

## ABSTRACT

MELLINGER, ALLYSON LEE. Developing High Resolution Mass Spectrometry Strategies for Studying Temporal Changes in Disease States. (Under the direction of Dr. David C. Muddiman).

Mass spectrometry is a versatile and sensitive tool for analytical measurements, utilized across the various facets of systems biology. The increasing performance of mass spectrometers has significantly expanded the capabilities of bioanalytical research. The accessibility of instruments capable of achieving high resolving powers at quick analysis speeds calls for expanded methodologies across the sampling of biofluids, tissues, and other sample matrices to study human health and disease. This work describes the development of strategies utilizing high resolution mass spectrometry to investigate how molecules change in different disease states over time.

In this work, shotgun proteomics was employed to analyze cerebrospinal fluid (CSF) collected from patients with the neurodegenerative disease amyotrophic lateral sclerosis (ALS). These samples were collected at various points in the patients' progression of the disease. Prior to analysis of these precious samples, a study was performed to determine the most effective data acquisition strategy for the identification and quantitation of peptides in cerebrospinal fluid. After identifying data independent acquisition as the preferred method (in terms of both peptide identifications and quantitative precision), we employed it to perform three main analyses of the ALS CSF data including: 1) an examination of peptide variability in CSF; 2) the association of inflammation with rate of disease progression; and 3) developing a mixed-effects model to best explain variation in the ALS- functional rating scale score of these patients. We found that intra-individual variability of abundances contributed just 4% to overall variance, indicating a tightly regulated CSF proteome. A moderately significant correlation ( $p < 0.1$ ) was observed between inflammation and disease progression rates of four patients. Finally, our mixed-effects model was developed using a least absolute shrinkage and selection operator for variable selection and

employed forward selection to minimize Akaike's information criterion for the final model, which utilized changes in abundances of 28 peptides to model ALS progression in these patients.

This work also describes the development of a functional mass spectrometry imaging (fMSI) method to study changes in the metabolic activity of glutathione in mouse mammary tissues. Glutathione plays a significant role in cancer cell viability and therapy response but has yet to be clearly linked with tumor clinical stage, histological grade, or therapy response in patients. It's possible the functional capacity (i.e., the tumor's ability to maintain glutathione levels over time) would be a better indicator of such characteristics. Measuring this capacity and its change over cross-sections of heterogeneous tumor tissue requires the temporal and spatial change measurements characteristic to fMSI techniques. We highlight the importance of such methods and describe the use of multiple infusions of differentially labeled isotopologues of glycine via stable-isotope labeling technology to track their time-dependent incorporation into the glutathione molecular pool within mouse tissues.

We first present a proof-of-concept study in mouse liver tissue and the development of the percent isotope enrichment tool within the MSiReader software to quantify and visualize changes. We utilized infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI) to perform these MSI studies and high-resolution accurate mass-mass spectrometry to fully resolve the three isotopologues of the enriched glutathione. These tools were then employed to map incorporation of the isotopologues into mouse 4T1 mammary tumor tissues. We then performed a voxel-by-voxel analysis to produce slope heatmaps for visualizing the time dependency of enrichment in various morphological areas of the tumor sections. We found areas where despite high steady-state levels of glutathione (mainly at tumor peripheries) we detected low metabolic

activity. We also mapped variability in glycine isotopologue delivery across tumors, allowing us to account for variations in tissue perfusion in observed metabolic activity.

© Copyright 2022 by Allyson Lee Mellinger

All Rights Reserved

Developing High Resolution Mass Spectrometry Strategies for Studying Temporal Changes in  
Disease States

by  
Allyson Lee Mellinger

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Chemistry

Raleigh, North Carolina  
2022

APPROVED BY:

---

David C. Muddiman  
Committee Chair

---

Erin S. Baker

---

Xiaoqing Liu

---

Yi Xiao

**DEDICATION**

I dedicate this document in full to my wonderful family, Mom, Dad, and Mandy. For 25 years of love and complete support.

## BIOGRAPHY

Allyson Mellinger was born in Port Orange, FL and spent many of her early years in the sunshine: going to the beach, swimming in the pool, and 95% of the time, playing soccer. At age seven, she was interviewed by the local newspaper and asked, “What do you want to be when you grow up?” Her answer was “to be a scientist, majoring in chemistry to study medicine and cure rare diseases” (following a long gold-medal soccer career, of course.) While the latter did not work out, she did follow through with a chemistry Bachelor of Science degree from Wake Forest University in 2018, accompanied by a minor in mathematics. While at WFU, she was lucky enough to meet Dr. Christa Colyer, who steered her on the path towards analytical chemistry with an NSF funded opportunity to perform academic, industrial, and international research. During her research internship at Ameritox in Greensboro, NC, Allyson was exposed to the beauty of mass spectrometry. With a little help from her Ameritox mentors, Dr. Oneka Cummings, Dr. Erin Strickland, and Dr. Greg McIntire, she was able to meet Dr. David Muddiman who agreed to advise her, alongside Dr. Michael Bereman (1981-2021) in pursuit of a Ph.D. in analytical chemistry.

## ACKNOWLEDGMENTS

To start off, I owe a huge thank you to all of my undergraduate mentors for giving me my start. To my Ameritox friends, Dr. Oneka Cummings, Dr. Erin Strickland, and Dr. McIntire: thank you for bringing me into this mass spec world and for letting me hang around you all even after my initial internship was over. I hope to continue seeing you around at future conferences!

To Dr. Christa Colyer: being assigned to you as an advisee was one of the best things to come out of my time at Wake Forest... worth every single penny of student loans! All of the advice, kindness, mentorship (even today)... priceless. You were such a great role model to me, in addition to singlehandedly putting me on the path I am today with the NSF grant invitation. I deeply appreciate it all and look forward to catching up on future W-S trips.

I would like to thank one of my first mentors in my proteomics endeavors, Dr. Jeff Enders. Thank you for all the training and counseling in the early days of my degree pursuit, all without complaint, with total patience, and with a sense of humor. Thanks also for emphasizing the importance of creating excellent figures and for getting our entire group started with CorelDraw. I have now grown a dependence on the software and will likely be a customer for life.

I would like to acknowledge and thank Dr. Taufika Williams for the mentorship, specifically in the Fall semester of 2020 during my stint as a METRIC TA in the proteomics group. Thank you for your patience with me and for never hesitating to provide assistance (or reassurance) when needed! I learned so much from you, even outside the science, and enjoyed our collaboration on the placenta project.

I would like to extend a huge thank you to every past colleague I would consider a “group member” I have ever worked with in my time at NCSU. Beginning with Rubia Martin, thank you for your kindness and collaboration during my time in the Bereman group. To Dr. Måns Ekelöf,

for teaching me the importance of humility and kindness, especially as an influential senior student. I hope we can continue the yearly ASMS reunions! Thank you to Dr. Sitora Khodjanizayova for your mentorship on the beginnings of the GSH project, for being such a great example of hard work, and for all the encouragement and friendship you've shown me even after leaving the lab. Thank you to Dr. Jackie Kalmar for the yearly Christmas cookies and for being the first to let me know that graduate school would be difficult (and that was okay.) Thank you to my neighbor Dr. Anqi Tu for setting an excellent standard as a researcher and presenter, and for putting things in a good perspective for me, especially towards the end of your grad school career. Thank you to Dr. Caleb Bagley, for many words of advice and (of course) for answering my very first way-too-long email of graduate school. Thank you to Ken Garrard, for always being around when someone needs a hand with something and for bringing the PIE tool to fruition within MSiReader.

I would also like to acknowledge many of my collaborators. Without them there would be very little to my projects. Thank you to Dr Emily Griffith for helping me with the statistics on a few different projects. It was a pleasure working with you and learning so much about how to look at data from you! To my collaborators in the Gamcsik group, Dr. Gamcsik and Zahid. Thank you so much for trusting me with this project, and for your continued excitement about its progress. It was an honor to work with you, and I appreciate all the encouragement along the way.

I would like to acknowledge and thank the current members of the Muddiman group. I am very proud of how hard you have all worked over the past couple years. To my fellow island-officer Hellena, (Dr Hongxia Bai !!!!) thank you for being calm, cool, and collected in the face of challenge and for inspiring me to always want to do the same! To Ying Xi, the new fearless leader of the group: Thank you for all the laughs and love, and for taking over the task of keeping our

(turbulent) folks in line. To Alex Sohn, one of the kindest people I know (and potentially most artistic... are you sure about science??) Thank you for making everyone in the group feel special and for all the selfless work you do to keep the group afloat. To my desk buddy Cristina Arciniega: thank you for the rays of sunshine and positivity you bring into the office, for laughing at my jokes, and for forgiving me (and not even acting upset!!) for the time I drenched your desk with my water bottle. Sorry again!!! To Kevan Knizner, thanks for the constant reminders to just be chill, for always being excited to talk science. And especially for making the meme board into everything I had hoped and dreamed it would become. To Tana Palomino, for loving Muf and talking cats with me!! And for becoming the next glycan person... I think we all thank you for that, especially since you've accomplished so much so far!! Which brings me to Mary Wang, another lucky glycan-er: thank you for being so sweet and for all the friendship and for talking TSwift with me! Thank you to Alena Joignant, for bringing the office together over fun personality quizzes. I'll miss trying to make you laugh from the nook! This goes for Olivia Dioli as well, thank you for keeping us laughing and happy with all your time in the office. To my mentee Russ Kibbe: you made this last year such a breeze for me by being an excellent scientist. Thank you so much for all the help and for getting excited about mouse tumors with me!!! And putting up with my constant stream of consciousness, I mean my unique style of mentoring... And to everyone: thank you so much for all the laughter and positivity in the past year. It made the return to normal office life so positive, special and wonderful. I never thought I would see the day where I was sad that grad school was ending, and I attribute a large portion of that feeling to you all. I will miss laughing and talking science with every single one of you! What will I do without you all laughing at my jokes and hyping me every minute of the day?!?!

I would like to thank all my non-science friends for all their help and support along the way. First, to those who managed to do this at a long distance. Thank you to Felicia for the visits and holiday talks and for always checking in on me. You're my oldest, closest friend and every time we catch up it feels as though no time has passed at all. Your constant support these years has meant the world to me. To Kristina, I've missed you so much in New York. Thank you for never being more than a phone call away and, of course, for consistently sending memes. They have really kept me going and laughing throughout these past years!

A few special thank you's to those I've met along the way: ToCarolynn, one of my first true friends here in Raleigh. Thank you for putting up with me as a roommate and for being so inclusive with your own friends. I think the world of you and can't wait to see you succeed with your own science. And to my earth sign girl gang - thank you for the hours (days, months, years) of venting and chatting. These moments are some of my favorites and are what really helped keep me going. I cannot wait for future visits and ASMS's together. Specifically: To (soon to be Dr/Mrs!!) Kaylie Kirkwood, my very first mentor. Thank you for putting up with me in my first few neurotic months as a graduate student. Right away, you set such a great standard for me as a scientist, which is incredible considering you were still an undergraduate student. I'm so thankful you decided to stay in Raleigh and we could continue our friendship outside the science. I've said this many times, but I think you have the best example of work-life balance I have ever seen and it set such a great example for me. I'm so proud and lucky to call you my friend. To Dr Melanie Odenkirk: our recruitment dinner is one of my favorite memories. I could tell even then that you are a force to be reckoned with. Thank you for the constant loyalty and perspective you can give on every situation. I don't think anyone gets as openly excited about others' successes as you, thank you for every note and congratulations over these past four years. I'm so proud of you and

all your work and growth these past four years. Finally, to Dr Crystal Pace: thank you for supporting me in literally every facet of my life. I am not sure where I would be if we had not taken that pottery class together. Thank you for letting me lean on you always, and for all the laughs and tears (and tears while laughing.) And for loving Muf more than anyone, which needs to be mentioned since it directly contributed to my ability to attend so many conferences this past year. Thank you for the years of setting a great example as a scientist and mentor. I've learned so much from you, again in all areas of my life. I look forward to seeing how you decorate my wing in your house for all the future visits.

