq2 (version 1.14.1) was performed on the non-rarefied dataset at each timepoint and at each taxonomic level using the Bioconductor R package in RStudio (version 1.0.136, RStudio Inc). This was used in favor of the Analysis of the

Composition of Microbiomes (ANCOM) due to the extremely low sensitivity of ANCOM when sample size is less than 20 per group (39).

8. PICRUSt2 Analysis of 16S rRNA gene data

We inferred the microbial gene content from the taxa abundance using the software package Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2; https://github.com/picrust/picrust2; v2.1.4-b) (40). This tool allows assessment of functional capacity of a microbiome using 16S rRNA sequencing data. We then used DESeq2 to identify genes that were differentially abundant between control and sleep-disrupted groups (notated with Enzyme Commission numbers).

9. Metabolomic Analysis

A total of 56 fecal samples (Pre-SD: *n* = 10/10 Con/SD; SD+2: *n* = 8/8 Con/SD; SD+4: *n* = 10/10 Con/SD) were processed for analysis of the fecal metabolome. Fecal samples were analyzed using an ultra-high performance liquid chromatography system coupled to a quadrupole-Orbitrap mass spectrometer (Q Exactive, Thermo Scientific, Waltham, MA, USA). Chromatographic separation was accomplished using a Kinetex C18 1.7 µm, 100 Å pore size, 2.1 mm (internal diameter) x 50 mm (length) column (Phenomenx, Torrance, CA, USA). The column was maintained at 40 °C during chromatographic separation. 5.0 µL of extract was injected per sample. Mobile phase composition was (A) water with 0.1 % formic acid (v/v) and (B) acetonitrile with 0.1 % formic acid (v/v) with a flow rate of 0.5 mL/min. Chromatographic elution was performed as follows: 0.00-0.50 min, 5% B; 0.50 - 4.00 min, 50% B; 4.00 - 5.00

min, 99% B; 5.00 - 7.00 min, 99% B; 7.00 -7.10 min, 5% B; 7.10 - 9.00 min, 5% B. Positive mode electrospray ionization was performed using a heated electrospray ionization source using the following source parameters: spray voltage, 3500 V; capillary temperature, 268.75 °C; sheath gas flow rate, 52.50 (arb. units); auxiliary gas flow rate, 13.75 (arb. units); probe heater temperature, 437.50 °C; and S-lens RF level, 50 (arb. units). Mass spectrometry data were collected using data-dependent acquisition. The MS1 scan range was set to 150-1,500 *m/z* with a resolution of 17,500 at 200 *m/z*. MS2 scans of the five most abundant ions in the previous MS1 scan, acquired in a data-dependent manner, were collected at a resolution of 17,500 at 200 *m/z*. MS1/MS2 automatic gain control target and maximum ion injection time were set to 5.0 E5 and 100 ms respectively. Higher-energy collision-induced dissociation was performed with a normalized collision energy stepped from 20, 30, to 40%.

The LC/MS/MS feature table, generated using Optimus (41) peak detection, was normalized to an internal standard followed by a row sum (total ion count) normalization and filtered to remove features present in less than two samples. The resulting table contained 1124 metabolites. PCoA plots were then generated using Bray-Curtis distance, and PERMANOVA was performed at each timepoint on the normalized feature table using the Vegan package (version 2.5-5) in RStudio. In order to identify metabolites that were different between sleep-disrupted and control groups, we used a multiple-method approach that included machine learning and nonparametric hypothesis testing. In order to first identify the group of metabolites that were the key drivers of differences between groups at each timepoint, Variable Selection Using Random Forests (VSURF, version 1.0.3) (42,43) analysis was performed using the VSURF.R package in RStudio. Briefly, this protocol uses multiple iterations of the random

forest supervised machine learning technique to isolate the most important drivers of separation between two groups by defining a threshold variable importance. Taking this list of suprathreshold features, we then performed Wilcoxon Rank Sum tests at each timepoint as a form of a 'post hoc' test to confirm differences between groups.

Features of interest were annotated using GNPS (version 1.3.0) (27), which allows MS1 and MS2 spectra to be shared between researchers, forming a large database. By matching an unknown spectrum to one or more in the database, and examining similarity to others within a molecular network, GNPS can be used to identify purported molecular structures of features from untargeted metabolomics. A molecular network was created using the online workflow at GNPS. The data were filtered by removing all MS/MS peaks within +/- 17 Da of the precursor *m/z*. MS/MS spectra were window-filtered by choosing only the top six peaks in the +/- 50 Da window throughout the spectrum. The data were then clustered with MS-Cluster with a parent mass tolerance of 0.1 Da and a MS/MS fragment ion tolerance of 0.1 Da to create consensus spectra. Further, consensus spectra that contained less than two spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.6 and more than four matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top ten most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.6 and at least four matched peaks. Results can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=6fb1d63a51764c7ea75a4e7256b6936a.

Individual features of interest from the feature table were then matched to nodes (clusters) in the network whose average *m/z* and RT were within 0.025 and 30 s, respectively, of the feature of interest. Features that were matched to multiple clusters using the aforementioned criteria were assigned to the cluster with the closest average *m/z* and RT.

10. Statistical Analysis and Software

All graphs depict the mean +/- SEM unless otherwise stated. All PCoA plots were generated using the EMPeror visualization tool as implemented in QIIME2 (44). Microbiome data processing and analysis, including PERMANOVA, were performed in QIIME2 as outlined above. Wilcoxon Rank-Sum tests, Kruskal-Wallace tests, VSURF, DESeq2 (with Benjamini Hochberg adjustment), heatmaps, and boxplots/scatterplots were performed or generated in RStudio (version 1.0.136, RStudio Inc). Two way ANOVA and mixed-effects models with Bonferroni post hoc testing of sleep, alpha diversity, and beta diversity measures, along with generation of all other graphs/figures, was performed using GraphPad PRISM (version 8.2.1; GraphPad Inc, San Diego, CA, USA). Test statistics generated by PERMANOVA, ANOVA, and mixed-effects models are reported in **Table 6.S1**.

11. Data Availability

Sequencing data and metadata are available on Qiita (45) under study ID 10777 and on EBI-ENA with accession number EBI: ERP113564 (46). The metabolomics dataset is publicly available in the MassIVE database under accession number MSV000080630

(https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=8f3141b17a1e4b5886df0d4c515f2a16) (47).

D. Results and Discussion

1. Experiment 1: The five-day sleep disruption protocol significantly reduces and fragments sleep

In Experiment 1, we performed a detailed analysis of sleep before, during, and after the sleep disruption protocol that was used in Experiment 2. We subjected male C57BL/6N mice to a five-day protocol whereby sleep was disrupted for 20 h/day, with *ad libitum* sleep opportunity from ZT2-ZT6 (see **Methods** and **Figure 6.1a**). Control animals were left undisturbed in cages with a motionless metal bar. We recorded sleep using implanted electroencephalogram (EEG)/electromyogram (EMG). Compared to control animals, sleep-disrupted animals had significantly less total sleep, NREM sleep, and REM sleep per 24 hours during the protocol (**Figure 6.2a-c**, F statistics can be found in **Table 6.S1**). While the amount of 24-hour NREM sleep recovered to the level of controls within the first day of recovery sleep (**Figure 6.2b**), there was a significant rebound in the amount of REM sleep on the first day of recovery sleep (**Figure 6.2c**). In order to observe sleep with greater resolution, we examined the fifth day of the sleep disruption protocol and the beginning of the first day of recovery sleep using two-hour time bins. It was evident that the majority of this 24-hour sleep loss occurred during the hours of the light period in which the motorized sleep disruption bar was moving (ZT0-ZT2 and ZT6-ZT12, **Figure 6.2e-g**). REM sleep was reduced to nearly zero percent while the motorized bar was moving, and this resulted in strong REM rebounds during the first two

hours of the *ad libitum* recovery windows (**Figure 6.2g**). Sleep disruption also resulted in more fragmented sleep. During the five days of sleep disruption, there was a significantly higher number of state changes in sleep-disrupted mice compared to controls (**Figure 6.2d,h**). Furthermore, there was an increase in the number of bouts of sleep and bouts of NREM in the sleep disruption group, accompanied by a significant decrease in the bout length, further suggesting fragmentation (**Figure 6.S1**). The number of REM bouts per 24 hours was significantly decreased on days the motorized bar was on, was significantly increased on the first recovery day, but no longer significantly different from control by the second recovery day (**Figure 6.S1**).

2. Experiment 2: Five days of sleep disruption creates changes in the fecal microbiome that last at least four days after disruption has ended

While previous work has demonstrated chronic (four week) sleep fragmentation (short disruption every two minutes) in mice alters the gut microbiome (15), it remains unknown what impact a sub-chronic, more severe sleep disruption paradigm may have. In Experiment 2, fecal samples were collected before sleep disruption (Pre-SD), at day two post-sleep disruption (SD+2), and at day four post-sleep disruption (SD+4) (see **Methods** and **Figure 6.1b**) to assess the fecal microbiome and fecal metabolome. Beta diversity, or the difference in diversity between two or more communities, was assessed at each experimental time point with weighted UniFrac distance, which takes into account both the abundances and phylogenetic relatedness of two communities (38,48). Principal coordinates analysis (PCoA) revealed no difference between control and sleep-disrupted groups at baseline, as expected (p = 0.877,

PERMANOVA; **Figure 6.3a**), but significant clustering of control mice and sleep-disrupted mice indicated a global difference in community structure at SD+2 (p = 0.018, PERMANOVA; **Figure 3b**) that was gone by SD+4 (p = 0.663, PERMANOVA; **Figure 6.3c**). The distance from baseline, the average weighted UniFrac distance between an individual post-sleep disruption and all individuals within the same group at Pre-SD, was increased at SD+2 and SD+4 (**Figure 6.3d**, **Right panel**). Furthermore, sleep disruption significantly increased the dissimilarity between individuals within the sleep-disrupted group at SD+2 and SD+4 (**Figure 6.3d**, **Left panel**). Therefore, five days of repeated sleep disruption had a "destabilizing" effect in that it not only shifted microbial communities away from controls, it increased dissimilarity within the group, and this effect lasted at least four days after return to *ad libitum* recovery sleep.

Multiple measures of alpha diversity, the microbial diversity within an individual community, were also examined because reductions in alpha diversity have been associated with pathological states such as inflammatory bowel syndrome (49), chronic stress (50), and obesity (51). Faith's phylogenetic diversity index was not affected by sleep disruption (**Figure 6.3e**, **Left**), consistent with results in different sleep disruption models (15). Pielou evenness (52), a measure of how evenly distributed species in a population are, was also unaffected by sleep disruption (**Figure 6.3e**, **Right**).

3. Multiple bacterial taxa are differentially abundant in the sleep-disrupted group

In order to investigate which bacterial taxa were driving the changes in beta diversity seen after sleep disruption, we performed analysis of differential abundance using the DESeq2 R package. Although DESeq2 may not adequately account for the compositionality of microbiome datasets (53), a recent study comparing multiple methods for differential abundance testing revealed DESeq2 to have superior sensitivity and specificity when the per-group $n < 20$, as it was in our case (39). We tested for differential abundance between control and SD groups at each taxonomic level, at each timepoint. Of the 142 originally identified taxa (includes all levels), 0, 16, and 6 were significantly different at the Pre-SD, SD+2, and SD+4 timepoints, respectively (FDR < 0.1, **Figure 6.4a**; **Table 6.S2**). The ratio of the two most prevalent phyla in the mammalian gut, the *Firmicutes:Bacteroidetes* (F:B) ratio, is a blunt measure of community shift. An increase in the F:B ratio has been seen in obesity (54,55), stress (56), as well as models of acute (14) and chronic (15) sleep disruption. We found a significant sleep disruption-induced increase in the F:B ratio (**Figure 6.4b**) that was significant at SD+2 but not at SD+4.

The increase in the F:B ratio was due to a significant increase in the relative abundance of *Firmicutes* at SD+2 (**Figure 6.4c**). Within the **Firmicutes** phylum, two major classes changed in different directions at SD+2. *Bacilli* were significantly decreased in sleep-disrupted mice (**Figure 6.4d**), while *Clostridia* were significantly increased (**Figure 6.4g**). The decrease in the class *Bacilli* appeared to mostly be due to significant decreases in the genus *Lactobacillus* (**Figure 6.4e**) and genus *Turicibacter* (**Figure 6.4f**). Within the *Clostridia* class, one unknown genus within the *Clostridiaceae* family was significantly decreased at SD+2 (**Figure 6.4h**), while other taxa within class *Clostridia* were significantly increased (**Table 6.S2**). The low

abundance phylum *Actinobacteria* (**Figure 6.4i**) was significantly decreased in the sleep-disrupted group at both SD+2 and SD+4. This decrease was evident in the genus *Bifidobacterium* within the *Actinobacteria* phylum. These results parallel the beta diversity findings in that the greatest magnitude of shift in the fecal microbiome was at SD+2, and while some measures recover, others persist into SD+4.

4. Five days of sleep disruption changes the fecal metabolome

Due to the increasing evidence supporting the role microbes play in generating or altering physiologically active metabolites, we examined the impact of sleep disruption on the fecal metabolome. Using untargeted LC-MS/MS mass spectrometry, we assessed the metabolomes of fecal pellets taken from the same mice at the same timepoints as those used for microbiome analysis. Normalized feature tables containing 1124 features were used for PCoA analysis at each timepoint to assess global changes due to sleep disruption. No separation was observed at Pre-SD (p = 0.881, PERMANOVA; **Figure 6.5a**), but a clear separation between sleep-disrupted and control mice was seen at SD+2 (p = 0.007, PERMANOVA; **Figure 6.5b**). This separation was no longer present at SD+4 (p = 0.381, PERMANOVA; **Figure 6.5c**). Of the 1124 molecular features assayed, 250 were identified as significantly changing over time, relative to Pre-SD baseline, in either the control group, the sleep-disrupted group, or both (Kruskal-Wallace FDR < 0.1, **Figure 6.5d**). Many features (101/250) significantly changed only in control animals, suggesting sleep disruption prevented a naturally occurring change. Conversely, 57/250 features significantly changed over time in the sleep-disrupted mice but not in the non-sleep-disrupted mice. We also compared sleep disruption to control groups at each

timepoint individually to assess the relative amount of differentially abundant features at each stage of the experiment, and the majority of significantly differentially abundant features (142/204; Wilcoxon Rank-Sum, $p < 0.05$) were found at SD+2 (**Figure 6.5e**), with 57 of those 142 decreased and 85 of the 142 increased in the sleep-disrupted group. Only 20/204 (13 decreased, 7 increased in the SD group) significantly differentially abundant features were found at Pre-SD, whereas 42/204 (29 decreased, 13 increased in the SD group) significantly differentially abundant features were found at SD+4. Overall, these results indicate that five days of sleep disruption results in a global shift in the fecal metabolome, both preventing naturally occurring shifts in the abundances of some metabolites and creating changes in others. While this global shift is present only at SD+2, some metabolites remained altered on the 4th day of *ad libitum* recovery sleep.