I am incredibly thankful and in debt to my first graduate co-chair, Dr Bereman. What an experience and opportunity for me to learn so early in my career, to witness real strength and courage in the face of the highest adversity. I learned so much more than chemistry and proteomics from our work together, invaluable lessons that I will carry through my life and time as a scientist. As importantly, I wish to thank the entire Bereman family for supporting our research and making possible all the precious time he spent helping answer my endless questions and walking me through projects.

A big thank you to my other wonderful committee members, who have set an excellent example for me on the NCSU campus. Thank you to Dr. Xiaojing Liu and Dr. Erin Baker for following along in my career these past years and for your excellent insight with my projects. A special thank you to Dr. Baker for specifically providing me a female role model and friend within the mass spectrometry community. Another special thank you to Dr. Yi Xiao for agreeing to be a late addition to my committee this year. I hope you all enjoy this work, and I look forward to showing you all my science!

The biggest thank you to my committee chair and PI, Dr. David Muddiman. I can never thank you enough for these past four years and for all I have learned and shaped up to be as a scientist. Thank you for all the encouragement, confidence, and support. Thank you for all the opportunities and for never discouraging me from chasing after them. Beyond the science, I have also learned so much about mentorship and independence from you, and I hope to one day be as meaningful and intentional of a mentor. I already look forward to catchups and piano bars at conferences in the future!!

To Brandon: Thank you for keeping me grounded and making me laugh as much as possible. It's been a hard four years for you as well, but your support means everything. And thank you for constantly reminding everyone how much smarter I am than you.

Finally, to my family: put simply, thank you for everything. I love you all, endlessly. Without you none of this would be possible for me (or at least not to the extent I feel I have succeeded.) The support did not start here, but really from all the years of knowing when to push me to be my best and as importantly when to keep me from holding myself back. It feels like everything in this document was really a group effort. To all my extended family, thank you for the constant excitement over my career path. More importantly, thank you for always providing a respite from graduate school filled always with love and laughter (which sounds cheesy, but that is literally how every visit seems to go.) To my Aunt Lynn and Auntie Shannon (and my W-S family down the road) for providing another level of this support and for setting such an important example for me of what it means to be a Ph.D. scientist. And for teaching me it is essential to prioritize happy hours, tailgates, kitchen dance parties and other things that make me happy. And finally, to my mom, dad, and sister Mandy. Sorry for all I've put you through!! Thank you for all the physical (moving and decorating my apartments) emotional (weekly phone calls, daily texts),

and financial (self-explanatory) support. And for setting the best possible examples of hard-work and caring for one another. I love you more than anything or anyone.

## TABLE OF CONTENTS

LIST OF TABLES .....	xv
LIST OF FIGURES .....	xvii
LIST OF PUBLICATIONS .....	xxi
<b>Chapter 1: Introduction</b> .....	<b>1</b>
1.1 Mass Spectrometry.....	1
1.1.1 High Resolution and Accurate Mass-Mass Spectrometry .....	2
1.2 Introduction to Amyotrophic Lateral Sclerosis.....	5
1.3 Shotgun Proteomics .....	6
1.3.1 Label-Free Sample Preparation.....	8
1.3.2 Liquid Chromatography Separation.....	9
1.3.3 Electrospray Ionization .....	10
1.3.4 MS2 Data Acquisition in Shotgun Proteomics .....	13
1.3.5 Peptide Identification and Quantification .....	14
1.4 Introduction to Cancer .....	17
1.5 Mass Spectrometry Imaging .....	17
1.5.1 Matrix-Assisted Laser Desorption Ionization.....	19
1.5.2 Infrared Matrix-Assisted Laser Desorption Electrospray Ionization.....	21
1.6 References.....	23
<b>Chapter 2: Elucidation of Effective Data Acquisition Parameters for Shotgun Proteomics to Assay CSF</b> .....	<b>36</b>
2.1 Introduction.....	36
2.2 Materials and Methods.....	37
2.2.1 Materials .....	37
2.2.2 Sample Preparation .....	38
2.2.3 LC-MS/MS .....	39
2.2.4 Chromatogram Library Generation.....	40
2.2.5 Spectral Library Generation.....	40
2.2.6 DIA database searching .....	41
2.2.7 DDA database searching.....	41
2.3 Results and Discussion .....	42

2.4 Conclusions.....	44
2.5 Acknowledgments.....	44
2.6 References.....	45
<b>Chapter 3: Analysis of the Proteome of CSF Derived from a Longitudinal Cohort to Understand Disease Progression .....</b>	<b>49</b>
3.1 Introduction.....	49
3.2 Materials and Methods.....	50
3.2.1 LC-MS/MS .....	50
3.2.2 Quality Control .....	51
3.2.3 Skyline .....	52
3.2.4 Multivariate Analysis.....	52
3.2.5 Global Inflammation Study.....	53
3.3 Results and Discussion .....	54
3.3.1 Peptide Variability in ALS.....	54
3.3.2 Longitudinal Biomarkers .....	58
3.4 Conclusions.....	64
3.5 Acknowledgements.....	64
3.6 References.....	65
<b>Chapter 4: A Detailed Method for Performing Discovery Proteomics of Human Placental Tissue.....</b>	<b>69</b>
4.1 Introduction.....	69
4.1.1 Key Points.....	69
4.1.2 Acronyms and Abbreviations .....	70
4.2 Materials and Methods.....	71
4.2.1 Materials .....	71
4.2.2 Chemicals.....	72
4.2.3 Collection of Tissues for Proteomic Analysis .....	73
4.2.3.1 Placental Collection and Storage .....	73
4.2.3.2 Placental Sampling Protocol.....	74
4.2.4 Preparation of Placental Tissues for Proteomic nanoLC-MS/MS Analysis .....	75
4.2.4.1 Preparation of Buffers Prior to Tissue Processing.....	75

4.2.4.2 Preparation of Reducing and Alkylating Agent Aliquots Prior to Proteolytic Digestion .....	77
4.2.4.3 Preparation of Protein Solubilization Buffers Prior to Proteolytic Digestion .....	77
4.2.4.4 Safety Precautions.....	79
4.2.4.5 Tissue Preparation.....	79
4.2.4.6 Filter-Aided Sample Preparation (FASP) and Proteolytic Digestion .....	82
4.2.5 Separation and Analysis by nanoLC-MS/MS.....	88
4.3 Quality Control (QC) .....	92
4.4 Label-Free Data Analysis .....	93
4.5 Acknowledgements.....	99
4.6 Data Availability Statement.....	99
4.7 References.....	100
<b>Chapter 5: Highlighting Functional Mass Spectrometry Imaging Methods in</b>	
<b>Bioanalysis .....</b>	<b>102</b>
5.1 Introduction.....	102
5.2 Why prioritize functional measurements? .....	102
5.3 Functional Imaging .....	103
5.4 Nucleic Acid Metabolism .....	106
5.5 Lipid Metabolism.....	109
5.6 Bioenergetics.....	112
5.7 Amino Acid, Peptide, and Protein Metabolism .....	113
5.8 Drug Uptake and Metabolism.....	116
5.9 Conclusions.....	117
5.10 References.....	119
<b>Chapter 6: Multiple Infusion Start Time Mass Spectrometry Imaging of Dynamic</b>	
<b>SIL-Glutathione Biosynthesis Using Infrared Matrix-Assisted Laser Desorption</b>	
<b>Electrospray Ionization .....</b>	<b>126</b>
6.1 Introduction.....	126
6.2 Materials and Methods.....	129
6.2.1 Mouse Tissue Labeling and Collection.....	129

6.2.2	Sample Preparation .....	130
6.2.3	Mass Spectrometry Imaging .....	131
6.2.4	Data Analysis and Preparation.....	131
6.2.5	Percent Isotopic Enrichment (PIE) Tool in MSiReader .....	132
6.2.6	Droplet Analysis of SIL-GSH solutions for PIE Tool Validation .....	133
6.3	Results and Discussion .....	134
6.3.1	Multiple Infusion Start Times of Glycine Isotopologues .....	134
6.3.2	Confirmed Detection of GSH and SIL-GSH in Mouse Liver Tissue Using Mass Measurement and Spectral Accuracy.....	135
6.3.3	Mapping of Labels Across Tissue Sections and Normalization .....	137
6.3.4	Development and Validation of MSiReader PIE Tool .....	140
6.3.5	Mapping Percent Enrichment of GSH Isotopologues in Mouse Liver Tissue.....	143
6.4	Conclusions.....	145
6.5	Acknowledgements.....	146
6.6	References.....	147
<b>Chapter 7: Mapping Glutathione Metabolic Activity in Mouse Mammary Tumors</b>		
<b>Using Functional Mass Spectrometry Imaging.....</b>		
7.1	Introduction.....	154
7.2	Materials and Methods.....	155
7.2.1	Animal Studies.....	155
7.2.2	Sample Preparation .....	156
7.2.3	Mass Spectrometry Imaging, Data Analysis, and Preparation .....	156
7.2.4	Histology and Immunohistochemistry .....	157
7.3	Results.....	157
7.3.1	Distribution and Labeling of Glutathione .....	157
7.3.2	Distribution and Uptake of Labeled Glycines .....	161
7.3.3	Time Dependence of Label Uptake and Incorporation.....	162
7.3.4	Necrosis and Neutrophil Infiltration Identified by Tumor Histology.....	165
7.4	Discussion and Conclusions .....	166
7.5	Acknowledgements.....	171

7.6 References.....	172
<b>Chapter 8: Investigating Glycerate Production from Intestinal Fructose Metabolism</b>	
<b>Elevated by Dietary Fat in Mouse Pancreas using Mass Spectrometry Imaging.....</b>	<b>176</b>
8.1 Introduction.....	176
8.2 Materials and Methods.....	177
8.2.1 Mass Spectrometry Imaging by IR-MALDESI-MSI.....	178
8.2.2 Data Collection and Analysis.....	179
8.3 Results and Discussion .....	180
8.3.1 Validating SIL-Fructose and SIL-Glycerate Detected in Mouse Intestine.....	180
8.3.2 Islet Estimation in Pancreas Tissues .....	183
8.3.3 SIL-Fructose Deposition in Gavaged Mouse Pancreas Sections .....	184
8.3.4 Increased Deposition of SIL-Fructose Derived Glycerate in HFD Mouse Pancreas.....	186
8.4 Conclusions.....	189
8.5 Acknowledgements.....	189
8.6 References.....	190
<b>APPENDICES .....</b>	<b>193</b>
Appendix A: Supplemental Information for Chapter 2 .....	194
Appendix B: Supplemental Information for Chapter 3.....	200
Appendix C: Supplemental Information for Chapter 6.....	205
Appendix D: Supplemental Information for Chapter 7 .....	208

## LIST OF TABLES

<b>Table 3.1</b>	Description of ALS patient cohort and digestion cycles .....	50
<b>Table 4.1</b>	Preparation of rinsing solution, protease inhibitor in water .....	75
<b>Table 4.2</b>	Preparation of lysis buffer, 50 mM ABC, and 1% SDC (w/v) with protease inhibitor (500 mL).....	76
<b>Table 4.3</b>	Preparation of reducing agent, dithiothreitol (500 mM) .....	77
<b>Table 4.4</b>	Preparation of alkylating agent, iodoacetamide (500 mM).....	77
<b>Table 4.5</b>	Preparation of 50 mM ABC with 1% (w/v) SDC .....	78
<b>Table 4.6</b>	Preparation of 50 mM ABC (only).....	78
<b>Table 4.7</b>	Preparation of 50 mM ABC and 8 M urea .....	78
<b>Table 4.8</b>	Rinsing blood from samples.....	80
<b>Table 4.9</b>	Tissue homogenization.....	80
<b>Table 4.10</b>	Tissue sonication .....	81
<b>Table 4.11</b>	Centrifugation and removal of lysate .....	82
<b>Table 4.12</b>	Aliquoting and diluting .....	83
<b>Table 4.13</b>	Denaturing proteins by breaking disulfide bonds between cysteine residues .....	83
<b>Table 4.14</b>	Alkylating reactive thiols to prevent reformation of disulfide bonds .....	84
<b>Table 4.15</b>	Transferring samples to MWCO filtration units .....	85
<b>Table 4.16</b>	Washing twice with buffer containing 50 mM ABC and 8 M urea .....	86
<b>Table 4.17</b>	Washing twice with buffer containing 50 mM ABC .....	87
<b>Table 4.18</b>	Proteolytic digestion.....	87
<b>Table 4.19</b>	Elution of tryptic peptides .....	88
<b>Table 4.20</b>	Quenching of enzyme.....	88
<b>Table 4.21</b>	nanoLC parameters for two-hour separation.....	89

<b>Table 4.22</b> Mass spectrometric parameters (Data Dependent Acquisition) .....	90
<b>Table 4.23</b> Processing workflow nodes and basic parameters .....	94
<b>Table 4.24</b> Consensus workflow nodes and basic parameters .....	96
<b>Table 4.25</b> Summary of protein identification results .....	98
<b>Table 4.26</b> Summary of hemoglobin identification across samples .....	99
<b>Table 5.1</b> Spatial resolutions of highlighted fMSI methods .....	104
<b>Table 6.1</b> Median percent isotope enrichment values of analyzed tissue.....	143
<b>Table A.1</b> Summary of raw file sources for building CSF spectral library.....	195
<b>Table B.1</b> Summary of peptides used for global inflammation study.....	202
<b>Table B.2</b> Summary of all fixed effects included in linear-mixed effects model to predict ALS-FRS score.....	204
<b>Table C.1</b> GSH isotopologue correction matrix.....	206

## LIST OF FIGURES

<b>Figure 1.1</b>	Summary of introduction sections and relation to following chapters.....	1
<b>Figure 1.2</b>	General components in a Thermo Fisher Scientific hybrid orbitrap mass spectrometer. ....	4
<b>Figure 1.3</b>	FASP workflow for enzymatic digestion of proteins .....	9
<b>Figure 1.4</b>	Electrospray ionization and proposed ion formation mechanisms.....	12
<b>Figure 1.5</b>	Tandem mass spectrometry via HCD of tryptic peptides.....	14
<b>Figure 1.6</b>	Duty cycle structures of data dependent and data independent acquisitions .....	15
<b>Figure 1.7</b>	Mass spectrometry imaging workflow .....	18
<b>Figure 1.8</b>	Laser ablation event in matrix-assisted laser desorption ionization.....	20
<b>Figure 1.9</b>	Infrared matrix-assisted laser desorption electrospray ionization source .....	22
<b>Figure 2.1</b>	Distribution of highly abundant proteins in human CSF. ....	36
<b>Figure 2.2</b>	Comparison of DIA and DDA for CSF shotgun proteomics... ..	42
<b>Figure 2.3</b>	Variability of peptide replicates between abundance quartiles. ....	43
<b>Figure 2.4</b>	Distributions of median centered peptide abundances. ....	44
<b>Figure 3.1</b>	Boxplot of logged peptide abundances of longitudinal CSF samples.....	52
<b>Figure 3.2</b>	ALS patient cohort and proteomics workflow. ....	55
<b>Figure 3.3</b>	Peptide variability in ALS patient cohort.....	57
<b>Figure 3.4</b>	Disease progression rates as a function of inflammatory response in ALS .....	59
<b>Figure 3.5</b>	Mixed-effects model generation to predict ALS-FRS score via peptide measurements. ....	61
<b>Figure 3.6</b>	Network interactions of proteins with peptide representation in model .....	63
<b>Figure 4.1</b>	Photograph of placenta tissue collection .....	75

<b>Figure 4.2</b> Photograph of tissue homogenization procedure. ....	81
<b>Figure 4.3</b> Processing workflow tree for Proteome Discoverer.....	94
<b>Figure 4.4</b> Consensus workflow tree for Proteome Discoverer.....	96
<b>Figure 5.1</b> Functional processes mapped by MSI methods .....	105
<b>Figure 5.2</b> MIMS imaging data showing heterogeneity in tumor cells .....	108
<b>Figure 5.3</b> Isotope ratio images of phosphatidylserine in the brain tissue of mice administered deuterated drinking water .....	111
<b>Figure 6.1</b> Summary of workflow employing MIST protocol for mapping of functional heterogeneity of GSH across mouse liver tissue via IR-MALDESI-MSI.....	130
<b>Figure 6.2</b> User-defined workflow settings for the PIE tool in MSiReader. ....	133
<b>Figure 6.3</b> Summary of glycine incorporation into reduced glutathione.....	135
<b>Figure 6.4</b> Mass spectrum of GSH from liver tissue sections. ....	136
<b>Figure 6.5</b> Normalization of GSH peaks in liver tissue section.....	138
<b>Figure 6.6</b> Normalized heatmaps of steady-state GSH isotopologues in labeled and unlabeled liver tissues .....	139
<b>Figure 6.7</b> Validation of PIE tool via IR-MALDESI droplet analysis .....	140
<b>Figure 6.8</b> Percent isotope enrichment of labeled liver section.....	144
<b>Figure 7.1</b> Summary of glycine isotopologues and MIST protocol for mouse tumor enrichment .....	158
<b>Figure 7.2</b> Optical image and GSH heatmaps for excised 4T1 tumor of MIST labeled mouse .....	159
<b>Figure 7.3</b> Box-and-whisker plots summarizing GSH PIE data from tumor replicates .....	160

<b>Figure 7.4</b> Positive mode heatmaps of glycine PIE and SIL-urea in mouse tumor .....	161
<b>Figure 7.5</b> Isotopologue incorporation and GSH slope heatmaps from mouse tumor .....	163
<b>Figure 7.6</b> Hematoxylin-eosin-stained tissue from mouse tumor.....	165
<b>Figure 8.1</b> Summary of pancreas study experimental design .....	178
<b>Figure 8.2</b> Heatmaps from SIL-fructose dosed mouse intestine.....	181
<b>Figure 8.3</b> Spectral accuracy in mouse intestine tissue .....	182
<b>Figure 8.4</b> Analysis of islet counts in mouse pancreas sections .....	184
<b>Figure 8.5</b> Heatmaps of SIL-fructose/glucose species detected in mouse pancreas tissues .....	185
<b>Figure 8.6</b> Abundance heatmaps of SIL-fructose derived glycerate detected in analyzed pancreas tissues.....	187
<b>Figure 8.7</b> Bar graph showing the average numbers of SIL-glycerate peaks detected by group normalized to pancreas tissue areas.....	189
<b>Figure C.1</b> Carbon and sulfur counting for GSH peak in liver tissue .....	205
<b>Figure C.2</b> Comparison of normalization methods of GSH in liver tissue. ....	206
<b>Figure C.3</b> PIE heatmaps of GSH for labeled and unlabeled liver replicates .....	207
<b>Figure D.1</b> Steady state heatmaps of GSH in tumor replicates .....	208
<b>Figure D.2</b> PIE heatmaps of GSH in tumor replicates.....	209
<b>Figure D.3</b> Steady-state and PIE heatmaps of GSH in negative control tumor.....	210
<b>Figure D.4</b> Positive mode heatmaps of glycine and SIL-urea in tumor replicates .....	211
<b>Figure D.5</b> Slope heatmaps of SIL-GSH in tumor replicates .....	212
<b>Figure D.6</b> Slope heatmaps of SIL-glycine .....	212

**Figure D.7** Discrete heatmap showing the number of GSH

isotopologues detected at each tissue voxel in liver section .....213

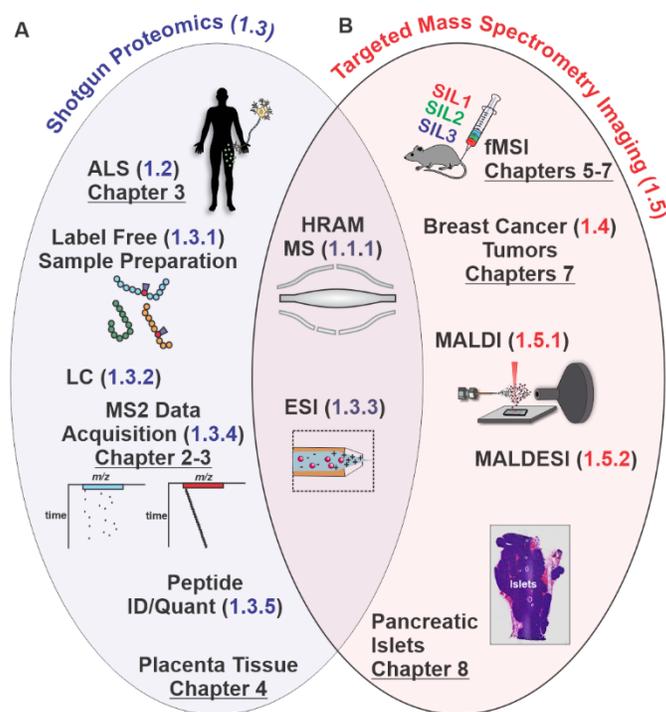
## LIST OF PUBLICATIONS

1. Bereman, M.S., Kirkwood, K.I., Sabaretnam, T., Furlong, S., Row, D.B., Guillemin, G.J., **Mellinger, A.L.**, Muddiman, D.C., “Metabolite Profiling Reveals Predictive Biomarkers and the Absence of  $\beta$ -Methyl Amino-L-Alanine in Plasma from Individuals Diagnosed with Amyotrophic Lateral Sclerosis”, *Journal of Proteome Research*, 2020, 19, 8, 3276–3285.
2. **Mellinger, A.L.**, Griffith, E.H., Bereman, M.S., “Peptide Variability and Signatures Associated with Disease Progression in CSF Collected Longitudinally from ALS Patients”, *Analytical and Bioanalytical Chemistry*, 2020, 412, 5465-5475.
3. **Mellinger, A.L.**, McCoy, K., Minior, D.A.T., Williams, T.I., “Discovery Proteomics of Human Placental Tissue”, *Rapid Communications in Mass Spectrometry*, 2021, e9189.
4. **Mellinger, A.L.**, Garrard, K.P., Rabbani, Z.N., Gamcsik, M.P., Muddiman, D.C., “Multiple Infusion Start Time Mass Spectrometry Imaging of Dynamic SIL-Glutathione Biosynthesis Using Infrared Matrix-Assisted Laser Desorption Electrospray Ionization”, *Journal of Proteome Research*, 2022, 21, 3, 747–757.
5. Yanru, W., Wong, C.W., Chiles, E.N., **Mellinger, A.L.**, Bae, H., Jung, S., Peterson, T., Wang, J., Negrete, M., Xiang, K., Huang, Q., Wang, L., Jang, C., Muddiman, D.C., Su, X., Williamson, I., Shen, X., “Glycerate Production from Intestinal Fructose Metabolism Elevated by Dietary Fat Induces Glucose Intolerance Through  $\beta$ -cell Damage,” *Cell Metabolism*, 2022, 34, 7, 1042-1053, e6.
6. **Mellinger, A.L.**, Muddiman, D.C., Gamcsik, M.P., “Highlighting Functional Mass Spectrometry Imaging (fMSI) Methods in Bioanalysis,” *Journal of Proteome Research*, 2022, 21, 1800-1807.
7. Kibbe, R.R., **Mellinger, A.L.**, Muddiman, D.C., “Novel Matrix Strategies for Improved Ionization and Spatial Resolution Using IR-MALDESI Mass Spectrometry Imaging,” *Journal of Mass Spectrometry*, 2022, 57(8), e4875.
8. **Mellinger, A.L.**, Kibbe, R.R., Rabbani, Z.N., Meritet, D., Muddiman, D.C., Gamcsik, M.P., “Mapping glutathione metabolic activity in mouse mammary tumors using functional mass spectrometry imaging”, *Free Radical Biology & Medicine*, 2022, submitted.