5. A subset of metabolites drive separation between sleep-disrupted and control groups at day two post-sleep disruption

Variable Selection Using Random Forests (VSURF) (42,43) was used to identify features that were important drivers of separation between sleep disruption and control groups at SD+2. VSURF identified 98 features that were above the threshold variable importance (suprathreshold) and that successfully distinguished the two groups on a heatmap (**Figure 6.6a-b**; **Table 6.S3**).

From here we sought to learn about the possible identities of these features of interest using GNPS (27). By matching an unknown spectrum or cluster of spectra to spectra in a large database, and examining their similarity to others within a molecular network, GNPS can be

used to identify molecular classes and annotate purported molecular structures of features from untargeted metabolomic datasets. This is a level 2 or 3 metabolite identification according to the 2007 Metabolomics Standards Initiative (57), where level 1 is considered a high confidence identification. Using GNPS to generate a molecular network for this dataset, the MS2 spectra of 21/98 suprathreshold SD+2 features were matched to annotated spectra, including 4 of the top 25 drivers identified by VSURF (**Figure 6.6b**). Examining only the top annotated features (**Figure 6.6c**), many features with spectral matches to di- and tripeptides, along with the lysine degradation product L-saccharopine, were significantly increased in the sleep-disrupted group (**Figure 6.6d-f**; **Figure 6.S2**).

We also noticed sleep disruption-induced changes in features with spectral matches to bacterially modified molecules including bile acids and urobilin. Two suprathreshold features with spectral matches to the bile acid cholic acid were significantly reduced, and two were significantly increased in sleep-disrupted mice at SD+2 (**Figure 6.6g-h**). Furthermore, two unannotated features with structural similarity to bile acids, as indicated by their presence in the same molecular networks as primary and secondary bile acids, were also significantly reduced (**Figure 6.S3**), indicating structural similarity to bile acids. Bile acids are commonly modified by bacteria in the gut lumen by the enzyme encoded by the gene *bile salt hydrolase* (BSH) (58,59), and have diverse signaling properties that involve the immune (60) and nervous systems (24,61). Thus, we used the software package PICRUSt2 to infer microbial gene content from the 16S rRNA gene data and assess inferred abundance of microbial BSH in our fecal samples. We found the inferred abundance of BSH (EC:3.5.1.24) was significantly reduced in the sleep-disrupted group at SD+2 (**Figure 6.6i**, **Table 6.S4**). These results provide evidence

that microbially modified, physiologically active classes of molecules are impacted by five days of sleep disruption, and that the microbiota have altered functional capacity to produce them.

Another class of molecules that was affected by sleep disruption was dietary-derived pentacyclic triterpenoids. Triterpenoids are plant-derived molecules, and some have been shown to have anti-inflammatory properties (62). We identified a molecular network containing 43 clusters, 12 of which were annotated as pentacyclic triterpenoids or close derivatives (**Figure 6.S4**). Of the 43 clusters in the network, 9 were matched to VSURF suprathreshold features. This includes seven that were suprathreshold at SD+2 (**Figure 6.S4b-h**) and two that were suprathreshold at SD+4 (**Figure 6.S4i,j**). A feature with a spectral match to sumaresinolic acid (**Figure 6.S4b**), along with unannotated feature 645 (**Figure 6.S4g**) were significantly reduced at SD+2. A feature matching corosolic acid (**Figure6. S4c**), along with unannotated features with the ID's 871, 204, 133 and 273 (**Figure 6.S4d-f,h**) were significantly increased at SD+2.

Another molecular network of interest contained two unannotated ions that only appeared in sleep-disrupted groups (**Figure 6.S5**). These metabolites therefore hold potential to act as markers for recent sleep disruption, and future additions to the GNPS database may result in level 2 or 3 annotation of clusters in the network.

6. Some changes to fecal metabolites are present at day four post-sleep disruption

Although no global change was seen on PCoA, we also ran VSURF analysis on the SD+4 feature table and identified 64 suprathreshold metabolites that were able to separate the control and sleep-disrupted groups (**Figure 6.S6a-b**; **Table 6.S3**). Seven of these metabolites

had MS2 spectra that matched reference spectral libraries in GNPS. Among the annotated features were molecules with spectral matches to hederagenin and wilforlide A (**Figure 6.S4i,j**), which were significantly reduced compared to controls at SD+4, and fell into the same molecular network as multiple metabolites that were suprathreshold at SD+2. Others that were significantly increased or decreased at SD+2 compared to controls did not recover by SD+4 (e.g. **Figure 6.S6d-f**). This indicates that, while no global changes were evident by day four of recovery sleep, sleep disruption did have an impact on some individual metabolites that persisted for at least four days.

E. Conclusions

Repeated sleep disruption is ubiquitous in modern society and has been linked to a multitude of health problems. Recent lines of scientific inquiry have established an important role for the gut microbiota in multiple facets of mammalian health and disease, many of which are also affected by sleep disruption. The present study took a detailed look at the impact of a sub-chronic, five-day sleep disruption protocol on the fecal microbiome as well as the fecal metabolome in mice, and found that repeated exposure to inadequate sleep had an impact on the microbiome and metabolome that lasted at least four days after the sleep disruption had ended. Importantly, sleep disruption reduced levels of beneficial bacterial genera, altered the metabolic function of the microbiome, and changed fecal levels of bacterially modified metabolites such as bile acids. These results can provide insights into possible mechanisms by which sleep disruption may impact host physiology.

The protocol used in this study resulted in severely disrupted sleep for five days. While NREM sleep was not completely deprived during the 20 hours per day the motorized bar was rotating, it was decreased and significantly fragmented. This disruption was fairly stable across the five days, indicating that the animals were unable to adapt to the paradigm or find strategies to improve sleep as the protocol went on. The repeated disruptions induced by each pass of the bar rarely allowed the mice to reach REM sleep, resulting in a near-total deprivation of REM during the 20 hours the bar was rotating. The subsequent REM rebound was evident during the first two hours of each four-hour *ad libitum* recovery as well as on the first 24-hour recovery day. This is a relevant model because repeated nights of inadequate sleep followed by a few days of recovery sleep is a common schedule in society today, and short sleep mixed with fragmentation is particularly prominent among groups with demanding work schedules such as on-call physicians (63) and active duty military personnel (64). By the second day of recovery sleep, nearly all sleep parameters had returned to control levels. Whether the specific characteristics of sleep disruption determine how the microbiome or metabolome changes is unknown and warrants further investigation.

Fecal samples taken from sleep-disrupted mice at day two post-sleep disruption (SD+2) had a significantly different microbial community structure than those taken from control mice, but measures of alpha diversity were largely unaffected. Sleep disruption also increased the dissimilarity of the fecal microbiome between animals within the sleep disruption group (within group distance) at both SD+2 and SD+4, indicating a “destabilizing” effect that persisted long into recovery sleep. This is particularly interesting because the sleep-wake pattern appeared to normalize by SD+2.

Changes to particular taxa observed in sleep-disrupted mice suggest the dysbiosis induced by repeated inadequate sleep may have a detrimental physiological impact. Differential abundance analysis of individual taxa revealed multiple bacterial taxa that were increased or decreased in the sleep-disrupted group compared to controls, including an increase in the *Firmicutes:Bacteroidetes* (F:B) ratio and a decrease in *Lactobacillus*, *Actinobacteria*, and *Bifidobacterium*, all of which have established physiological impacts. An increase in the F:B ratio is a blunt measure of community shift and has been seen in many pathological states including obesity (54,55), chronic stress (56), as well as other models of sleep disruption (14,15). The phylum *Actinobacteria*, genus *Bifidobacterium*, and genus *Lactobacillus* were all low in sleep-disrupted mice. Previous studies in rodents and humans support a positive role for these taxa in stress resilience (13,65) and anxiety-like behavior (66-68). Therefore, an increased F:B ratio, along with reduced *Actinobacteria*, *Bifidobacterium*, and *Lactobacillus* indicates a state whereby ability to cope with a proinflammatory or anxiety-inducing stimulus may be reduced. Indeed, sleep deprivation results in altered responsiveness of the hypothalamic-pituitary-adrenal (HPA) axis (69-71), increased inflammation (72,73), and potentiated effects of a chemical stressor in a model of colitis in mice (74), indicating that sleep deprivation may be a factor promoting stress vulnerability. This is supported by a human study that observed that preexisting complaints of poor sleep increased the risk of posttraumatic stress disorder (PTSD) and other stress-related psychiatric disorders following trauma exposure (75).

As one of the principle mechanisms by which a change in the fecal microbiota can impact host physiology is via change in the molecules they produce, we also examined the fecal metabolome in this study. An untargeted metabolomic approach allows for wide surveillance

of the molecular environment as well as discovery of new molecular classes of interest (76). Although untargeted mass spectrometry cannot confirm exact structures of metabolites of interest without secondary targeted standard assays, GNPS (27) allows us to infer the general class of many interesting features based on spectral matches and molecular networking. Using this approach, we identified multiple classes of molecules significantly impacted by sleep disruption, including bile acids, which are microbially modified and physiologically relevant.

Multiple results from this study suggest microbiota-influenced bile acid metabolism was impacted by sleep disruption. Primary bile acids are cholesterol derivatives that are synthesized in the mammalian liver and excreted into the intestinal lumen to aid in lipid emulsification and absorption. In the intestine, primary bile acids are dehydroxylated and deconjugated by the gut microbiota, creating secondary bile acids and greatly enhancing the diversity of the bile acid pool (77). Mounting evidence in the past decade has described bile acids as versatile signaling molecules, with receptors throughout the mammalian organism (25,61,77). Some bile acids can act as anti-inflammatory and immunoregulatory agents in the intestinal tract and the central nervous system by activating the bile acid receptors FXR (farnesoid X receptor) and TGR5 (Takeda G protein-coupled receptor 5) (24,60). Furthermore, bile acid receptors play a role in glucose, lipid, and cholesterol metabolism (78,79). Two molecules with spectral matches to cholic acid and two unannotated molecules within molecular networks that contained multiple primary and secondary bile acids were decreased in the sleep-disrupted group at SD+2.

Importantly, analysis of the inferred gene content also revealed a reduction in the abundance of the microbial *bile salt hydrolase* (BSH) gene in the microbiomes of sleep-disrupted mice at SD+2. BSH catalyzes the critical first step in microbial bile acid metabolism,

and multiple lines of evidence suggest these enzymes may be the "gatekeepers" of host-microbiome crosstalk (59). In a previous experiment, feeding *Escherichia coli* engineered to overexpress *Lactobacillus* BSH to mice protected them from weight gain, and curbed lipid and cholesterol metabolism (80). A reduction in the fecal bile acid pool due to a reduction in microbial BSH, therefore, could result in a proinflammatory, metabolically dysregulated state in the host, both of which are reported in models of sleep disruption (1-3).

We also noticed a general increase in the abundance of metabolites with spectral matches to tripeptides and dipeptides. This could indicate an increase in host mucosal proteolysis or in microbial proteolysis (81). Host and microbial proteolytic enzymes play a role in gastrointestinal physiology, including activating signaling pathways (e.g. protease-activated receptors or PAR's) controlling inflammation and gut barrier function (81) as well as modulation of dorsal root ganglion neuron excitability (82). Furthermore, commensal bacteria have been shown to create molecules with potent protease inhibitory activity (83), so a shift in microbial community structure could have a direct impact on host proteolysis and physiology.

A third class of molecules that was impacted by sleep disruption was pentacyclic triterpenoids and close derivatives. The molecular family of dihydroxylated pentacyclic triterpenoids, judged by spectral matches to sumaresinolic acid and corosolic acid, were decreased and increased, respectively, at SD+2, while the dihydroxylated and monohydroxylated spectral features, with spectral matches to hederagenin and wilforlide A, were significantly decreased at SD+4. Also, four unannotated spectra within the network were significantly increased, and one was decreased, at SD+2. Triterpenoids are a class of diverse, plant-derived molecules that have been traditionally studied for their anti-tumor or anti-

inflammatory properties (62,84,85). Shifts in the balance of this molecular network could therefore have impacts on host physiology. While it has been shown that administration of triterpenoid molecules can modulate the microbiota (86), and that certain pentacyclic triterpenoids are metabolized by the microbiota (87), it is unclear whether the changes seen in this family of molecules due to sleep disruption were due to changes in the microbiome. In order to evaluate the biological impacts and therapeutic potential of the molecules discovered in this study, follow-up studies will need to be done to verify the structures of the features discovered here as well as to quantify their concentrations in the gut.

Our results are consistent with a study by Poroyko *et al*. investigating the gut microbiota in a mouse model of obstructive sleep apnea (15). In that study, four weeks of chronic sleep fragmentation caused global shifts in the fecal microbiome, as well as an increase in the F:B ratio, similar to this study. Furthermore, a study of acute sleep loss (four-hour sleep opportunity) in humans also observed an increase in the F:B ratio, but not a global shift in beta diversity (14). Recently, a study in rats found marked shifts in the fecal microbiome and urinary metabolites after a seven day severe stress/REM deprivation protocol (88). Taken with the present results, a link between inadequate sleep and the fecal microbiome appears to be present across species and sleep disruption protocols. Importantly, the present study expands on these findings to include the fecal metabolome, which has important implications as an effector system of microbe-host interactions (20). We found that, similar to the microbiome, the fecal metabolome (1124 measured metabolites) displayed a global shift when comparing sleep-disrupted mice to control mice at SD+2, but not SD+4.

It is of note that a study published by Zhang et al. (16) used a similar sleep restriction protocol (20h/day sleep disruption using a rotating bar for 7 days) but found no changes in the fecal microbiome. There are a few potential reasons for this discrepancy. First, the study by Zhang and colleagues used rats while ours used mice. Second, the rats in both the sleep restriction and control groups were manipulated every day to collect body weight measurements and fecal pellets. We chose to leave the animals relatively undisturbed throughout the sleep disruption protocol. Our automated protocol allowed mice to remain in the same home cage to have undisturbed sleep opportunity and to have limited contact with human researchers, which can affect the microbiome (89,90). However, this approach did introduce some limitations to the experiment because it did not allow for constant monitoring of food intake or fecal microbiome/metabolome during the sleep disruption period.

Overall, this study characterizes the impact of inadequate sleep on fecal microbiome as well as the fecal metabolome, a potential effector system in microbe-host interactions. The changes to microbiome and metabolome were present on the second day of recovery sleep, and some changes persisted until at least the fourth day of recovery sleep, despite the recovery of most sleep within two days of the cessation of sleep disruption. This is particularly interesting in view of the observation that some of the neurobehavioral impairments observed after a week of short sleep do not recover after a ‘weekend’ of recovery sleep despite reduction of subjective sleepiness (91,92). Our findings also suggest that changes seen in particular bacteria and bacterially-influenced signaling molecules such as bile acids suggest a proinflammatory, metabolically dysregulated state in the days following a five-day sleep disruption protocol. Interventions designed to maintain the fecal microbiome and metabolome, or to proactively

offset the negative impacts of dysbiosis, should be investigated to promote resilience to repeated sleep disruption, a problem that is ubiquitous in modern society.