## CHAPTER 1: Introduction

### 1.1 Mass Spectrometry

Mass spectrometry (MS) is an increasingly vital analytical tool across all facets of biological research. The technique is used to measure relative signals or abundances of gas-phase ions separated by their mass-to-charge ratios (denoted  $m/z$ ). It can also be used to collect structural information on the ions via induced fragmentation. The primary components that make up a mass spectrometer platform are the 1) ionization source 2) mass analyzer and 3) detection system. Various types of each component exist, allowing for numerous molecules and compound classes to be measured and characterized for a wide variety of sample matrices. Mass spectrometry is key to the work presented here, and details describing the ionization sources and mass analyzers



**Figure 1.1.** Contents of introduction and how they relate to the following work. **A)** Several shotgun proteomics projects will be presented towards elucidating disease signatures of amyotrophic lateral sclerosis. **B)** Mass spectrometry imaging (MSI), specifically functional MSI (fMSI), methods were developed and used to study changes in glutathione flux over time in breast cancer tumor models.

relevant to this work will be expanded upon in the following sections. A summary of the following introduction is shown below in **Figure 1.1**. This figure outlines each of the following introductory subjects presented within this section and their context or relation to the subsequent chapters of this work.

### *1.1.1 High Resolution and Accurate Mass-Mass Spectrometry*

While there are various mass analyzers of great importance in the mass spectrometry field, the Orbitrap will be the focus of subsequent information presented due to its relevance to this work. Advances in separation and ionization techniques for complex samples developed a need for improved speed and accuracy of detection. The invention of the Orbitrap mass analyzer by Alexandrov Makarov in 2000<sup>1</sup> and its eventual integration into bench top mass spectrometers for commercialization quickly satisfied this need across the mass spectrometry community.<sup>2</sup>

The premise of the Orbitrap was based on the work of Kingdon from 1923<sup>3</sup> with fundamental changes to make it more amenable to MS measurements. These included: 1) the thin trapping wire became a spindle shaped electrode, and 2) the measurement of axial oscillation frequency of trapped ions around the central electrode was used rather than the rotation frequency.<sup>1</sup> Importantly, the addition of a low-pressure cell known as the C-trap allows for the accumulation and collisional cooling of ions outside of the Orbitrap (labeled in **Figure 1.2**.) This allowed the analyzer to be coupled with continuous ionization sources (e.g., electrospray, *1.3.3*) as ion packets are stored and cooled while previous transients take place.<sup>2,4</sup> Ions are detected by measuring the induced image current of ionic axial oscillations via two outer electrodes. Using fast Fourier transform algorithms, the periodic signals of these oscillations that are recorded in the time domain are mathematically transformed into frequencies. Frequency is related to  $m/z$  based on the following equation:

$$\omega = \sqrt{(z/m)k} \quad (1.1)$$

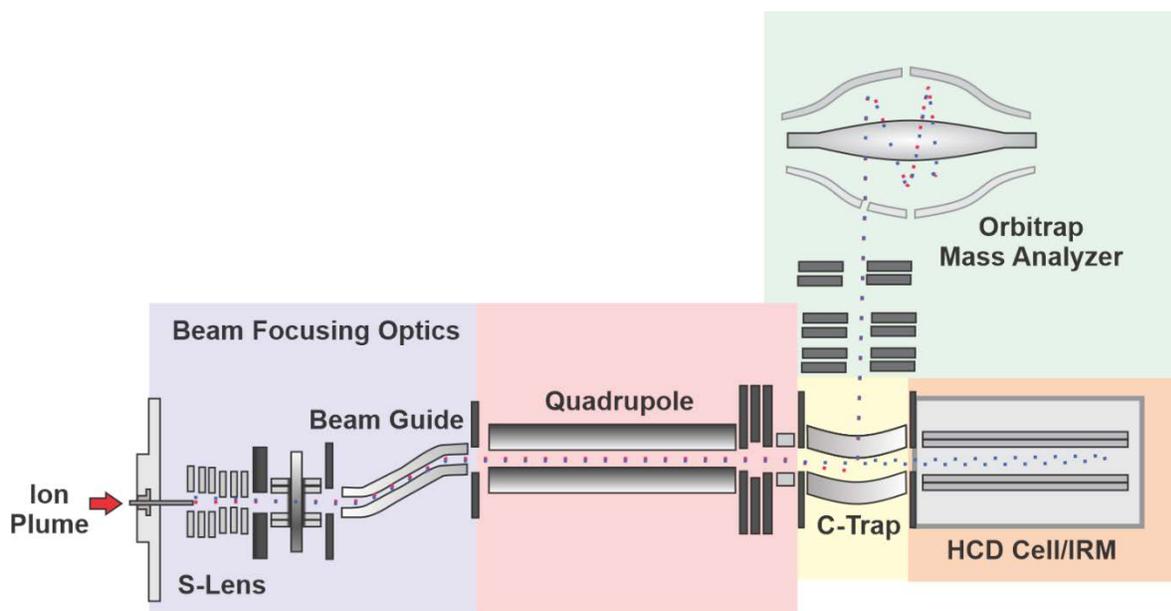
where  $\omega$  represents the axial oscillation frequency and  $k$  represents field curvature. Axial oscillation frequency is useful due to its independence from the initial position and velocities of ion packets. High resolving power and mass accuracy are characteristics of Fourier transform-based mass spectrometers. Resolving power is the ability of a system to distinguish peaks with similar  $m/z$ . Mass accuracy refers to the ratio of the error in a measured  $m/z$  to the true  $m/z$ .<sup>5</sup>

In the form of a benchtop mass spectrometer, the Orbitrap technology was revolutionary as an accessible and relatively low-maintenance way to make fast and high resolution measurements with high mass accuracy (within 5 ppm via external calibration). High mass accuracy can improve confidence in identifying molecules from their measured  $m/z$  value. In addition, an Orbitrap system performs these high resolving power measurements at considerably fast acquisition rates (up to 41 Hz). This is helpful in applications such as proteomics where multiple measurements can be made across chromatographic peaks when coupled with nanoflow liquid chromatography (1.3.2), while still maintaining desired sensitivity and high mass accuracy.<sup>6</sup>

Several other important improvements were made to the original commercial Orbitrap mass spectrometers, also highlighted in **Figure 1.2**. The first was the equipping of a higher energy collision-induced dissociation (HCD) cell. This cell provides beam-type collisional dissociation at a higher energy than the traditional “collision-induced dissociation” fragmentation method, which can provide structural information about the fragmented molecules.<sup>7</sup> The addition of a quadrupole mass analyzer in the Q Exactive series of Orbitrap mass spectrometers allowed for the filtering of species outside specified precursor  $m/z$  ranges prior to Orbitrap detection. It also allows users to isolate selected precursors for fragmentation in the HCD cell, improving subsequent identification. This filtering and fragmentation process is known as tandem mass spectrometry

(MS/MS)<sup>8-10</sup> and is discussed further in the context of proteomics in 1.3.4. Measurements in this work were made on instruments of this series, specifically the Q Exactive HF and HF-X (“HF” referring to the higher field strength of these Orbitrap analyzers) mass spectrometers, along with an Exploris 240 mass spectrometer. Rather than an HCD cell, the latter instrument of the Exploris series is equipped with an “ion routing multipole” (IRM). The IRM serves both as an HCD cell and to trap and cool ions prior to routing stored them back to C-trap.<sup>11</sup> The distinctive ion paths for MS1 scans collected with these instruments is illustrated in **Figure 1.2**.

For reference, the Q Exactive HF can collect spectra with a resolving power of up to 240,000 at 200  $m/z$  at a speed of 1.5 Hz. Fragmentation spectra can also be collected at fast scan speeds but with lower resolving powers. In the context of proteomics (1.3) this quick speed of isolation and fragmentation led to high increases in peptide identifications by maximizing the duty



**Figure 1.2.** General schematic of a Thermo Fisher Scientific hybrid quadrupole orbitrap mass spectrometer. Important hardware components such as the mass filtering quadrupole (pink), collisional cooling C-trap (yellow), and the collision cell (termed the Higher Energy Collisional Dissociation, HCD, Cell in the Q Exactive Series and Ion Routing Multiple, IRM, in the Exploris series) for ion fragmentation (orange) are labeled and color coded, in addition to the characteristic high field Orbitrap mass analyzer (green). Ion beam paths for a typical MS1 scan are outlined with dotted lines in the Q Exactive Series (pink) and the Exploris series (blue.)

cycle of the instrument within the width of chromatographic peaks.<sup>10</sup> With improvements to the efficiency of ion transfers and peak selection algorithms, the Q Exactive HF-X mass spectrometer can collect MS/MS scans at speeds above 40 Hz while maintaining a resolving power of 7,500 at 200 *m/z*. In addition, the QE HF-X is equipped with an electrodynamic ion funnel and high-capacity transfer tube, replacing the S-lens of the QE HF and the Exploris 240 instruments as a brighter ion source.<sup>9</sup> These main components and their configuration within the mass spectrometers are summarized in **Figure 1.2**.

## 1.2 Introduction to Amyotrophic Lateral Sclerosis

Amyotrophic Lateral Sclerosis (ALS) is a disease characterized by rapid degeneration of both upper and lower motor neurons in the brain and spinal cord. Patients experience progressive muscle weakness, spasticity, and eventually total body paralysis. Following the loss of diaphragm and other muscles that control breathing, patients typically pass away from respiratory distress just only 3-5 years after symptom onset.<sup>12,13</sup> No therapy has been successful at reversing, halting, or significantly slowing disease progression. Translation of therapies have been hindered for various reasons over the years but largely due to the poor understanding of the underlying pathological mechanisms and to the lack of biomarkers for accurate diagnosis, prognosis, or measurement of therapeutic response.<sup>14,15</sup>

Only 5-10% of ALS cases are classified as “familial” and are associated with family inheritance. This leaves most cases classified as “sporadic,” where the disease is seemingly onset completely at random. Clinicians can only diagnose ALS by excluding similar diseases, a procedure that takes 1-1.5 years on average. This delay wastes valuable time, induces anxiety and, most notably, prevents early intervention.<sup>16,17</sup> Once diagnosed, disease progression is monitored using a 12-question survey called the ALS-Functional Rating Scale (ALS-FRS) where the degree

of impairment in everyday functions (e.g., walking, writing) is gauged on a 5-point scale (0-4 points)<sup>18,19</sup>. The survey is fast and cheap, but limitations include its subjectivity and nonlinear scale with degree of disability. An objective and quantitative assessment for the disease and its progression with time could accelerate accurate diagnosis, assist in effective therapy development, and be a useful tool in disease categorization and clinical trials.<sup>20,21</sup> This is the focus of the work presented in **Chapters 2 and 3** and will be further discussed there, as outlined by **Figure 1.1A**. The following topics in **1.3** introduce relevant techniques utilized to perform this work, along with the work presented in **Chapter 4**.

### **1.3 Shotgun Proteomics**

The term “proteomics” was first defined in the 1990’s and its coinage is credited to Marc Wilkins in 1994.<sup>22–24</sup> The year 2002 marked the notion that studying proteins was important and the broad definition of proteomics was given as the study of the “identities, quantities, structures, and biochemical and cellular functions of all proteins in an organism, organ, or organelle, and how these properties vary in space, time, and physiological state.”<sup>25</sup> Proteins play a crucial part in the central dogma of biology and maintain a high degree of interaction with each other and with other molecular components in a system.<sup>26</sup> Proteins exhibit rapid fluctuations in response to physiological changes within an organism (e.g., disease, nutritional, or pharmacological stimuli).<sup>27</sup> With their high degree of involvement in all major cellular processes,<sup>28,29</sup> they provide a measurable avenue for connecting the relationship between static genetic coding within cells to dynamic organism phenotypes. Resolving the genotype-phenotype link is an essential goal in human health and disease research.<sup>29</sup> The mapping of the human genome<sup>30–32</sup> set a template for discovery of proteins within humans. This achievement, along with major advances in proteomic

technologies (particularly in the field of mass spectrometry) over the past few decades, has significantly shaped the focus on proteomics.

There are three popular ways of using mass spectrometry to analyze proteins in various sample matrices. These include: 1) bottom-up 2) middle-down and 3) top-down proteomics. In the first technique, proteins are enzymatically digested, and the resulting peptides are separated (typically by liquid chromatography, *1.3.2*, or capillary electrophoresis) and analyzed by mass spectrometry. Middle-down proteomics is similar in that the proteins are enzymatically digested, but the size of the resulting peptides is typically larger. Finally, top-down proteomics investigates intact proteins in a sample without any digestion step. While the intact peptides or proteins can be measured directly, tandem mass spectrometry is usually employed for the production and detection of peptide or protein fragments. The fragmentation patterns produced act as unique pieces for the identification of the species.

“Shotgun proteomics” is the term used to describe the bottom-up analysis of a mixture of proteins. This coinage was established based on its similarities with shotgun genome sequencing. Herein, the protein pieces (i.e., peptides) are used to infer identification of the whole protein contained in the sample prior to digestion.<sup>33</sup> There are many advantages to using shotgun proteomics over other analytical techniques. By digesting proteins into peptides, the range of physiochemical properties of the analytes is reduced. In other words, the target analytes are switched from proteins ranging in molecular weights from 10s to 1000s of kDa to peptides with a relatively narrow range of sizes and chemical properties, making them much simpler to separate and ionize by common techniques.<sup>34</sup> The general throughput of mass spectrometry analyses is a key advantage. Shotgun proteomics is considered the “workhorse”<sup>34</sup> of proteomics and will be discussed further as it is the method of choice in the following work.

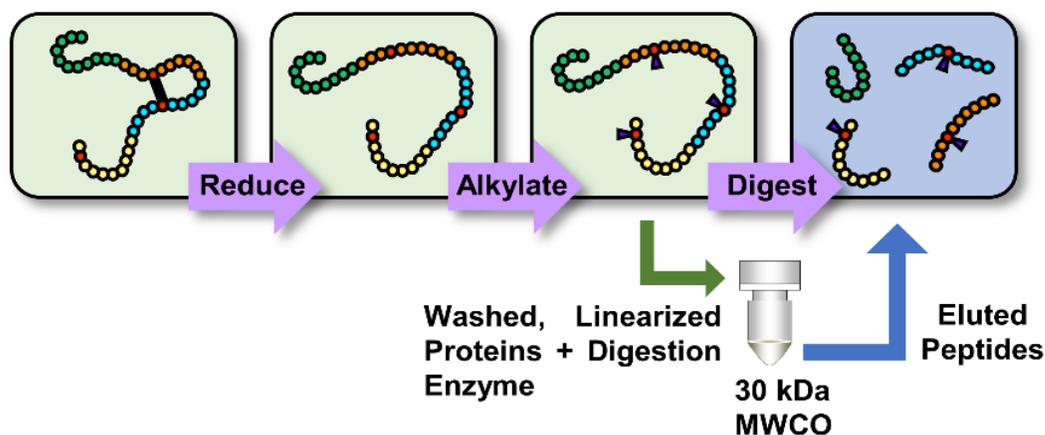
### *1.3.1 Label-Free Sample Preparation*

To generate a sample containing suitable peptides for separation, ionization, and analysis via MS, proteins in the starting matrix must be appropriately solubilized and digested, with interfering substances removed prior to analysis. Many variations of sample preparation protocols exist, and both the procedures and reagents utilized may be optimized on a per laboratory basis. However, any detergents and small molecules added to the sample prior to digestion can significantly hinder LC-MS/MS analyses. SDS in particular is difficult to remove from samples, is known to suppress peptide ion signals, and can collect on emitters, LC lines and columns, and MS inlets.<sup>35</sup> Therefore, it is imperative to filter or cleanse the samples for the removal of interfering molecules prior to analysis.

For this purpose, the filter-aided sample preparation (FASP) procedure has gained significant popularity in the field since its characterization by Wisniewski and coworkers. In this procedure a molecular weight cutoff filter (commonly referred to as a “spin filter”) is used to capture all molecules above the cutoff size (i.e., the large proteins) allowing them to be washed of detrimental small molecules that will be rinsed through the filter. Prior to placement on the filter, an important step to prepare the proteins is to linearize them and ensure all cleavable sites are exposed for the enzyme to digest. For this purpose, dithiothreitol (DTT) is used to reduce disulfide bonds between cysteines. Subsequently, an alkylation reaction is typically performed using iodoacetamide (IAM) to ensure the newly exposed reactive thiols do not reform these bonds.

Once the linearized proteins are transferred to the filter, a urea wash can be used to swap detergent molecules that were used for protein solubilization (e.g., SDS). The enzymatic digestion can then occur on the filter, cleaving the clean proteins into the desired final peptides, which are now appropriately sized to elute through the filter for collection.<sup>36</sup> Trypsin is the most used enzyme

for digestion. Its popularity is a result of its reliable cleavage at arginine and lysine carboxyls. This produces highly specific peptides capable of maintaining multiple charges and of an appropriate length (6-25 amino acids) for effective ionization.<sup>34</sup> These steps are summarized in **Figure 1.3**. A more detailed description of this process is presented in **Chapter 4**, specifically describing the preparation of human placental tissues for shotgun proteomics.



**Figure 1.3.** Schematic of enzymatic digestion of proteins into peptides over a molecular weight cutoff filter in the FASP protocol. Proteins are first denatured and linearized before being transferred to filter for wash and digestion.

### 1.3.2 Liquid Chromatography Separation

High performance liquid chromatography (HPLC) emerged in the mid 1970's as a powerful way to separate peptides by an array of properties, including differences in charge, size, isoelectric point, and hydrophobicity. The most commonly used separation technique in shotgun proteomics is reversed-phase liquid chromatography (RPLC).<sup>37</sup> In this technique, peptides are loaded onto a chromatography column packed with hydrophobic stationary phase. Mixed organic/aqueous phase solvents (e.g., acetonitrile and water) are then pumped through the column in gradient fashion to differentially elute the loaded peptides. In a gradient elution RPLC separation, the organic phase strength is increased as a function of time to allow for separation based on the peptides' hydrophobic character.

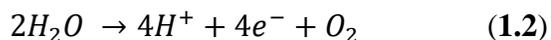
Advances within liquid chromatography technology have greatly furthered the power of shotgun proteomics. Today, LC-MS/MS shotgun proteomics experiments utilize C18 columns that are <100  $\mu\text{m}$  inner diameters packed with 3-5  $\mu\text{m}$  particles and flow rates in the hundreds of nanoliters per minute range. Longer analytical columns (15-100 cm) are also utilized to increase plate number and improve peptide resolution.<sup>6</sup> With these improvements, peak capacity and ionization efficiency can reach results similar to those achieved with multidimensional separations, even with low volumes of loaded sample.<sup>38,39</sup> Factors such as mobile phase compositions, flow rates, gradient length, column length and particle size, and column temperature are easily adjusted and optimized to reduce peptide chromatographic peaks widths, maximizing both the quality and quantity of peptide identifications in complex samples.<sup>40-42</sup> Hundreds of thousands of peptides are now routinely resolved, identified, and quantified with high sensitivity using these separation technologies coupled with mass spectrometry.

One advantage of RPLC as a peptide separation technique prior to ionization and mass analysis is its compatibility with electrospray ionization (ESI). After separation, peptides can be directly electrosprayed from emitter tips coupled to the analytical column. This ionization method and the mechanisms involved are further described in the following section (1.3.3). The resolution of peptides prior to ionization by electrospray is important to a successful comprehensive analysis of a complex sample of peptides.

### *1.3.3 Electrospray Ionization*

The introduction and augmentation of electrospray ionization was vital to the development and popularization of biological mass spectrometry.<sup>44</sup> Operating at atmospheric pressure, this method of ionization occurs by applying a high voltage to a capillary emitter containing the chargeable aqueous analytes. ESI is characterized as a “soft” method of producing ions, meaning

the technique itself induces very little (if any) fragmentation of analytes prior to their entrance into the mass spectrometer. In positive mode ionization, protons primarily contribute to charge production in solution. In portions this work, the source of these protons is from the oxidation of water at the metal surface of the capillary (**Equation 1.2**).



In this mode the ESI capillary acts as an anode while the mass spectrometer inlet acts as a cathode and electrons flow through wires connecting the systems, thus creating an electrochemical cell which facilitates the transport of the positively charged gaseous ion droplets towards the inlet of the mass spectrometer as they resolve current in the circuit (**Figure 4**). In negative mode, water is reduced via the application of a negative voltage, and therefore the opposite is true.<sup>43</sup> This mode was also used in a portion of work discussed herein.