F. Acknowledgements

This research was funded by the Office of Naval Research Grant # N00014-15-1-2809, with support from National Institutes of Health Training Grant T32HL007909. The authors would like to acknowledge Dr. Gail Ackermann for organizing and coordinating data processing and analysis, and Chris Olker and Eun Joo Song for assistance scoring sleep.

Chapter 6, in full, has been submitted for publication of the material as it may appear in Sleep, 2019, Bowers, S. J; Vargas, F; Gonzalez, A; He, S; Jiang, P; Dorrestein, P. C; Knight, R; Wright, K. P; Lowry, C. A; Fleshner, M; Vitaerna, M. H; Turek, F. W.; Oxford University Press, 2019. The dissertation author was a primary investigator and author of this paper.

G. Figures and Tables

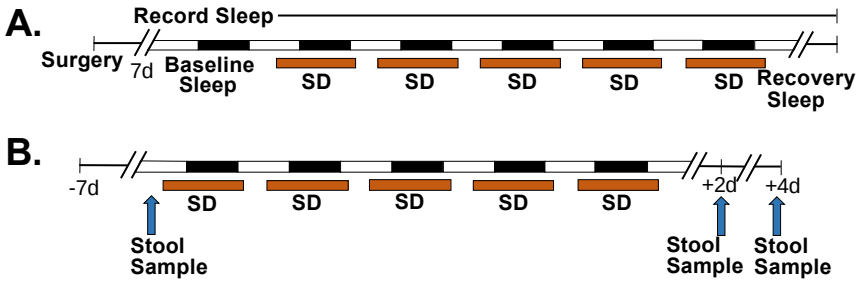


Figure 6.1. Experimental timelines. **a**) Experiment 1. Adult male C57BL/6N mice (*n* = 3, Control; *n* = 4, Sleep Disruption) received surgical implants of sleep recording devices. After recovery from surgery, mice were subjected to five days of repeated sleep disruption and two days of *ad libitum* recovery sleep. Sleep was disrupted for 20 h/day, with an *ad libitum* sleep window between ZT2-ZT6. Sleep was recorded throughout the experiment. **b**) Experiment 2. Non-instrumented adult male C57BL/6N mice (*n* = 10/group) were subjected to the same sleep disruption protocol, but with four days of recovery sleep. Stool samples were collected before sleep disruption, on day 2 post-sleep disruption and on day 4 post-sleep disruption (arrows). Abbreviations: SD, Sleep Disruption.

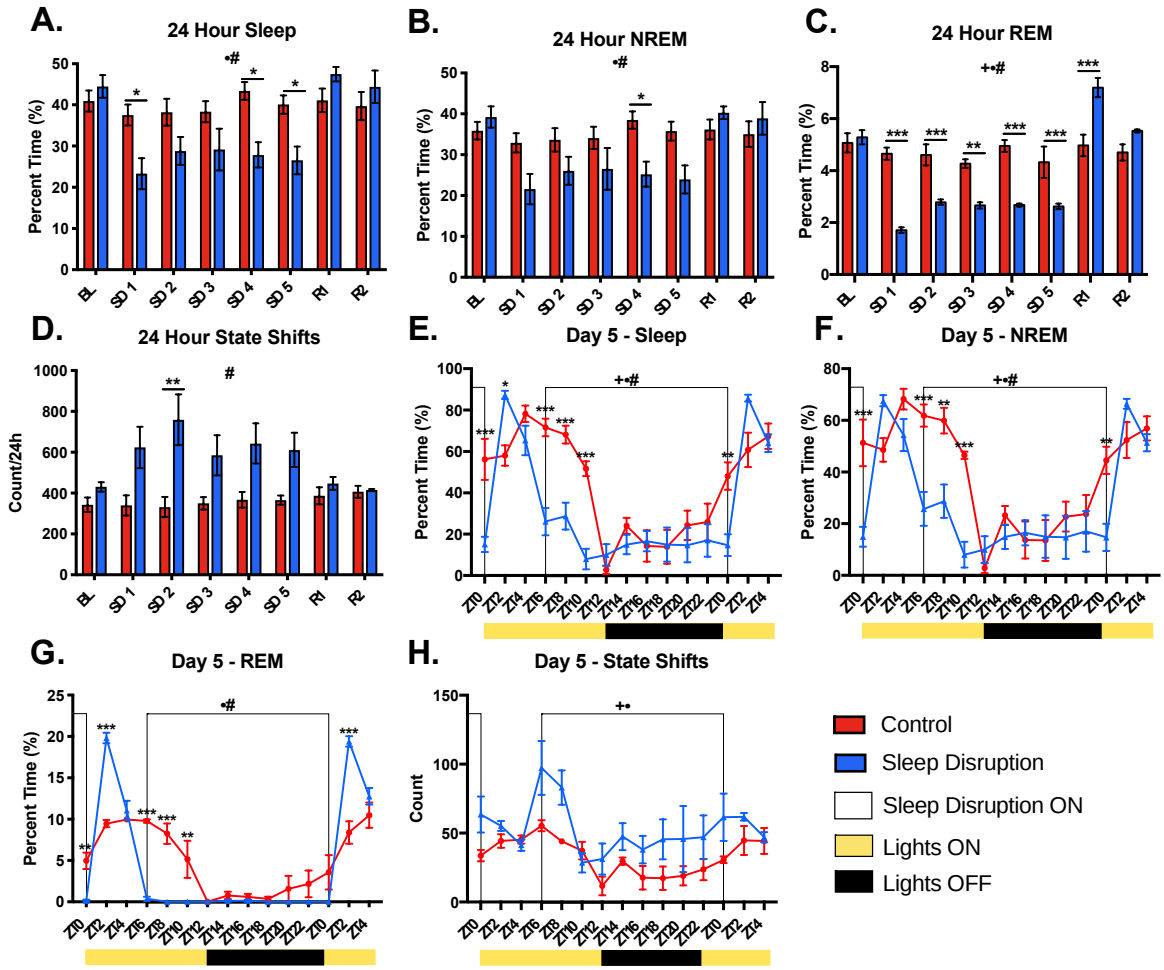


Figure 6.2. Effect of sleep disruption protocol on sleep measures. **a-d**) 24-hour totals of total sleep, non-rapid eye movement sleep (NREM), rapid eye movement sleep (REM), and state shifts. There was a significant decrease in sleep, NREM, and REM during the sleep disruption protocol, and an increase in state shifts. **e-h**) Two-hour bins of total sleep, NREM, REM, and state shifts from the fifth day of the sleep disruption protocol through ZT4 of the first day of recovery sleep. Yellow bars under the x axis indicate the lights being on, while black bars indicate the lights being off. Abbreviations: BL, baseline; SD, sleep disruption; R, recovery; ZT, zeitgeber time. n = 3-4/group. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (Bonferroni post hoc test); +$p < 0.05$ (overall effect of SD over entire time interval, Mixed-effects model); •$p<0.05$ (overall effect of Time over entire time interval, Mixed-effects model); #$p < 0.05$ (SDxTime interaction over entire time interval, Mixed-effects model).

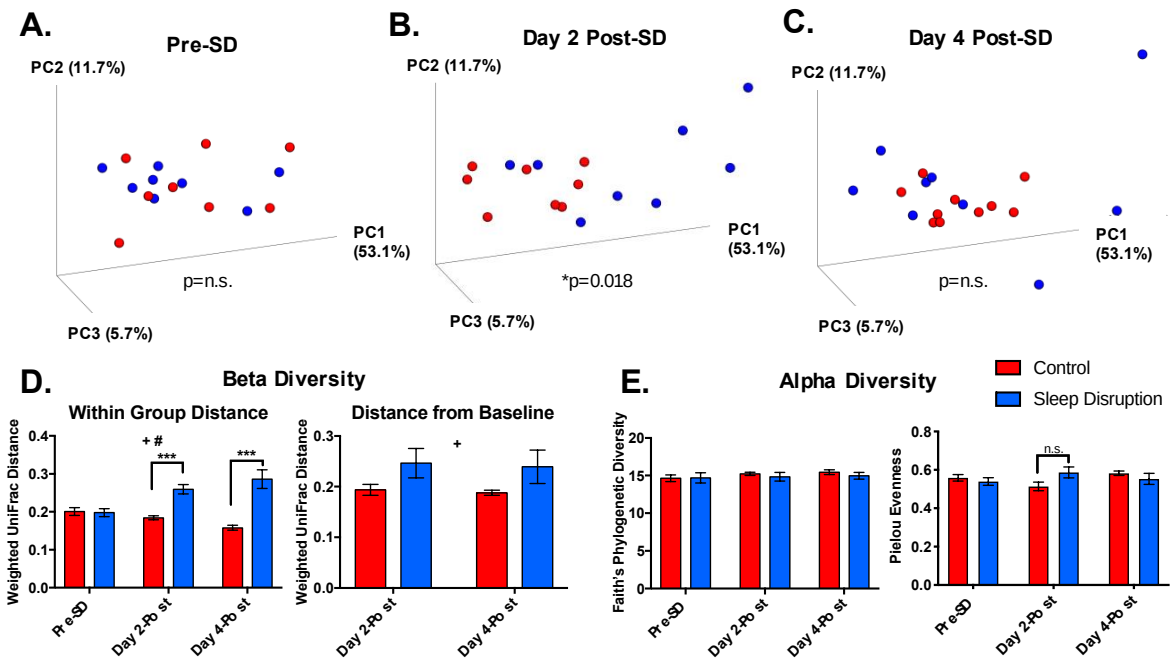


Figure 6.3. Effect of sleep disruption on microbiome beta and alpha diversity. **a-c**) Principal coordinates analysis (PCoA) plots using weighted UniFrac distance. A significant difference between sleep disruption and control groups at day 2 post-sleep disruption was detected using PERMANOVA. **d**) Average weighted UniFrac distance from an individual to all individuals within the same group (left) and from an individual post-sleep disruption to each individual pre-sleep disruption (right) is increased at both day 2 and day 4 post-sleep disruption. **e**) Faith's Phylogenetic Diversity (left) and Pielou Evenness (right) were unchanged throughout the experiment. SD, Sleep Disruption. n = 8-10/group. *$p < 0.05$ (PERMANOVA); **$p < 0.01$, ***$p < 0.001$ (Bonferroni post hoc test); +$p < 0.05$ (Overall effect of SD, Mixed-effects model); #$p < 0.05$ (SD x Time interaction, Mixed-effects model).

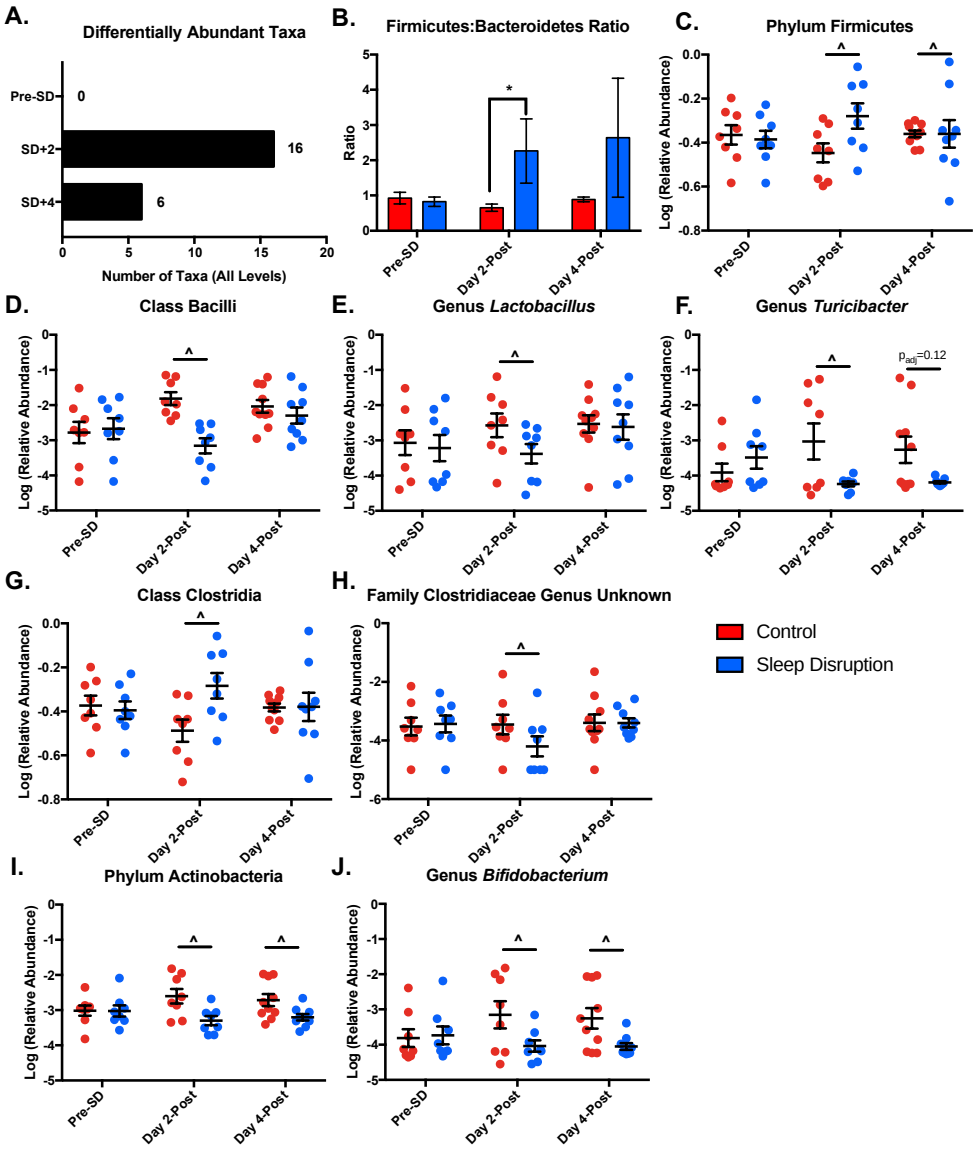


Figure 6.4. Effect of sleep disruption on individual microbial taxa. At each timepoint, DESeq2 was performed at each taxonomic level to identify taxa differentially abundant between sleep disrupted and control groups. **a**) Summary of significantly different taxa by timepoint. **b**) The ratio of relative abundances of the phyla *Firmicutes* to *Bacteroidetes* was significantly increased at day 2 post-sleep disruption in sleep-disrupted animals. This increase was mostly driven by a significant increase in *Firmicutes* (**c**). Within the *Firmicutes* phylum, the class *Bacilli* (**d**), genus *Lactobacillus* (**e**), and genus *Turicibacter* (**f**) were reduced at day 2 post-sleep disruption. The class *Clostridia* was increased (**g**) and an unknown genus within the *Clostridiaceae* family (**h**) was decreased in sleep-disrupted animals at day 2 post-sleep disruption. Both the phylum *Actinobacteria* (**i**), and the genus *Bifidobacterium* (**j**) were decreased at day 2 and day 4 post-sleep disruption in the sleep disrupted group. n = 8-10/group. Data represent means ± SEM. *p < 0.05 (Wilcoxon Rank-Sum Test); ^FDR < 0.1 (DESeq2).