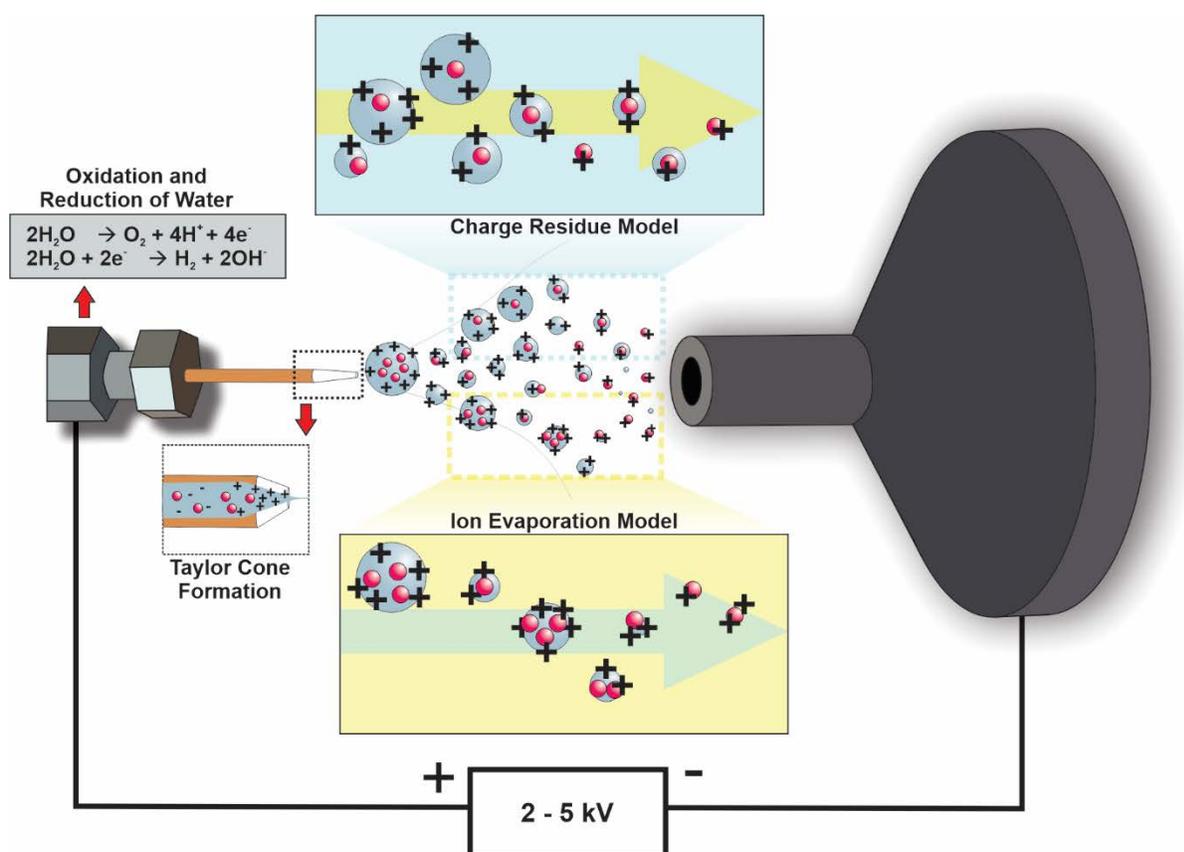
When the voltage is applied, the charged solution at the capillary tip forms a conical interface known as the “Taylor Cone”<sup>45</sup> from which the highly charged analyte containing droplets are emitted in a fine mist or “spray.” With the application of this electric charge, each of the droplets reach their Rayleigh<sup>46</sup> limit (**Equation 1.3**) of charge-to-surface area ( $q_r$ ) and consequently divide into smaller droplets by the surmounting of Coulombic forces over droplet surface tension. Repeated fissions and evaporations lead to the charged nanometer sized droplets that produce the gas-phase ions detected by the mass spectrometer.<sup>43</sup>

$$q_r = \sqrt{64\pi^2 \epsilon_0 \gamma r^3} \quad (1.3)$$

The exact mechanism of progeny droplet production is still debated, but the community generally agrees that the ion evaporation model (IEM) proposed by Iribarne and Thomson.<sup>47</sup> applies best to smaller molecules while the charge reduction model (CRM), the first model proposed by Dole and coworkers<sup>48</sup>, applies to larger globular macromolecules. The IEM states that charged analyte

containing droplets reach a high enough electric field that solvated ions are ejected from the surface of the droplet. In the CRM, each nanodrop contains a single analyte to which charge is transferred as the solvent surrounding it evaporates and undergoes fission cycles.<sup>43</sup> Both models are shown in **Figure 4**. As a gaseous spray, the analytes become suitable for detection by the mass spectrometer.

Advantages of the ESI source for proteomics includes 1) the softness of the source 2) the characteristic ability to routinely obtain multiply charged species and 3) the ease with which it is interfaced with LC separation systems. The softness of ESI facilitates the analysis of intact peptides by preventing their fragmentation prior to reaching the mass spectrometer. Obtaining highly charged species aids the dissociation of ions in tandem mass spectrometry experiments. It



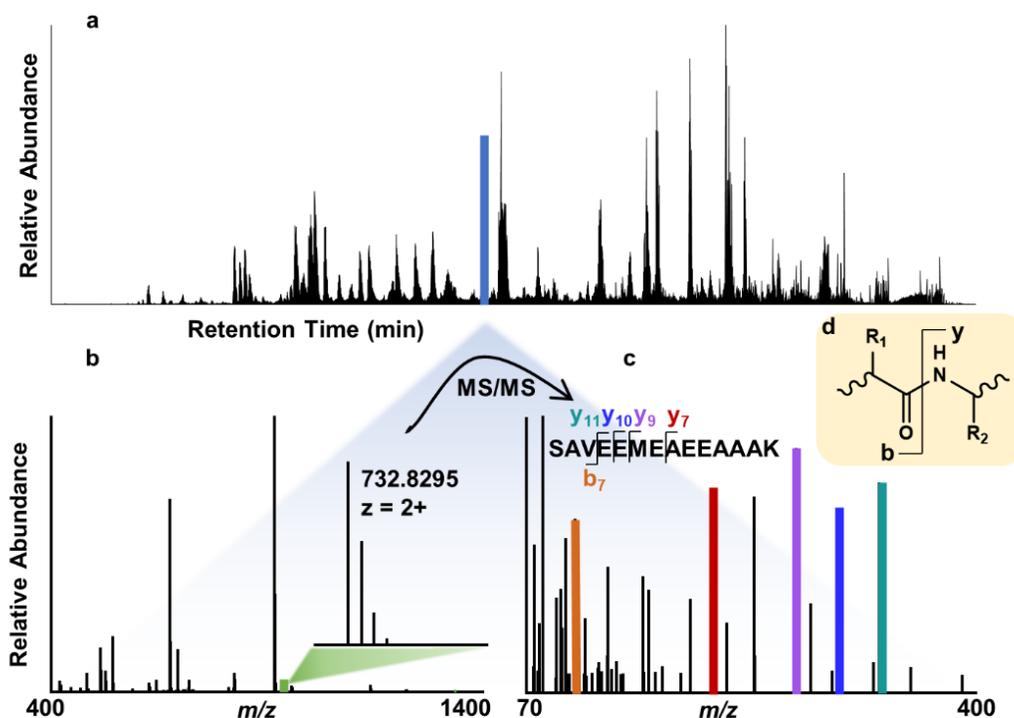
**Figure 1.4.** Schematic of electrospray ionization and the formation of ions via the charge residue model (CRM) and the ion evaporation model (IEM).

also allows for analysis of analytes with higher masses on instruments where the measurable  $m/z$  range is limited.<sup>43</sup>

#### *1.3.4 MS2 Data Acquisition in Shotgun Proteomics*

The traditional mode for collecting peptide identifications is data dependent acquisition (DDA). In this method, a duty cycle begins with a full scan of all eluting intact precursors at a given time in the chromatographic analysis and is followed by a user-defined number of fragmentation scans, before the cycle repeats. The most abundant precursor ions in the full scan are selected for MS/MS. These precursor isotopic distributions are recognized by the instrument software as multiply charged peptide species and are isolated by the quadrupole. A narrow enough window ( $\sim 2.0 m/z$ ) is used to isolate the isotopic distribution of that species without, in theory, simultaneously isolating other species that could convolute the fragmentation spectra. The isolated species are then subject to collisions for fragmentation and these fragments are mass analyzed. The “N” number of precursor species that can be selected for fragmentation from the MS1 scan is dependent on instrument scan speed, but typically ranges from 10-40 within a 1-2 second duty cycle. These methods are labeled Top N (e.g., “Top 20”) with regards to the selection of the N highest abundant species in the precursor scan.<sup>34</sup> This process is summarized in **Figure 5**.

Alternatively, a data independent acquisition (DIA) approach was proposed by Venable and coworkers in 2004. In this method, precursors are fragmented within larger isolation windows (in the approximate range of 4 – 25  $m/z$  width) completely independent of what species are detected in MS1 scans.<sup>49</sup> The number and isolation width of the scans are balanced to maintain an appropriate duty cycle for multiple measurements across chromatographic peaks, but still survey and fragment the entire  $m/z$  range of interest. As a result, multiple precursors are expected to be



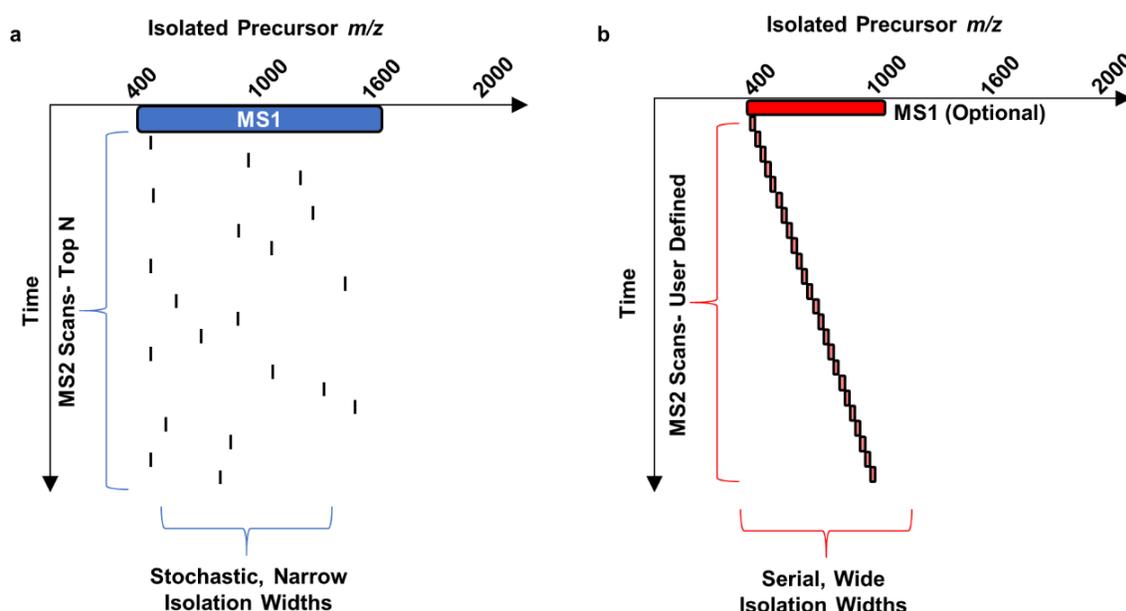
**Figure 1.5.** Components identifying a peptide via tandem mass spectrometry. **a)** Total ion chromatogram collected over a LC gradient separation of a sample. Eluting precursors are further separated by their mass-to-charge ratio. **b)** Precursor, or MS1, scan of peptides eluting in the highlighted retention time (blue) in **a**. Inset shows an enlarged isotopic distribution of the doubly charged precursor peak at highlighted (green)  $m/z$ . **c)** HCD fragmentation spectrum of precursor from **b**. Identified fragments are highlighted in the spectrum. **d)** HCD fragmentation along the peptide backbone primarily results in b- and y- ions.

fragmented in each scan. DIA methods can potentially alleviate issues associated with DDA via this comprehensive, user-defined, and serial fragmentation of the entire precursor  $m/z$  space of interest. A visual comparison of DIA and DDA duty cycles and window widths is shown in **Figure**