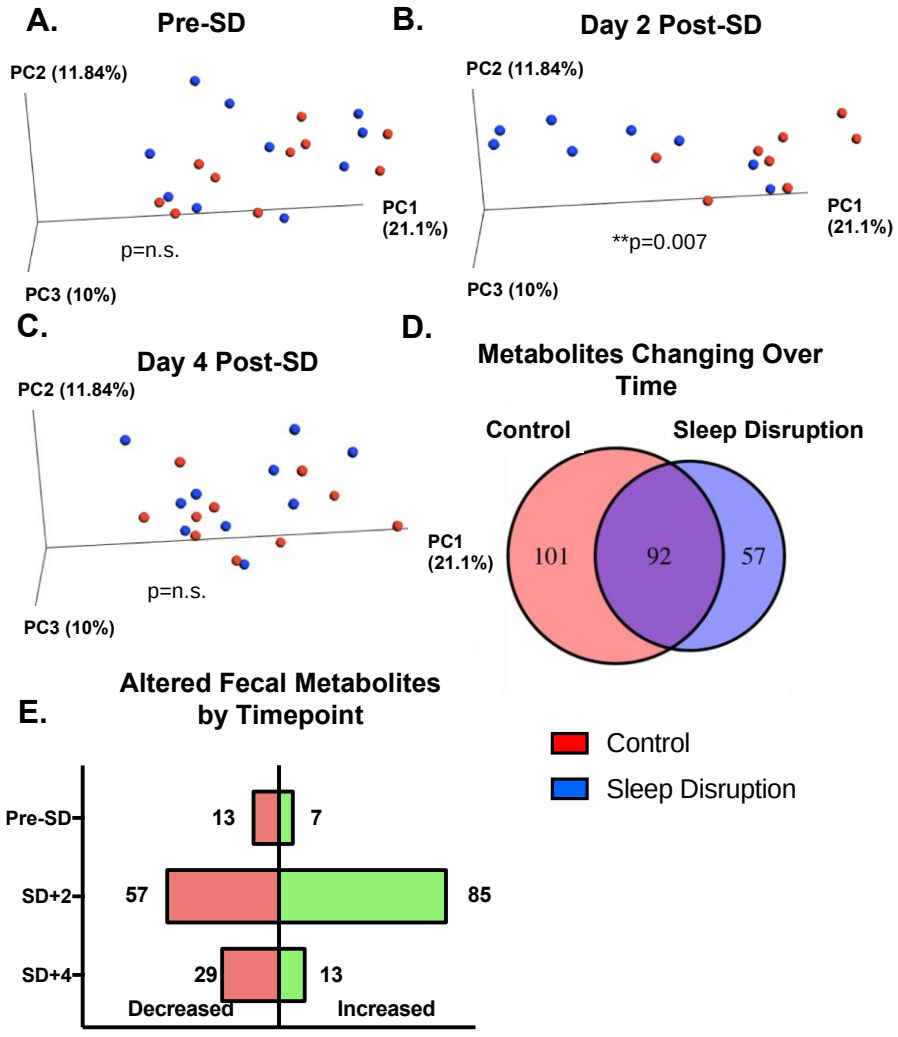
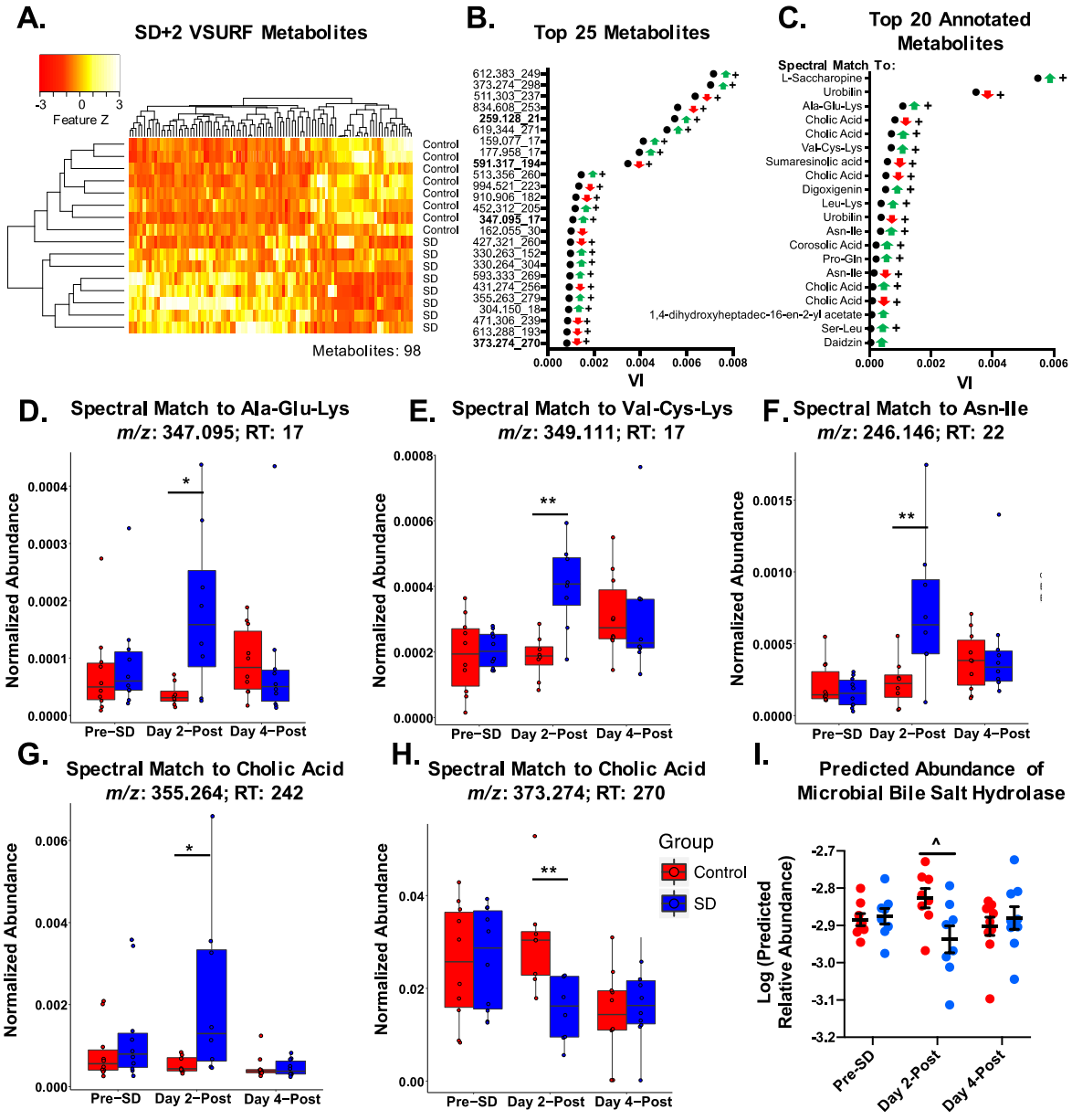


Figure 6.5. Effect of sleep disruption on the fecal metabolome. **a**,**b**,**c**) Principal coordinates analysis (PCoA) plots using Bray Curtis distance. PERMANOVA detected a significant difference between sleep disruption and control groups day 2 post-sleep disruption (SD), but not pre-SD or at day 4 post-SD. **d**) Kruskal-Wallace tests were run within the control group and within the SD group to determine metabolites significantly changing over the course of the experiment (FDR < 0.1). The number of metabolites found to have an effect of time only in the control group (left number), an effect of time only in the sleep-disrupted group (right number), or in both groups (middle number) is depicted in the Venn diagram. **e**) Wilcoxon Rank-Sum tests were performed at each timepoint to quantify the number of metabolites increased (right, green bars) or decreased (left, pink bars) in the sleep disruption group at each timepoint (uncorrected $p < 0.05$). SD, sleep disruption. n = 8-10/group. **$p < 0.01$ (PERMANOVA.

Figure 6.6. Metabolites that drive changes seen at day 2 post-sleep disruption. Variable Selection Using Random Forests (VSURF) was performed at the second day post-sleep disruption (SD+2) to identify metabolites that are the most important drivers (above a threshold variable importance) of separation between sleep-disrupted (SD) and control groups. **a**) Heatmap of the 98 suprathreshold metabolites. **b**) Variable importance scores of the top 25 suprathreshold metabolites (*m/z*_RT), along with direction of change (arrows, green/up = increased in SD group, red/down = decreased in SD group). Bold indicates metabolites that were annotated using Global Natural Products Social Molecular Networking (GNPS). **c**) Top 20 annotated metabolites. Normalized abundance (peak intensity normalized to total ion count) of metabolites with spectral matches to **d**) Ala-Glu-Lys, **e**) Val-Cys-Lys, and **f**) Asn-Ile were increased at SD+2 but not at day SD+4. **g**,**h**) One metabolite with a spectral match to cholic acid was increased (**g**) and another was decreased (**h**) at SD+2. **i**) Inferred abundance of the microbial bile salt hydrolase gene (EC:3.5.1.24) was also reduced at SD+2. For (**d**-**h**), boxes indicate median, 25^{th} and 75^{th} quantiles; whiskers indicate 2*IQR from edges of box. For (**i**), data represent mean ± SEM. Abbreviations: VI, variable importance; SD, sleep disruption; *m/z*, mass to charge ratio; RT, retention time (seconds). n = 8-10/group. +FDR < 0.1, *p < 0.05, **p < 0.01 (Wilcoxon-Rank Sum test); ^FDR < 0.1 (DESeq2).


A.
SD+2 VSURF Metabolites
Feature Z
-3
0
3
Control
Control
Control
Control
Control
Control
Control
SD
SD
SD
SD
SD
SD
SD
Metabolites: 98
B.
Top 25 Metabolites
612.383_249
373.274_298
511.303_237
834.608_253
259.125_21
619.344_271
159.077_17
177.958_17
591.317_194
513.356_260
994.521_223
910.906_182
452.312_205
347.095_17
162.055_30
427.321_260
330.263_152
330.264_304
593.333_269
431.274_256
355.263_279
304.150_18
471.306_239
613.288_193
373.274_270
0.000
0.002
0.004
0.006
0.008
VI
C.
Top 20 Annotated Metabolites
Spectral Match To:
L-Saccharopine
Urobilin
Ala-Glu-Lys
Cholic Acid
Cholic Acid
Val-Cys-Lys
Sumaresinolic acid
Cholic Acid
Digoxigenin
Leu-Lys
Urobilin
Asn-Ile
Corosolic Acid
Pro-Gln
Asn-Ile
Cholic Acid
Cholic Acid
1,4-dihydroxyheptadec-16-en-2-yl acetate
Ser-Leu
Daidzin
0.000
0.002
0.004
0.006
VI
D.
Spectral Match to Ala-Glu-Lys
m/z: 347.095; RT: 17
Normalized Abundance
0.0000
0.0001
0.0002
0.0003
0.0004
*
Pre-SD
Day 2-Post
Day 4-Post
E.
Spectral Match to Val-Cys-Lys
m/z: 349.111; RT: 17
Normalized Abundance
0.0000
0.0002
0.0004
0.0006
0.0008
**
Pre-SD
Day 2-Post
Day 4-Post
F.
Spectral Match to Asn-Ile
m/z: 246.146; RT: 22
Normalized Abundance
0.0000
0.0005
0.0010
0.0015
**
Pre-SD
Day 2-Post
Day 4-Post
G.
Spectral Match to Cholic Acid
m/z: 355.264; RT: 242
Normalized Abundance
0.000
0.002
0.004
0.006
*
Pre-SD
Day 2-Post
Day 4-Post
H.
Spectral Match to Cholic Acid
m/z: 373.274; RT: 270
Normalized Abundance
0.00
0.02
0.04
**
Group
Control
SD
Pre-SD
Day 2-Post
Day 4-Post
I.
Predicted Abundance of Microbial Bile Salt Hydrolase
Log (Predicted Relative Abundance)
-2.7
-2.8
-2.9
-3.0
-3.1
-3.2
^
Pre-SD
Day 2-Post
Day 4-Post

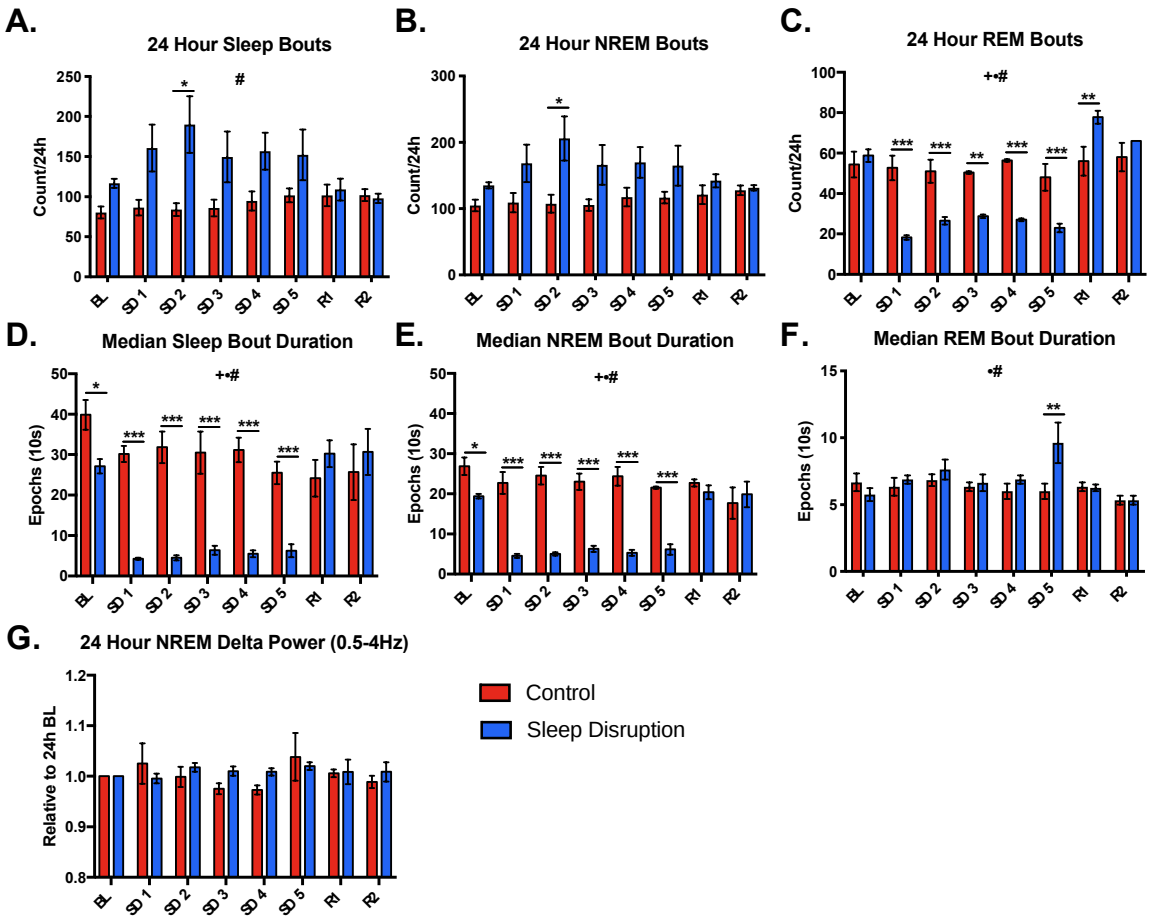


Figure 6.S1. Effect of sleep disruption protocol on sleep fragmentation measures and delta power. **a**,**b**) There was a significant increase in the 24-hour totals of total sleep bouts (**a**) and non-rapid eye movement sleep (NREM) bouts (**b**) during the sleep disruption protocol in the sleep disruption (SD) group. **c**) Rapid eye movement (REM) bouts were decreased during the SD protocol, and increased on the first day of recovery in the SD group. **d**,**e**,**f**) Median sleep bout duration (**d**) and NREM bout duration (**e**) were significantly decreased in the SD group during the SD protocol, while the median REM bout duration (**f**) was unaffected in all days except for SD5. **g**) There was no change in 24-hour NREM delta power due to SD. Abbreviations: BL, baseline; SD, sleep disruption; R, recovery; ZT, zeitgeber time. n = 3-4/group. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (Bonferroni post hoc test); +$p < 0.05$ (overall effect of SD, Mixed-effects model); •$p < 0.05$ (overall effect of Time, Mixed-effects model); #$p < 0.05$ (SDxTime interaction, Mixed-effects model).

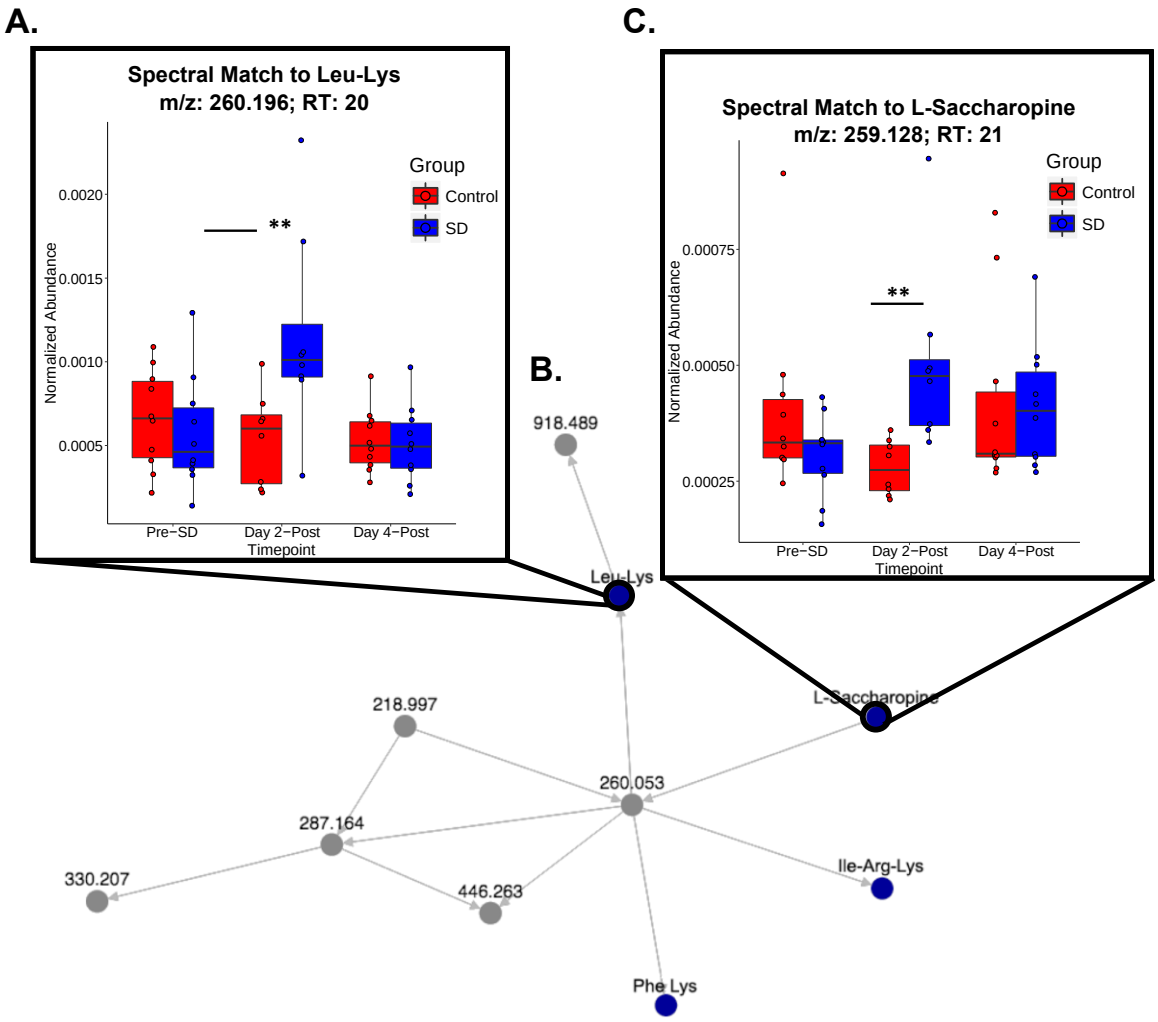


Figure 6.S2. Sleep disruption changes fecal levels of molecules related to protein metabolism. **a**) Normalized abundance (peak intensity normalized to total ion count) of a metabolite with a spectral match to Leu-Lys was increased at day 2 post-sleep disruption (SD). **b**) Global Natural Products Social Molecular Networking (GNPS)-generated molecular network containing multiple annotated (blue) and unannotated (grey) clusters. Numbers next to grey clusters indicate average parent mass of the spectra in the cluster. Metabolites that were above the threshold variable importance in Variable Selection Using Random Forests (VSURF) analysis are outlined with a black circle. Length of grey lines connecting clusters indicates relative similarity of the MS*2* spectra. Arrowheads point towards clusters with a larger *m/*z. **c**) Normalized abundance of a metabolite with a spectral match to L-saccharopine was increased at day 2 post-SD. Boxes indicate median, 25th and 75th quantiles; whiskers indicate 2*IQR from edges of box. Abbreviations: VI, variable importance; SD, sleep disruption; *m/z*, mass to charge ratio; RT, retention time (seconds). n = 8-10/group. *$p < 0.05$, **$p < 0.01$ (Wilcoxon-Rank Sum test).

Figure 6.S3. Sleep disruption decreases fecal levels of unknown molecules in networks with bile acids. **a**,**d**) Global Natural Products Social Molecular Networking (GNPS)-generated molecular networks containing multiple annotated (blue) and unannotated (grey) clusters. Numbers next to grey clusters indicate average parent mass of the spectra in the cluster. Metabolites that were above the threshold variable importance in Variable Selection Using Random Forests (VSURF) analysis are outlined with a black circle. Length of grey lines connecting clusters indicates relative similarity of the MS^2 spectra. Arrowheads point towards clusters with a larger *m/z*. **b**) Normalized abundance (peak intensity normalized to total ion count) of an unannotated metabolite with ID 516 was decreased at day 2 post-sleep disruption (SD). **c**) Normalized abundance (peak intensity normalized to total ion count) of an unannotated metabolite with ID 512 was decreased at day 2 post-SD. Boxes indicate median, 25^{th} and 75^{th} quantiles; whiskers indicate 2*IQR from edges of box. Abbreviations: VI, variable importance; SD, sleep disruption; *m/z*, mass to charge ratio; RT, retention time (seconds). n = 8-10/group. *$p < 0.05$, **$p < 0.01$ (Wilcoxon-Rank Sum test).

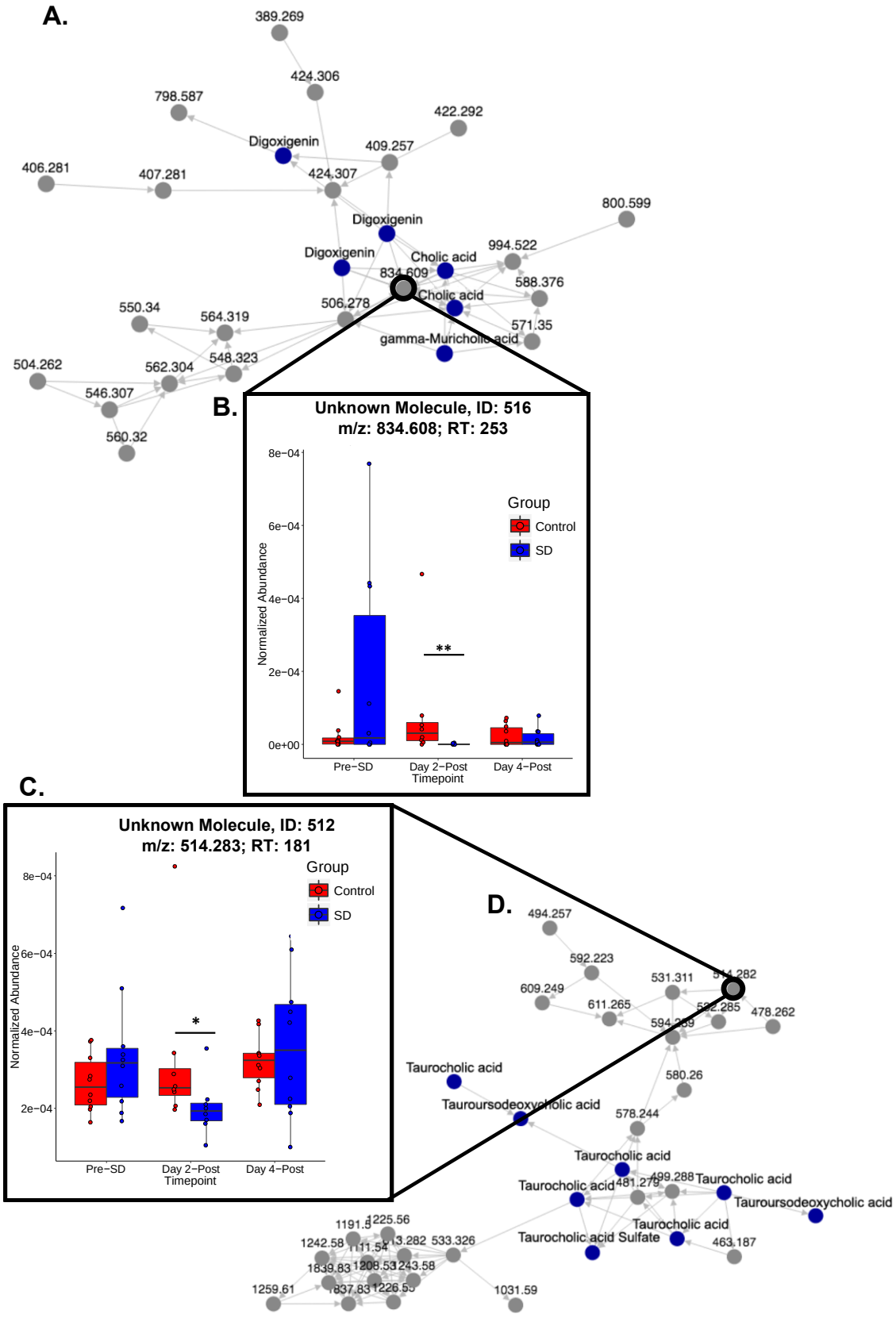
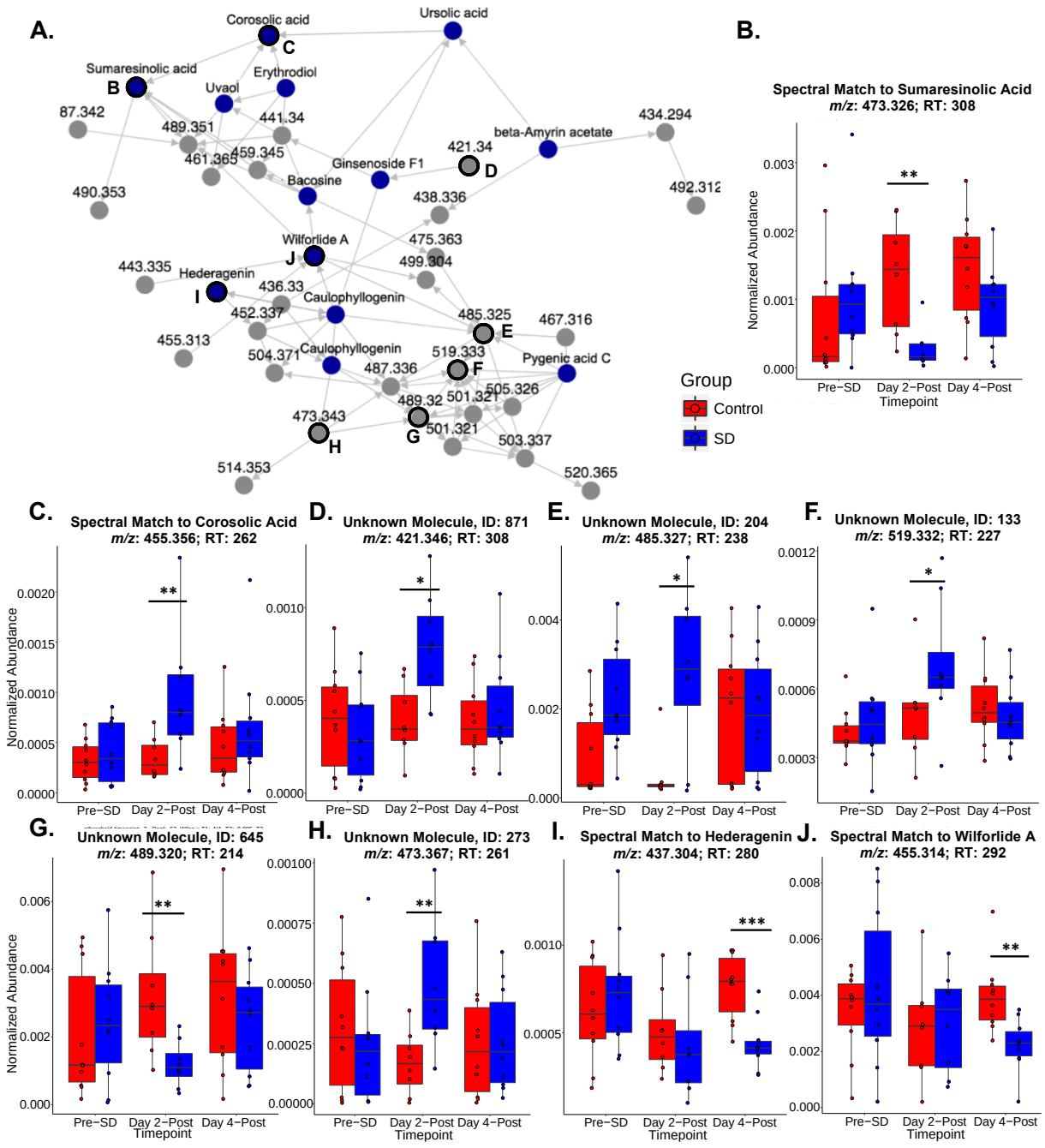