**1.6. Chapters 2-4** of this work further detail implementation of these techniques.

### 1.3.5 Peptide Identification and Quantification

Since hundreds of thousands of spectra are collected in any given experiment, manually identifying all peaks would be time-consuming and impractical. Hence, database searching is used to make these identifications. Spectrum centric algorithms, such as SEQUEST<sup>50</sup> for DDA data or DIA-Umpire<sup>51</sup> for DIA data, compare the collected spectra to theoretical spectra from protein



**Figure 1.6.** Example duty cycle structures of DDA and DIA methods. Each duty cycle takes approximately 1 second to complete before repeating to collect enough data points across a chromatographic peak. Each scan occurs on a ms time scale. **a)** Example of a DDA “Top 20” experiment. MS2 isolations are dependent on precursors detected in MS1 scan. **b)** Example of a DIA experiment with 24 windows isolating 25  $m/z$  width portions of the precursor space.

sequences predicted based on the organism’s genomic sequences. The user narrows the search space by inputting constraints, based on how the peptides of interest were generated and measured, from the theoretical proteins contained in the organism database (e.g., digestion enzyme, mass tolerance, fixed and variable post translational modifications, etc.). These comparisons are scored by the algorithm and the best scoring theoretical peptide matching a given MS/MS spectrum is reported as a “peptide spectral match” (PSM). An alternative option for DIA data is to search via peptide centric algorithms (e.g., EncyclopeDIA<sup>52</sup> and OpenSWATH<sup>53</sup>). Instead of matching each collected MS/MS spectrum to a peptide contained in a database, these algorithms attempt to search collected data for evidence of pre-specified peptides of interest, such as those contained in a database or in a spectral library.<sup>52,54</sup> In this case, the evidence is scored and ranked so the best scoring spectra for a given peptide of interest is reported as a peptide identification.<sup>54</sup>

In both cases, peptide identifications run the risk of error and false identifications. Therefore, further statistical analysis must be performed in order to correctly report identifications.<sup>55</sup> The most common method of reporting significance within a list of scored and ranked PSMs is to use a “decoy database” to model and test the null hypothesis that each identification was made by chance. This database contains shuffled or reversed amino-acid sequences from the target protein database. The querying of these decoy or “pseudo” peptides can help establish cutoff scores for evidence of confident identifications. Most commonly, false discovery rate (FDR) is employed for multiple hypothesis testing correction and can be estimated using the incorrect PSMs.<sup>56,57</sup> As the main objective of proteomics is to investigate on the protein level, it is important that these statistical significance tests are made on the PSM, peptide, and protein levels of identification.<sup>58</sup>

Within label-free proteomics, peptide quantitation may be made via spectral counting (specific to DDA data) or intensity-based methods. The first utilizes the number of PSMs to assess peptide quantity.<sup>59</sup> While this process involves more straightforward approaches to data extraction, caveats include biased quantitation towards peptides with different physicochemical properties and the increased risk of missing error associated with nonsystematic errors.<sup>34,60</sup> Another option is to use the extracted ion chromatograms (XICs) and compare integrated chromatographic peaks corresponding to the same peptide between analyses. Using peak areas for quantification is much more feasible with the improvements made in both high resolution accurate mass technology and the increased accessibility of faster and higher quality data extraction and analysis software.<sup>60,61</sup> Measurements of peptides made in this work are all based on peptide abundance and calculated peak areas.

## 1.4 Introduction to Cancer

This section and the following **1.5** serve to introduce information and techniques relevant to **Chapters 5-7** in this work (**Figure 1.1B.**) As one of the leading causes of death worldwide, cancer is a truly devastating disease, and is therefore the focus of techniques developed in the work within these chapters. In the United States alone, national expenditures reached \$150.8 billion in 2018 and will continue to rise with population age increases.<sup>62,63</sup> Afflicted tissues are characterized by rampant cellular proliferation and disrupted energy metabolism. There is a continued need for linking molecular profiles and potential biomarkers with tissue function in addition to assessment of variability across cellular distributions. Satisfying these needs would further treatment research and precision medicine capabilities in providing personalized cancer therapies. The ability to assess the spatial variability across the inflicted cellular distributions is particularly pertinent. Elucidating specific mechanisms of tumor growth will be highly relevant to therapeutic opportunities. In particular, the role of glutathione and oxidative stress remains a mysterious system open to opportunities.<sup>64</sup> This is discussed in more detail in **Chapter 7**. In this work specifically, we explore the utility of functional mass spectrometry imaging (fMSI) towards exploring these opportunities, specifically beginning in mouse models of breast cancer tumors.

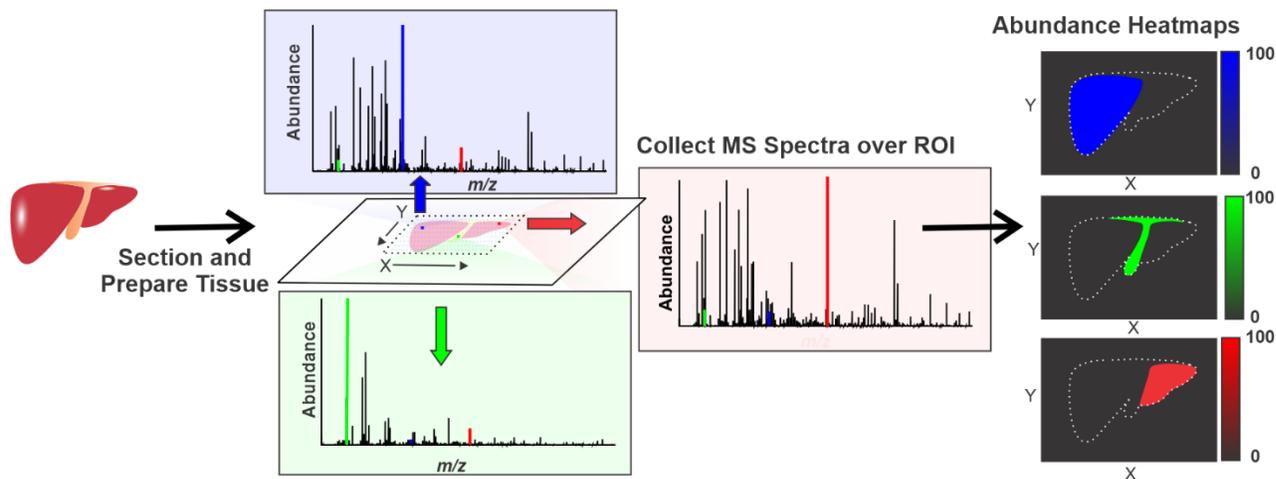
## 1.5 Mass Spectrometry Imaging

While LC-MS based proteomics provides the vital information on total protein levels in homogeneous samples, mass spectrometry may also be applied to provide spatial information alongside the molecular information measured. This technique is typically dubbed “mass spectrometry imaging” (MSI) or “imaging mass spectrometry” (IMS), although the former avoids confusion with the more common and first proposed acronym “Ion-Mobility MS,” also IMS. A multitude of ionization sources exist for this purpose and may typically be characterized as “laser-

based” or not (e.g., desorption electrospray ionization, DESI). Several relevant laser-based ionization techniques are specifically discussed in the following sections *1.5.1* and *1.5.2*.

The main MSI principle is the collection of MS data at the various X and Y locations of two-dimensional samples. The collected abundances are computationally mapped back to these locations to create a “heatmap” where areas of higher relative abundance of a molecule are typically “brighter” and thus “darker” where the molecule is less abundant or even absent. A heatmap can be generated for each detected  $m/z$  peak to visualize its spatial distribution. These heatmaps may also be compared with others to visualize colocalization or compared with histological data in multimodal techniques. A generalized workflow is summarized in **Figure 1.7**.

However, this field is not without its challenges. While usually involving a more direct mode of ionization and sampling, the typical lack of separation prior to ionization may lead to increased signal suppression. Additionally, although minimal sample preparation is typically employed, tissue sectioning and optimizing organic matrix selection (specific to matrix-assisted techniques, *1.5.1*) present further unique challenges (particularly in reproducibility) in performing MSI analyses. Finally, the high dimensionality and quantity of data produced, especially as MS instrumentation advances, in such experiments requires constant improvement of data analysis



**Figure 1.7.** General mass spectrometry imaging workflow.

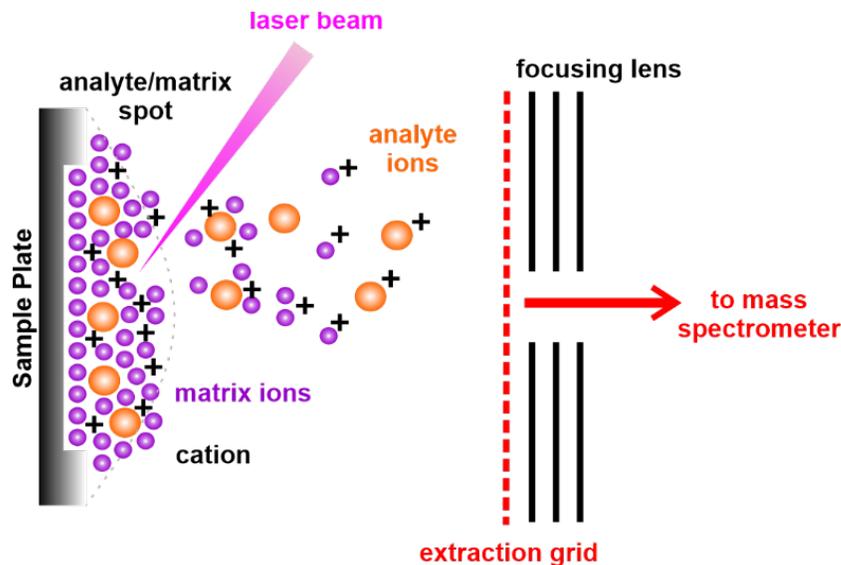
tools that can handle these large data sets. Despite such challenges, MSI has been significantly advanced and useful in a variety of imaging applications.<sup>65</sup> In **Chapter 5**, we specifically expand upon the use of MSI for functional measurements in a technique dubbed functional MSI (fMSI.)

### *1.5.1 Matrix-Assisted Laser Desorption Ionization*

Another common soft method of ionization called matrix-assisted laser desorption ionization (MALDI) was introduced by Hillencamp and Karas in the mid-1980's as a version of ultraviolet laser desorption ionization (UV-LDI) but with the application of an organic acid matrix. This novel technique was revolutionary in its ability to ionize large and nonvolatile species (>10 kDa) without fragmentation so that intact molecular ions could be detected.<sup>66</sup> The organic acid matrix is applied to absorb the energy of the laser used to irradiate the sample, resulting in ablation of both analyte and ionized matrix clusters from the sample.