A.
389.269
424.306
798.587
422.292
Digoxigenin
409.257
406.281
407.281
424.307
800.599
Digoxigenin
994.522
Digoxigenin
Cholic acid
834.609
588.376
Cholic acid
550.34
506.278
564.319
571.35
gamma-Muricholic acid
504.262
562.304
548.323
546.307
560.32
B.
Unknown Molecule, ID: 516
m/z: 834.608; RT: 253
Normalized Abundance
8e−04
6e−04
4e−04
2e−04
0e+00
Group
Control
SD
**
Pre−SD
Day 2−Post
Day 4−Post
Timepoint
C.
Unknown Molecule, ID: 512
m/z: 514.283; RT: 181
Normalized Abundance
8e−04
6e−04
4e−04
2e−04
Group
Control
SD
*
Pre−SD
Day 2−Post
Day 4−Post
Timepoint
D.
494.257
592.223
609.249
531.311
611.265
478.262
Taurocholic acid
580.26
Tauroursodeoxycholic acid
578.244
Taurocholic acid
Taurocholic acid
Taurocholic acid
Tauroursodeoxycholic acid
Taurocholic acid
Taurocholic acid Sulfate
463.187
1225.56
1191.5
1242.58
533.326
1839.83
1259.61
1031.59

Figure 6.S4. Sleep disruption changes fecal levels of molecules related to pentacyclic triterpenoids. **a**) Global Natural Products Social Molecular Networking (GNPS)-generated molecular network containing multiple annotated (blue) and unannotated (grey) clusters. Numbers next to grey clusters indicate average parent mass of the spectra in the cluster. Metabolites that were above the threshold variable importance in Variable Selection Using Random Forests (VSURF) analysis are outlined with a black circle. Letters next to the black circles indicate the panel of the figure corresponding to the cluster. Length of grey lines connecting clusters indicates relative similarity of the MS^2 spectra. Arrowheads point towards clusters with a larger *m/z*. **b**,**c**) Normalized abundance (peak intensity normalized to total ion count) of a metabolite with a spectral match to sumaresinolic acid (**b**) was decreased at day 2 post-sleep disruption, while a spectral match to corosolic acid (**c**) was increased at SD+2. **d**,**e**,**f**,**h**) Unannotated molecules with ID 871 (**d**), 204 (**e**), 133 (**f**), and 273 (**h**) were increased at SD+2. **g**,**i**,**j**) Unannotated molecule ID 645 (**g**) was decreased at SD+2. Metabolites matching hederagenin (**i**), and wilforlide A (**j**) were decreased at SD+4. Boxes indicate median, 25^{th} and 75^{th} quantiles; whiskers indicate points within 2*IQR from edges of box. Abbreviations: VI, variable importance; SD, sleep disruption; *m/z*, mass to charge ratio; RT, retention time (seconds). n = 8-10/group. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (Wilcoxon-Rank Sum test).


A.
Corosolic acid
C
Ursolic acid
Sumaresinolic acid
B
Erythrodiol
Uvaol
87.342
441.34
489.351
461.365
459.345
beta-Amyrin acetate
434.294
421.34
D
Ginsenoside F1
Bacosine
490.353
492.312
438.336
Wilforlide A
J
475.363
499.304
443.335
Hederagenin
I
436.33
Caulophyllogenin
452.337
485.325
E
467.316
455.313
504.371
Caulophyllogenin
519.333
Pygenic acid C
487.336
F
505.326
489.32
501.321
473.343
H
G
501.321
503.337
514.353
520.365
Group
Control
SD
B.
Spectral Match to Sumaresinolic Acid
m/z: 473.326; RT: 308
Normalized Abundance
Pre−SD
Day 2−Post
Day 4−Post
Timepoint
C.
Spectral Match to Corosolic Acid
m/z: 455.356; RT: 262
D.
Unknown Molecule, ID: 871
m/z: 421.346; RT: 308
E.
Unknown Molecule, ID: 204
m/z: 485.327; RT: 238
F.
Unknown Molecule, ID: 133
m/z: 519.332; RT: 227
G.
Unknown Molecule, ID: 645
m/z: 489.320; RT: 214
H.
Unknown Molecule, ID: 273
m/z: 473.367; RT: 261
I.
Spectral Match to Hederagenin
m/z: 437.304; RT: 280
J.
Spectral Match to Wilforlide A
m/z: 455.314; RT: 292

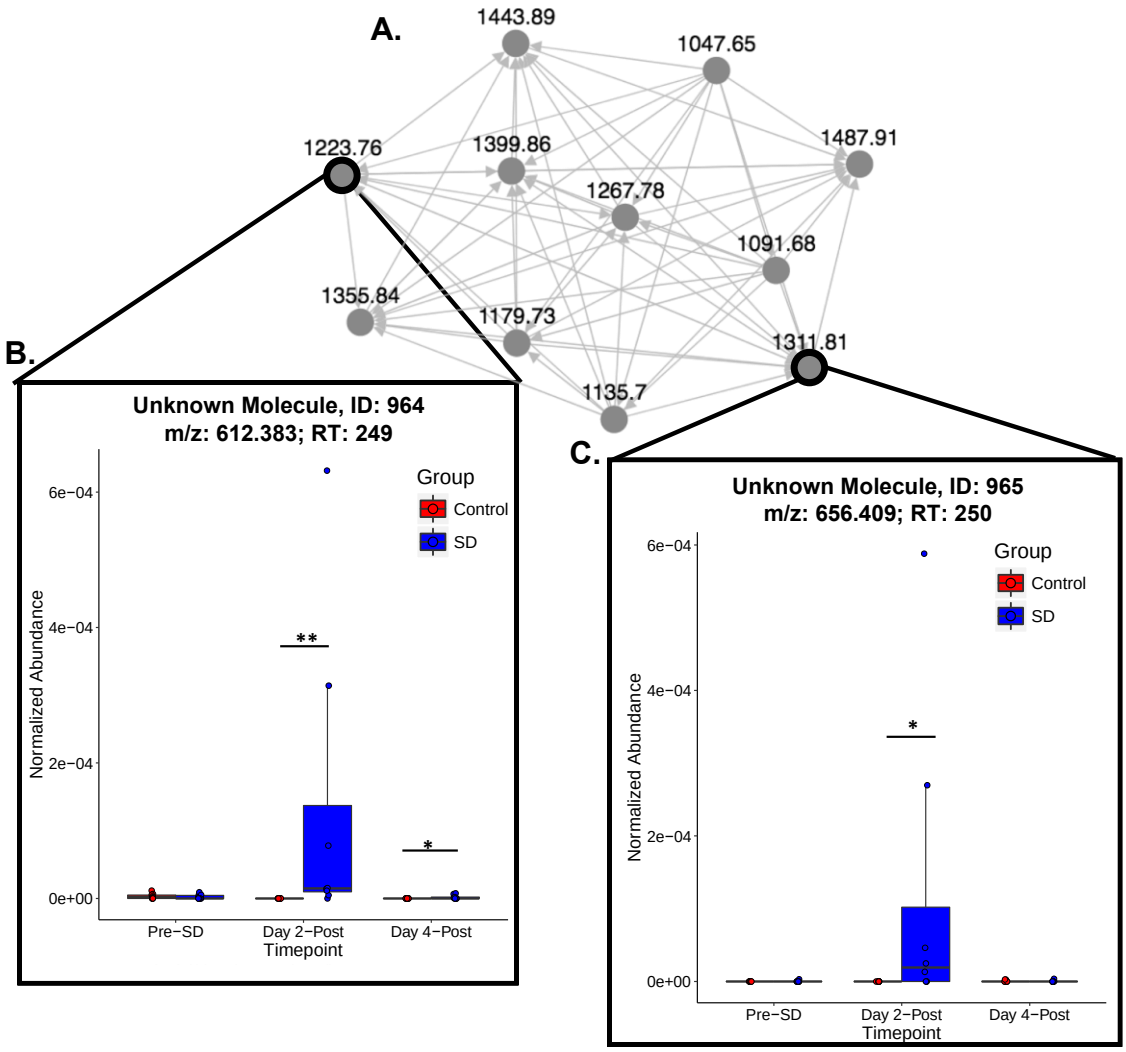


Figure 6.S5. Two unknown fecal metabolites are present only in sleep-disrupted mice. **a**) Global Natural Products Social Molecular Networking (GNPS)-generated molecular network containing multiple unannotated (grey) clusters. Numbers next to grey clusters indicate average parent mass of the spectra in the cluster. Metabolites that were above the threshold variable importance in Variable Selection Using Random Forests (VSURF) analysis are outlined with a black circle. Length of grey lines connecting clusters indicates relative similarity of the MS^2 spectra. Arrowheads point towards clusters with a larger *m/z*. **b**) Normalized abundance (peak intensity normalized to total ion count) of an unannotated metabolite with ID 964 was increased at day 2 post-sleep disruption and at day 4 post-sleep disruption. **c**) Normalized abundance (peak intensity normalized to total ion count) of an unannotated metabolite with ID 965 was increased at day 2 post-sleep disruption but not day 4 post-sleep disruption. Boxes indicate median, 25^{th} and 75^{th} quantiles; whiskers indicate points within 2*IQR from edges of box. Abbreviations: VI, variable importance; SD, sleep disruption; *m/z*, mass to charge ratio; RT, retention time (seconds). n = 8-10/group. *$p < 0.05$, **$p < 0.01$ (Wilcoxon-Rank Sum test).

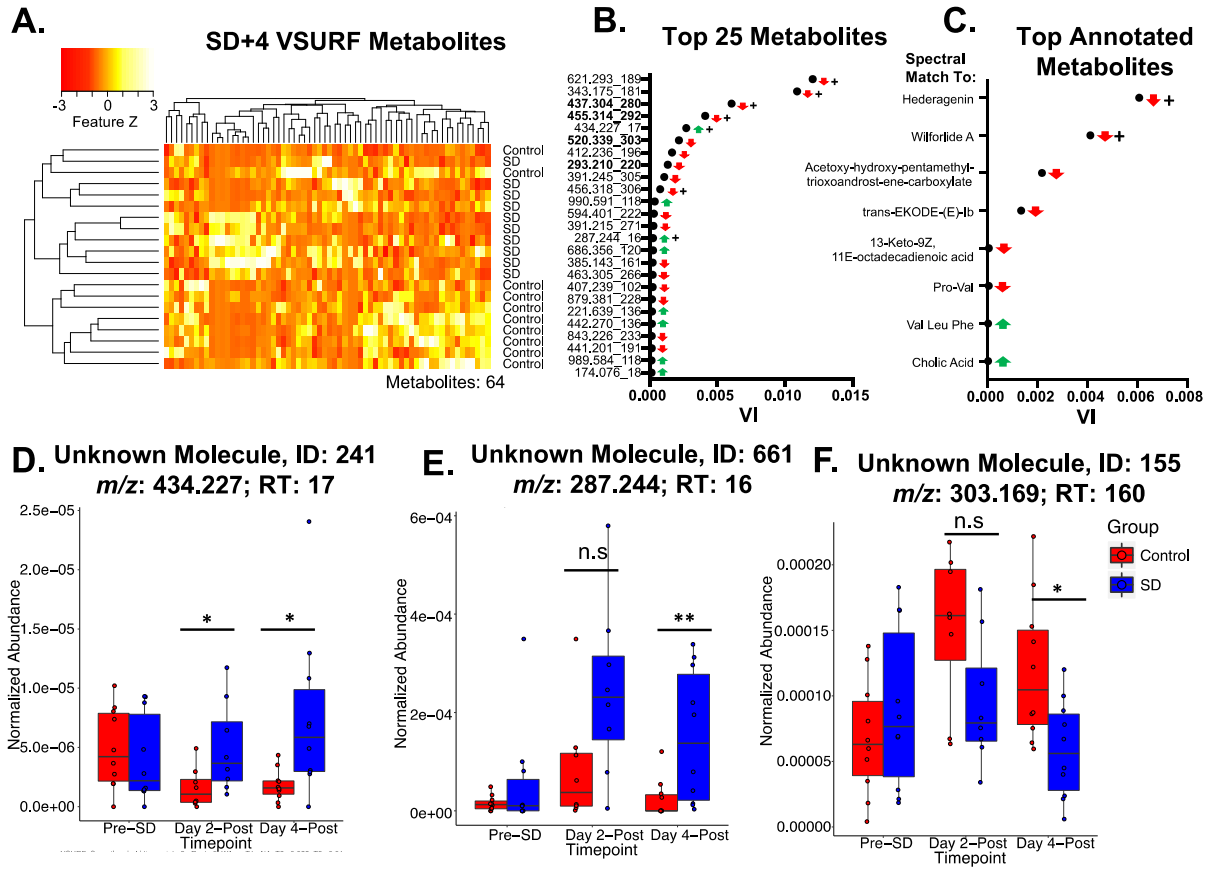


Figure 6.S6. Metabolites that are changed at day 4 post-sleep disruption. **a**) Heatmap of the 64 metabolites that were above threshold variable importance in Variable Selection Using Random Forests (VSURF) analysis. **b**) Variable importance (VI) scores of the top 25 suprathreshold metabolites (*m/z*_RT). Bold indicates metabolites that were annotated using Global Natural Products Social Molecular Networking (GNPS). **c**) VI scores of annotated metabolites. **d**) Normalized abundance (peak intensity normalized to total ion count) of an unannotated metabolite with ID 241 was increased in sleep-disrupted relative to control mice at both day 2 post-sleep disruption (SD) and day 4 post-SD. **e**) An unannotated metabolite with ID 661 was trending towards an increase in sleep-disrupted compared to control at day 2 post-sleep disruption and was increased at day 4 post-SD. **f**) An unannotated metabolite with ID 155 was trending towards a decrease in sleep-disrupted mice compared to control mice at day 2 post-SD and was decreased at day 4 post-SD. Boxes indicate median, 25th and 75th quantiles; whiskers indicate 2*IQR from edges of box. Abbreviations: VI, variable importance; SD, sleep disruption; *m/z*, mass to charge ratio; RT, retention time (seconds). n = 8-10/group. *$p < 0.05$, **$p < 0.01$ (Wilcoxon-Rank Sum test).

Table 6.S2. Differentially Abundant Bacterial Taxa Post-Sleep Disruption

Day 2 Post-SD1

Taxon	Control Relative Abundance	Sleep Disrupted Relative Abundance	Fold Difference (SD vs Con)	Adjusted p
Phylum Actinobacteria	0.0050 ± 0.0020	0.0006 ± 0.0002	-0.8800	0.0004
Class Actinobacteria	0.0044 ± 0.0020	0.0001 ± 0.0001	-0.9773	0.012
Order Bifidobacteriales	0.0042 ± 0.0021	0.0001 ± 0.0001	-0.9762	0.028
Family Bifidobacteriaceae	0.0042 ± 0.0021	0.0001 ± 0.0001	-0.9762	0.032
Genus Bifidobacterium	0.0042 ± 0.0021	0.0001 ± 0.0001	-0.9762	0.046
Phylum Firmicutes	0.370 ± 0.036	0.558 ± 0.072	0.508	0.002
Class Bacilli	0.027 ± 0.010	0.001 ± 4.2e-4	-0.963	1.9e-6
Order Lactobacillales	0.013 ± 0.008	0.001 ± 4.2e-4	-0.923	0.003
Family Lactobacillaceae	0.013 ± 0.008	0.001 ± 3.7e-4	-0.923	0.023
Genus Lactobacillus	0.012 ± 0.008	0.001 ± 3.7e-4	-0.917	0.047
Order Turicibacterales	0.014 ± 0.008	0.000 ± 0.000	-1.000	0.015
Family Turicibacteraceae	0.014 ± 0.008	0.000 ± 0.000	-1.000	0.003
Genus Turicibacter	0.014 ± 0.008	0.000 ± 0.000	-1.000	0.047
Class Clostridia	0.340 ± 0.037	0.553 ± 0.072	0.627	0.051
Order Clostridiales	0.340 ± 0.037	0.553 ± 0.072	0.627	0.028
Family Clostridiaceae, Genus unknown	0.003 ± 0.002	0.001 ± 0.0005	-0.667	0.047

Day 4 Post-SD

Taxon	Control Relative Abundance	Sleep Disrupted Relative Abundance	Fold Difference (SD vs Con)	Adjusted p
Phylum Actinobacteria	0.0038 ± 0.0014	0.0006 ± 0.0002	-0.8421	0.004
Class Actinobacteria	0.0029 ± 0.0014	4.9e-5 ± 3.8e-5	-0.9831	0.005
Order Bifidobacteriales	0.0028 ± 0.0013	4.9e-5 ± 3.8e-5	-0.9825	0.002
Family Bifidobacteriaceae	0.0028 ± 0.0013	4.9e-5 ± 3.8e-5	-0.9825	0.015
Genus Bifidobacterium	0.0028 ± 0.0013	4.9e-5 ± 3.8e-5	-0.9825	0.005
Phylum Firmicutes	0.439 ± 0.015	0.475 ± 0.073	0.082	0.045
Order Turicibacterales	0.010 ± 0.007	0.000 ± 0.000	-1.00	0.087

[1]DESeq2 was performed at each taxonomic level to determine taxa differentially abundant between sleep-disrupted and control groups. Taxa significant at an FDR < 0.1 are listed above as the mean relative abundance + SEM, along with Benjamini Hochberg-adjusted p values. Fold difference: (SD-Con)/Con. Abbreviations: SD, Sleep Disruption. *n* = 8-10/group.

Table 6.S1. F Statistics for PERMANOVA, ANOVA, and Mixed-effects Models: See attached spreadsheet

Table 6.S3. VSURF Suprathreshold Metabolites[1]: See attached spreadsheet

Table 6.S4. PICRUSt2 Results: See attached spreadsheet.

H. References

1. Spiegel K, Leproult R, Van Cauter E. Impact of sleep debt on metabolic and endocrine function [Internet]. *The Lancet*. 1999.**354**:1435–1439. Available from: http://dx.doi.org/10.1016/s0140-6736(99)01376-8.

2. Irwin MR. Why Sleep Is Important for Health: A Psychoneuroimmunology Perspective [Internet]. *Annual Review of Psychology*. 2015.**66**:143–172. Available from: http://dx.doi.org/10.1146/annurev-psych-010213-115205.

3. Mullington JM, Haack M, Toth M, Serrador JM, Meier-Ewert HK. Cardiovascular, inflammatory, and metabolic consequences of sleep deprivation. *Prog Cardiovasc Dis*. 2009 Jan;**51**(4):294–302. PMCID: PMC3403737.

4. Krause AJ, Ben Simon E, Mander BA, Greer SM, Saletin JM, Goldstein-Piekarski AN, Walker MP. The sleep-deprived human brain [Internet]. *Nature Reviews Neuroscience*. 2017.**18**:404–418. Available from: http://dx.doi.org/10.1038/nrn.2017.55.

5. Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, Al-Soud WA, Sørensen SJ, Hansen LH, Jakobsen M. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One*. 2010 Feb 5;**5**(2):e9085. PMCID: PMC2816710.

6. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JI. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*. 2013 Sep 6;**341**(6150):1241214. PMCID: PMC3829625.

7. Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, Sitaraman SV, Knight R, Ley RE, Gewirtz AT. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science*. 2010 Apr 9;**328**(5975):228–231. PMCID: PMC4714868.

8. Cryan JF, Dinan TG. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat Rev Neurosci*. 2012 Oct;**13**(10):701–712. PMID: 22968153.

9. Bercik P, Denou E, Collins J, Jackson W, Lu J, Jury J, Deng Y, Blennerhassett P, Macri J, McCoy KD, Verdu EF, Collins SM. The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice. *Gastroenterology*. 2011 Aug;**141**(2):599–609, 609.e1–3. PMID: 21683077.

10. Colpitts SL, Kasper EJ, Keever A, Liljenberg C, Kirby T, Magori K, Kasper LH, Ochoa-Repáraz J. A bidirectional association between the gut microbiota and CNS disease in a biphasic murine model of multiple sclerosis [Internet]. *Gut Microbes*. 2017;8:561–573. Available from: http://dx.doi.org/10.1080/19490976.2017.1353843.

11. Koopman M, El Aidy S, MIDtrauma consortium. Depressed gut? The microbiota-diet-inflammation trialogue in depression. *Curr Opin Psychiatry*. 2017 Sep;**30**(5):369–377. PMID: 28654462.

12. Crumeyrolle-Arias M, Jaglin M, Bruneau A, Vancassel S, Cardona A, Daugé V, Naudon L, Rabot S. Absence of the gut microbiota enhances anxiety-like behavior and neuroendocrine response to acute stress in rats. *Psychoneuroendocrinology*. 2014 Apr;**42**:207–217. PMID: 24636517.

13. Hemmings SMJ, Malan-Müller S, van den Heuvel LL, Demmitt BA, Stanislawski MA, Smith DG, Bohr AD, Stamper CE, Hyde ER, Morton JT, Marotz CA, Siebler PH, Braspenning M, Van Criekinge W, Hoisington AJ, Brenner LA, Postolache TT, McQueen MB, Krauter KS, Knight R, Seedat S, Lowry CA. The Microbiome in Posttraumatic Stress Disorder and Trauma-Exposed Controls: An Exploratory Study. *Psychosom Med*. 2017 Oct;**79**(8):936–946. PMCID: PMC5763914.

14. Benedict C, Vogel H, Jonas W, Woting A, Blaut M, Schürmann A, Cedernaes J. Gut microbiota and glucometabolic alterations in response to recurrent partial sleep deprivation in normal-weight young individuals. *Mol Metab*. 2016 Dec;**5**(12):1175–1186. PMCID: PMC5123208.

15. Poroyko VA, Carreras A, Khalyfa A, Khalyfa AA, Leone V, Peris E, Almendros I, Gileles-Hillel A, Qiao Z, Hubert N, Farré R, Chang EB, Gozal D. Chronic Sleep Disruption Alters Gut Microbiota, Induces Systemic and Adipose Tissue Inflammation and Insulin Resistance in Mice. *Sci Rep*. 2016 Oct 14;**6**:35405. PMCID: PMC5064361.

16. Zhang SL, Bai L, Goel N, Bailey A, Jang CJ, Bushman FD, Meerlo P, Dinges DF, Sehgal A. Human and rat gut microbiome composition is maintained following sleep restriction. *Proc Natl Acad Sci U S A*. 2017 Feb 21;**114**(8):E1564–E1571. PMCID: PMC5338418.

17. Ray K. Gut microbiota. Host-microbe interactions and the enteric nervous system: a new connection? Nature reviews. *Gastroenterology & hepatology*. 2015;**12**:311. PMID: 25917439.

18. Mukherji A, Kobiita A, Ye T, Chambon P. Homeostasis in intestinal epithelium is orchestrated by the circadian clock and microbiota cues transduced by TLRs. *Cell*. 2013 May 9;**153**(4):812–827. PMID: 23663780.

19. Lowry CA, Smith DG, Siebler PH, Schmidt D, Stamper CE, Hassell JE Jr, Yamashita PS, Fox JH, Reber SO, Brenner LA, Hoisington AJ, Postolache TT, Kinney KA, Marciani D, Hernandez M, Hemmings SMJ, Malan-Muller S, Wright KP, Knight R, Raison CL, Rook GAW. The Microbiota, Immunoregulation, and Mental Health: Implications for Public Health. *Curr Environ Health Rep*. 2016 Sep;**3**(3):270–286. PMCID: PMC5763918.

20. De Vadder F, Kovatcheva-Datchary P, Goncalves D, Vinera J, Zitoun C, Duchampt A, Bäckhed F, Mithieux G. Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell*. 2014 Jan 16;**156**(1-2):84–96. PMID: 24412651.

21. Stilling RM, van de Wouw M, Clarke G, Stanton C, Dinan TG, Cryan JF. The neuropharmacology of butyrate: The bread and butter of the microbiota-gut-brain axis? *Neurochem Int*. 2016 Oct;**99**:110–132. PMID: 27346602.

22. Furusawa Y, Obata Y, Fukuda S, Endo TA, Nakato G, Takahashi D, Nakanishi Y, Uetake C, Kato K, Kato T, Takahashi M, Fukuda NN, Murakami S, Miyauchi E, Hino S, Atarashi K, Onawa S, Fujimura Y, Lockett T, Clarke JM, Topping DL, Tomita M, Hori S, Ohara O, Morita T, Koseki H, Kikuchi J, Honda K, Hase K, Ohno H. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature*. 2013 Dec 19;**504**(7480):446–450. PMID: 24226770.

23. Govindarajan K, MacSharry J, Casey PG, Shanahan F, Joyce SA, Gahan CGM. Unconjugated Bile Acids Influence Expression of Circadian Genes: A Potential Mechanism for Microbe-Host Crosstalk. *PLoS One*. 2016 Dec 1;**11**(12):e0167319. PMCID: PMC5132238.

24. Yanguas-Casás N, Barreda-Manso MA, Nieto-Sampedro M, Romero-Ramírez L. TUDCA: An Agonist of the Bile Acid Receptor GPBAR1/TGR5 With Anti-Inflammatory Effects in Microglial Cells. *J Cell Physiol*. 2017 Aug;**232**(8):2231–2245. PMID: 27987324.

25. Kuipers F, Bloks VW, Groen AK. Beyond intestinal soap—bile acids in metabolic control [Internet]. *Nature Reviews Endocrinology*. 2014;**10**:488–498. Available from: http://dx.doi.org/10.1038/nrendo.2014.60.

26. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, Mohney RP, Small KS, Bell JT, Steves CJ, Valdes AM, Spector TD, Menni C. The fecal metabolome as a functional readout of the gut microbiome [Internet]. *Nature Genetics*. 2018;**50**:790–795. Available from: http://dx.doi.org/10.1038/s41588-018-0135-7.

27. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten

RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol*. 2016 Aug 9;**34**(8):828–837. PMCID: PMC5321674.

28. Gao V, Turek F, Vitaterna M. Multiple classifier systems for automatic sleep scoring in mice [Internet]. *Journal of Neuroscience Methods*. 2016;**264**:33–39. Available from: http://dx.doi.org/10.1016/j.jneumeth.2016.02.016.

29. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010 May;**7**(5):335–336. PMCID: PMC3156573.

30. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y,

Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019 Aug;**37**(8):852–857. PMID: 31341288.

31. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton [Internet]. *Aquatic Microbial Ecology*. 2015;**75**:129–137. Available from: http://dx.doi.org/10.3354/ame01753.

32. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012 Aug;**6**(8):1621–1624. PMCID: PMC3400413.

33. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* [Internet]. 2017 Mar;**2**(2). Available from: http://dx.doi.org/10.1128/mSystems.00191-16 PMCID: PMC5340863.

34. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea [Internet]. *The ISME Journal*. 2012;**6**:610–618. Available from: http://dx.doi.org/10.1038/ismej.2011.139.

35. Mirarab S, Nguyen N, Warnow T. SEPP: SATé-enabled phylogenetic placement. *Pac Symp Biocomput*. 2012;247–258. PMID: 22174280.

36. Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T. Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science*. 2009 Jun 19;**324**(5934):1561–1564. PMID: 19541996.

37. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007 Aug;**73**(16):5261–5267. PMCID: PMC1950982.

38. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. *ISME J*. 2011 Feb;**5**(2):169–172. PMCID: PMC3105689.

39. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vázquez-Baeza Y, Birmingham A, Hyde ER, Knight R. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*. 2017 Mar 3;**5**(1):27. PMCID: PMC5335496.

40. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*. 2013 Sep;**31**(9):814–821. PMCID: PMC3819121.

41. Protsyuk I, Melnik AV, Nothias L-F, Rappez L, Phapale P, Aksenov AA, Bouslimani A, Ryazanov S, Dorrestein PC, Alexandrov T. 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. *Nat Protoc*. 2018 Jan;**13**(1):134–154. PMID: 29266099.

42. Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests [Internet]. *Pattern Recognition Letters*. 2010;31:2225–2236. Available from: http://dx.doi.org/10.1016/j.patrec.2010.03.014.

43. Genuer R, Poggi J-M, Tuleau-Malot C. VSURF: An R Package for Variable Selection Using Random Forests [Internet]. *The R Journal*. 2015;**7**:19. Available from: http://dx.doi.org/10.32614/rj-2015-018.

44. Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience*. 2013 Nov 26;**2**(1):16. PMCID: PMC4076506.

45. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat Methods*. 2018 Oct;**15**(10):796–798. PMCID: PMC6235622.

46. Bowers S. 16S microbiome dataset of mice subjected to 5 day sleep disruption protocol. EBI-ENA. (accession no. ERP113564). {Unpublished raw data}. 2019.

47. Vitaterna M. Metabolomic study on the effect of sleep disruption on mice. In Database https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=8f3141b17a1e4b5886df0d4c515f2a16 (accession no. MSV000080630) {unpublished raw data.}. Retrieved from Database https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=8f3141b17a1e4b5886df0d4c515f2a16 (accession no. MSV000080630) {unpublished raw data.}2019.

48. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005 Dec;**71**(12):8228–8235. PMCID: PMC1317376.

49. Codling C, O'Mahony L, Shanahan F, Quigley EMM, Marchesi JR. A molecular analysis of fecal and mucosal bacterial communities in irritable bowel syndrome. *Dig Dis Sci*. 2010 Feb;**55**(2):392–397. PMID: 19693670.

50. Reber SO, Siebler PH, Donner NC, Morton JT, Smith DG, Kopelman JM, Lowe KR, Wheeler KJ, Fox JH, Hassell JE Jr, Greenwood BN, Jansch C, Lechner A, Schmidt D, Uschold-Schmidt N, Füchsl AM, Langgartner D, Walker FR, Hale MW, Lopez Perez G, Van Treuren W, González A, Halweg-Edwards AL, Fleshner M, Raison CL, Rook GA, Peddada SD, Knight R, Lowry CA. Immunization with a heat-killed preparation of the environmental bacterium Mycobacterium vaccae promotes stress resilience in mice. *Proc Natl Acad Sci U S A*. 2016 May 31;**113**(22):E3130–9. PMCID: PMC4896712.

51. Yun Y, Kim H-N, Kim SE, Heo SG, Chang Y, Ryu S, Shin H, Kim H-L. Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort [Internet]. *BMC Microbiology*. 2017;**17**, Available from: http://dx.doi.org/10.1186/s12866-017-1052-0.

52. Pielou EC. *An introduction to mathematical ecology*. John Wiley & Sons; 1969.

53. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional [Internet]. *Frontiers in Microbiology*. 2017;**8**,. Available from: http://dx.doi.org/10.3389/fmicb.2017.02224

54. Koliada A, Syzenko G, Moseiko V, Budovska L, Puchkov K, Perederiy V, Gavalko Y, Dorofeyev A, Romanenko M, Tkach S, Sineok L, Lushchak O, Vaiserman A. Association between body mass index and Firmicutes/Bacteroidetes ratio in an adult Ukrainian population [Internet]. *BMC Microbiology*. 2017;**17**, Available from: http://dx.doi.org/10.1186/s12866-017-1027-1.

55. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*. 2005 Aug 2;**102**(31):11070–11075. PMCID: PMC1176910.

56. Gautam A, Kumar R, Chakraborty N, Muhie S, Hoke A, Hammamieh R, Jett M. Altered fecal microbiota composition in all male aggressor-exposed rodent model simulating features of post-traumatic stress disorder. *J Neurosci Res*. 2018 Jul;**96**(7):1311–1323. PMID: 29633335.

57. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, -M. Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. Proposed minimum reporting standards for chemical analysis [Internet]. *Metabolomics*. 2007;**3**:211–221. Available from: http://dx.doi.org/10.1007/s11306-007-0082-2.

58. Russell DW, Setchell KD. Bile acid biosynthesis. *Biochemistry*. 1992 May 26;**31**(20):4737–4749. PMID: 1591235.

59. Foley MH, O'Flaherty S, Barrangou R, Theriot CM. Bile salt hydrolases: Gatekeepers of bile acid metabolism and host-microbiome crosstalk in the gastrointestinal tract [Internet]. *PLOS Pathogens*. 2019;**15**:p.e1007581. Available from: http://dx.doi.org/10.1371/journal.ppat.1007581.

60. Vavassori P, Mencarelli A, Renga B, Distrutti E, Fiorucci S. The bile acid receptor FXR is a modulator of intestinal innate immunity. *J Immunol*. 2009 Nov 15;**183**(10):6251–6261. PMID: 19864602.

61. Mertens KL, Kalsbeek A, Soeters MR, Eggink HM. Bile Acid Signaling Pathways from the Enterohepatic Circulation to the Central Nervous System. *Front Neurosci*. 2017 Nov 7;**11**:617. PMCID: PMC5681992.

62. Banno N, Akihisa T, Tokuda H, Yasukawa K, Higashihara H, Ukiya M, Watanabe K, Kimura Y, Hasegawa J-I, Nishino H. Triterpene acids from the leaves of Perilla frutescens and their anti-inflammatory and antitumor-promoting effects. *Biosci Biotechnol Biochem*. 2004 Jan;**68**(1):85–90. PMID: 14745168.

63. Wali SO, Qutah K, Abushanab L, Basamh R 'a, Abushanab J, Krayem A. Effect of on-call-related sleep deprivation on physicians' mood and alertness. *Ann Thorac Med*. 2013 Jan;**8**(1):22–27. PMCID: PMC3573553.

64. Peterson AL, Goodie JL, Satterfield WA, Brim WL. Sleep Disturbance during Military Deployment [Internet]. *Military Medicine*. 2008;**173**:230–235. Available from: http://dx.doi.org/10.7205/milmed.173.3.230.

65. Karl JP, Margolis LM, Madslien EH, Murphy NE, Castellani JW, Gundersen Y, Hoke AV, Levangie MW, Kumar R, Chakraborty N, Gautam A, Hammamieh R, Martini S, Montain SJ, Pasiakos SM. Changes in intestinal microbiota composition and metabolism coincide with increased intestinal permeability in young adults under prolonged physiological stress. Am J Physiol Gastrointest Liver Physiol. 2017 Jun 1;**312**(6):G559–G571. PMID: 28336545.

66. Bercik P, Park AJ, Sinclair D, Khoshdel A, Lu J, Huang X, Deng Y, Blennerhassett PA, Fahnestock M, Moine D, Berger B, Huizinga JD, Kunze W, McLean PG, Bergonzelli GE, Collins SM, Verdu EF. The anxiolytic effect of Bifidobacterium longum NCC3001 involves vagal pathways for gut-brain communication. *Neurogastroenterol Motil*. 2011 Dec;**23**(12):1132–1139. PMCID: PMC3413724.

67. Messaoudi M, Violle N, Bisson J-F, Desor D, Javelot H, Rougeot C. Beneficial psychological effects of a probiotic formulation (Lactobacillus helveticus R0052 and

Bifidobacterium longum R0175) in healthy human volunteers. *Gut Microbes*. 2011 Jul;**2**(4):256–261. PMID: 21983070.

68. Steenbergen L, Sellaro R, van Hemert S, Bosch JA, Colzato LS. A randomized controlled trial to test the effect of multispecies probiotics on cognitive reactivity to sad mood. *Brain Behav Immun*. 2015 Aug;**48**:258–264. PMID: 25862297.

69. Bassett SM, Lupis SB, Gianferante D, Rohleder N, Wolf JM. Sleep quality but not sleep quantity effects on cortisol responses to acute psychosocial stress [Internet]. *Stress*. 2015;**18**:638–644. Available from: http://dx.doi.org/10.3109/10253890.2015.1087503.

70. Guyon A, Morselli LL, Balbo ML, Tasali E, Leproult R, L'Hermite-Balériaux M, Van Cauter E, Spiegel K. Effects of Insufficient Sleep on Pituitary-Adrenocortical Response to CRH Stimulation in Healthy Men. *Sleep* [Internet]. 2017 Jun 1;**40**(6). Available from: http://dx.doi.org/10.1093/sleep/zsx064 PMCID: PMC6075556.

71. Suchecki D, Tiba PA, Tufik S. Paradoxical sleep deprivation facilitates subsequent corticosterone response to a mild stressor in rats. *Neurosci Lett*. 2002 Mar 1;**320**(1-2):45–48. PMID: 11849760.

72. Aho V, Ollila HM, Rantanen V, Kronholm E, Surakka I, van Leeuwen WMA, Lehto M, Matikainen S, Ripatti S, Härmä M, Sallinen M, Salomaa V, Jauhiainen M, Alenius H, Paunio T, Porkka-Heiskanen T. Partial sleep restriction activates immune response-related gene expression pathways: experimental and epidemiological studies in humans. *PLoS One*. 2013 Oct 23;**8**(10):e77184. PMCID: PMC3806729.

73. Vgontzas AN, Zoumakis E, Bixler EO, Lin H-M, Follett H, Kales A, Chrousos GP. Adverse effects of modest sleep restriction on sleepiness, performance, and inflammatory cytokines. *J Clin Endocrinol Metab*. 2004 May;**89**(5):2119–2126. PMID: 15126529.

74. Tang Y, Preuss F, Turek FW, Jakate S, Keshavarzian A. Sleep deprivation worsens inflammation and delays recovery in a mouse model of colitis [Internet]. *Sleep Medicine*. 2009;10:597–603. Available from: http://dx.doi.org/10.1016/j.sleep.2008.12.009.

75. Bryant RA, Creamer M, O'Donnell M, Silove D, McFarlane AC. Sleep disturbance immediately prior to trauma predicts subsequent psychiatric disorder. *Sleep*. 2010 Jan;**33**(1):69–74. PMCID: PMC2802249.

76. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectrom*. 2016 Dec;**27**(12):1897–1905. PMCID: PMC5110944.

77. Thomas C, Pellicciari R, Pruzanski M, Auwerx J, Schoonjans K. Targeting bile-acid signalling for metabolic diseases. *Nat Rev Drug Discov*. 2008 Aug;**7**(8):678–693. PMID: 18670431.

78. Parséus A, Sommer N, Sommer F, Caesar R, Molinaro A, Ståhlman M, Greiner TU, Perkins R, Bäckhed F. Microbiota-induced obesity requires farnesoid X receptor. *Gut*. 2017 Mar;**66**(3):429–437. PMCID: PMC5534765.

79. Li F, Jiang C, Krausz KW, Li Y, Albert I, Hao H, Fabre KM, Mitchell JB, Patterson AD, Gonzalez FJ. Microbiome remodelling leads to inhibition of intestinal farnesoid X receptor signalling and decreased obesity. *Nat Commun*. 2013;**4**:2384. PMCID: PMC6595219.

80. Joyce SA, MacSharry J, Casey PG, Kinsella M, Murphy EF, Shanahan F, Hill C, Gahan CGM. Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut. *Proc Natl Acad Sci U S A*. 2014 May 20;**111**(20):7421–7426. PMCID: PMC4034235.

81. Antalis TM, Shea-Donohue T, Vogel SN, Sears C, Fasano A. Mechanisms of disease: protease functions in intestinal mucosal pathobiology. *Nat Clin Pract Gastroenterol Hepatol*. 2007 Jul;**4**(7):393–402. PMCID: PMC3049113.

82. Sessenwein JL, Baker CC, Pradhananga S, Maitland ME, Petrof EO, Allen-Vercoe E, Noordhof C, Reed DE, Vanner SJ, Lomax AE. Protease-Mediated Suppression of DRG Neuron Excitability by Commensal Bacteria. *J Neurosci*. 2017 Nov 29;**37**(48):11758–11768. PMCID: PMC5707769.

83. Guo C-J, Chang F-Y, Wyche TP, Backus KM, Acker TM, Funabashi M, Taketani M, Donia MS, Nayfach S, Pollard KS, Craik CS, Cravatt BF, Clardy J, Voigt CA, Fischbach MA. Discovery of Reactive Microbiota-Derived Metabolites that Inhibit Host Proteases. *Cell*. 2017 Jan 26;**168**(3):517–526.e18. PMCID: PMC5302092.

84. Banno N, Akihisa T, Tokuda H, Yasukawa K, Taguchi Y, Akazawa H, Ukiya M, Kimura Y, Suzuki T, Nishino H. Anti-inflammatory and Antitumor-Promoting Effects of the Triterpene Acids from the Leaves of Eriobotrya japonica [Internet]. *Biological & Pharmaceutical Bulletin*. 2005;**28**:1995–1999. Available from: http://dx.doi.org/10.1248/bpb.28.1995

85. Wang F, Hua H, Pei Y, Chen D, Jing Y. Triterpenoids from the resin of Styrax tonkinensis and their antiproliferative and differentiation effects in human leukemia HL-60 cells. *J Nat Prod*. 2006 May;**69**(5):807–810. PMID: 16724846.

86. Chen L, Brar MS, Leung FCC, Hsiao WLW. Triterpenoid herbal saponins enhance beneficial bacteria, decrease sulfate-reducing bacteria, modulate inflammatory

intestinal microenvironment and exert cancer preventive effects in ApcMin/+ mice. *Oncotarget*. 2016 May 24;**7**(21):31226–31242. PMCID: PMC5058752.

87. Li F, Wang D, Xu P, Wu J, Liu L, Liu X. Identification of the metabolites of anti-inflammatory compound clematichinenoside AR in rat intestinal microflora. *Biomed Chromatogr*. 2013 Dec;**27**(12):1767–1774. PMID: 23852993.

88. Ma W, Song J, Wang H, Shi F, Zhou N, Jiang J, Xu Y, Zhang L, Yang L, Zhou M. Chronic paradoxical sleep deprivation-induced depression-like behavior, energy metabolism and microbial changes in rats. *Life Sci*. 2019 May 15;**225**:88–97. PMID: 30953642.

89. Franklin CL, Ericsson AC. Microbiota and reproducibility of rodent models. *Lab Anim*. 2017 Mar 22;**46**(4):114–122. PMCID: PMC5762113.

90. Ma BW, Bokulich NA, Castillo PA, Kananurak A, Underwood MA, Mills DA, Bevins CL. Routine habitat change: a source of unrecognized transient alteration of intestinal microbiota in laboratory mice. *PLoS One*. 2012 Oct 17;**7**(10):e47416. PMCID: PMC3474821.

91. Belenky G, Wesensten NJ, Thorne DR, Thomas ML, Sing HC, Redmond DP, Russo MB, Balkin TJ. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J Sleep Res*. 2003 Mar;**12**(1):1–12. PMID: 12603781.

92. Pejovic S, Basta M, Vgontzas AN, Kritikou I, Shaffer ML, Tsaoussoglou M, Stiffler D, Stefanakis Z, Bixler EO, Chrousos GP. Effects of recovery sleep after one work week of mild sleep restriction on interleukin-6 and cortisol secretion and daytime sleepiness and performance. *Am J Physiol Endocrinol Metab*. 2013 Oct 1;**305**(7):E890–6. PMCID: PMC3798707.