To prepare samples for MALDI analysis, analytes are mixed or sprayed with the organic matrix. The matrix is selected based on the wavelength of the laser used for ablation and the sample types analyzed. In UV-MALDI for example, this is commonly pi-conjugated acids such as 2,5-dihydroxybenzoic acid (DHB) or  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA). In infrared (IR)-MALDI, matrices such as glycerol or succinic acid, whose O-H stretching bonds absorb in the IR-region, may be selected. After matrix application, the surrounding solvent is evaporated from the sample, leaving behind analytes co-crystallized with the matrix molecules. This might be in the form of a droplet spotted on a flat microscope slide, or a biological tissue sectioned to an even thickness as in mass spectrometry imaging (MSI).<sup>67</sup> The prepared sample is then ready for ablation by the source laser, which is focused on the sample surface and fired to release the analytes and matrix within intact clusters (**Figure 1.8**). While contributing greatly to the softness of this ionization source, the application of organic matrices results in disadvantages such as ion

suppression and isobaric interferences within the measured spectra. This ionization is traditionally performed under vacuum, although the use of atmospheric pressure (AP)-MALDI sources has grown in popularity.<sup>68</sup>



**Figure 1.8.** Schematic of laser ablation event on a matrix assisted laser desorption ionization (MALDI) source.

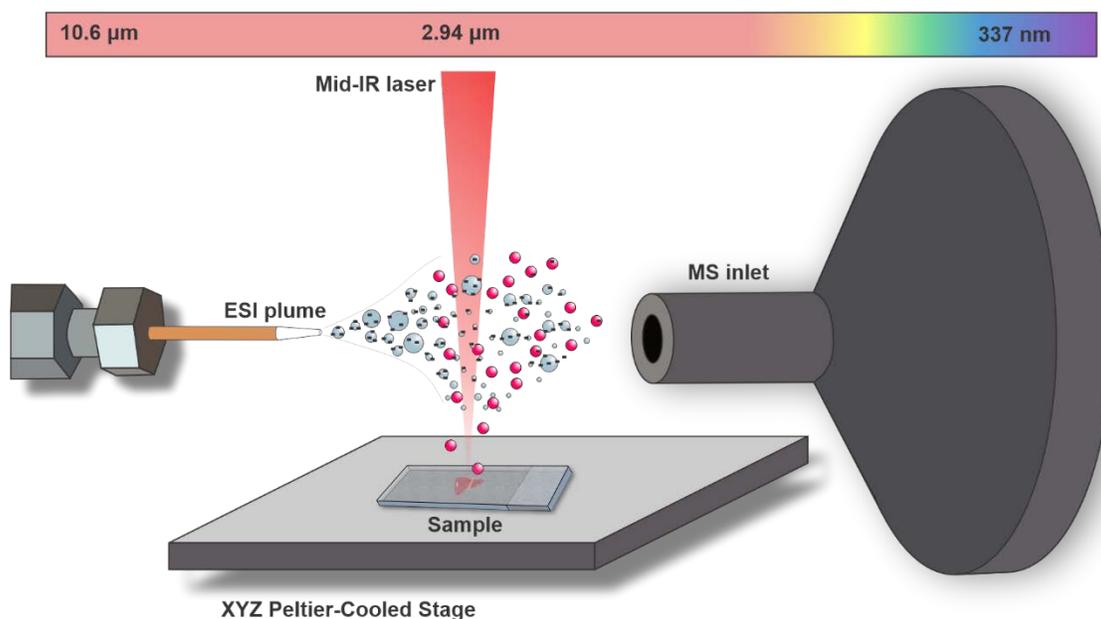
There are several theories of ion formation within MALDI, all of which involve interactions between the organic matrix and analytes, including via gas phase protonation and the “lucky survivor theory”, although the mechanism is highly affected by the various conditions involved sample preparation and laser characteristics (i.e., pulse length, wavelength, energy, etc.)<sup>69–71</sup> Characteristic to MALDI is the generation of predominately singly charged ions, both in the positive and negative modes. The gas-phase protonation theory states that the charged matrix and neutral analyte clusters collide in the gas phase to undergo a series of proton transferring reactions. This is dependent on the proton affinity of the analytes, which must be higher than that of the matrix for the reaction enthalpy to favor the proton transfer. The lucky survivor theory states that analytes may be multiply charged prior to both sample preparation and laser ablation and will undergo a series of neutralization/counter-neutralization reactions in the gas phase. Based on this

theory, the singly charged analytes are the “lucky survivors” of these neutralization reactions, likely due to slower rate of removal of the final single charge from their sufficiently reduced neutralization cross-section and thus increased probability of survival.<sup>71</sup>

MALDI eventually became the most popular ionization source used in MSI applications, first popularized by Caprioli and coworkers in the late 1990’s. Its high tolerance for salts and contaminants within the sample coupled with the ease with which laser firing may be rastered across a two-dimensional tissue makes it amenable to such analyses.<sup>65,67</sup>

### *1.5.2 Infrared Matrix-Assisted Laser Desorption Electrospray Ionization*

The infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI) source is an effective tool for informative MSI of biological tissues across many molecular species and tissue types. It combines advantages of both ESI and MALDI techniques, resulting in a soft ionization source<sup>72</sup> tolerant of high sample salt and contaminant content, and primarily used for MSI and high throughput screening applications.<sup>73</sup> In this work, MALDESI with an IR laser and an applied ice matrix, rather than an organic energy absorbing matrix as in MALDI, was utilized for MSI applications. This ice matrix facilitates the desorption of sample material as the mid-IR laser (2.97  $\mu\text{m}$  wavelength) pulse excites the O-H stretching modes of the water. The laser pulse ablates the sample material, allowing desorbed neutrals to enter the gas-phase and partition into the orthogonal electrospray where they are ionized via mechanisms similar to traditional ESI.<sup>72</sup> The utility of the ice matrix along with the ambient nature of the source allows biological samples analyzed by IR-MALDESI with minimal sample preparation.<sup>74</sup> A schematic of the IR-MALDESI source is shown in **Figure 1.9**. Advantages of this ionization source are further discussed in the context of this work within **Chapter 6**.



**Figure 1.9.** Schematic of IR-MALDESI source. An orthogonal electrospray plume (negative mode shown here) is established orthogonal to an XY translational stage, where the sample is placed. An ice layer is formed by cooling the stage and exposing it to controlled humidity prior to ablation via the mid IR laser. Ablated neutrals enter the ESI plume for ionization prior to entering the mass spectrometer inlet for mass analysis.

## 1.6 References

- (1) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72* (6), 1156–1162. <https://doi.org/10.1021/ac991131p>.
- (2) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. The Orbitrap: A New Mass Spectrometer. *Journal of Mass Spectrometry* **2005**, *40* (4), 430–443. <https://doi.org/10.1002/jms.856>.
- (3) Kingdon, K. H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys. Rev.* **1923**, *21* (4), 408–418. <https://doi.org/10.1103/PhysRev.21.408>.
- (4) Hardman, M.; Makarov, A. A. Interfacing the Orbitrap Mass Analyzer to an Electrospray Ion Source. *Anal. Chem.* **2003**, *75* (7), 1699–1705. <https://doi.org/10.1021/ac0258047>.
- (5) Scigelova, M.; Hornshaw, M.; Giannakopoulos, A.; Makarov, A. Fourier Transform Mass Spectrometry. *Mol Cell Proteomics* **2011**, *10* (7). <https://doi.org/10.1074/mcp.M111.009431>.
- (6) Makarov, A.; Scigelova, M. Coupling Liquid Chromatography to Orbitrap Mass Spectrometry. *Journal of Chromatography A* **2010**, *1217* (25), 3938–3945. <https://doi.org/10.1016/j.chroma.2010.02.022>.
- (7) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nat Methods* **2007**, *4* (9), 709–712. <https://doi.org/10.1038/nmeth1060>.
- (8) Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wiegand, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass Spectrometry-Based Proteomics Using

- Q Exactive, a High-Performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol Cell Proteomics* **2011**, *10* (9). <https://doi.org/10.1074/mcp.M111.011015>.
- (9) Kelstrup, C. D.; Bekker-Jensen, D. B.; Arrey, T. N.; Hogrebe, A.; Harder, A.; Olsen, J. V. Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J. Proteome Res.* **2018**, *17* (1), 727–738. <https://doi.org/10.1021/acs.jproteome.7b00602>.
- (10) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-Filter, High-Performance Quadrupole and an Ultra-High-Field Orbitrap Analyzer. *Mol Cell Proteomics* **2014**, *13* (12), 3698–3708. <https://doi.org/10.1074/mcp.M114.043489>.
- (11) Knizner, K. T.; Bagley, M. C.; Garrard, K. P.; Hauschild, J.-P.; Pu, F.; Elsen, N. L.; Williams, J. D.; Muddiman, D. C. Optimized C-Trap Timing of an Orbitrap 240 Mass Spectrometer for High-Throughput Screening and Native MS by IR-MALDESI. *J. Am. Soc. Mass Spectrom.* **2022**, *7*.
- (12) Zufiría, M.; Gil-Bea, F. J.; Fernández-Torrón, R.; Poza, J. J.; Muñoz-Blanco, J. L.; Rojas-García, R.; Riancho, J.; López de Munain, A. ALS: A Bucket of Genes, Environment, Metabolism and Unknown Ingredients. *Progress in Neurobiology* **2016**, *142*, 104–129. <https://doi.org/10.1016/j.pneurobio.2016.05.004>.
- (13) Boillée, S.; Vande Velde, C.; Cleveland, D. W. ALS: A Disease of Motor Neurons and Their Nonneuronal Neighbors. *Neuron* **2006**, *52* (1), 39–59. <https://doi.org/10.1016/j.neuron.2006.09.018>.
- (14) Gagliardi, D.; Meneri, M.; Saccomanno, D.; Bresolin, N.; Comi, G. P.; Corti, S. Diagnostic and Prognostic Role of Blood and Cerebrospinal Fluid and Blood Neurofilaments in

- Amyotrophic Lateral Sclerosis: A Review of the Literature. *International Journal of Molecular Sciences* **2019**, *20* (17), 4152. <https://doi.org/10.3390/ijms20174152>.
- (15) Petrov, D.; Mansfield, C.; Moussy, A.; Hermine, O. ALS Clinical Trials Review: 20 Years of Failure. Are We Any Closer to Registering a New Treatment? *Front. Aging Neurosci.* **2017**, *9*. <https://doi.org/10.3389/fnagi.2017.00068>.
- (16) Brown, R. H.; Al-Chalabi, A. Amyotrophic Lateral Sclerosis. *N Engl J Med* **2017**, *377* (2), 162–172. <https://doi.org/10.1056/NEJMra1603471>.
- (17) Zarei, S.; Carr, K.; Reiley, L.; Diaz, K.; Guerra, O.; Altamirano, P. F.; Pagani, W.; Lodin, D.; Orozco, G.; Chinae, A. A Comprehensive Review of Amyotrophic Lateral Sclerosis. *Surg Neurol Int* **2015**, *6*. <https://doi.org/10.4103/2152-7806.169561>.
- (18) *ALS Functional Rating Scale*. <https://www.outcomes-umassmed.org/ALS/alsscale.aspx> (accessed 2019-10-15).
- (19) Cedarbaum, J. M.; Stambler, N.; Malta, E.; Fuller, C.; Hilt, D.; Thurmond, B.; Nakanishi, A. The ALSFRS-R: A Revised ALS Functional Rating Scale That Incorporates Assessments of Respiratory Function. BDNF ALS Study Group (Phase III). *J. Neurol. Sci.* **1999**, *169* (1–2), 13–21. [https://doi.org/10.1016/s0022-510x\(99\)00210-5](https://doi.org/10.1016/s0022-510x(99)00210-5).
- (20) Turner, M. R.; Kiernan, M. C.; Leigh, P. N.; Talbot, K. Biomarkers in Amyotrophic Lateral Sclerosis. *The Lancet Neurology* **2009**, *8* (1), 94–109. [https://doi.org/10.1016/S1474-4422\(08\)70293-X](https://doi.org/10.1016/S1474-4422(08)70293-X).
- (21) Bereman, M. S.; Beri, J.; Enders, J. R.; Nash, T. Machine Learning Reveals Protein Signatures in CSF and Plasma Fluids of Clinical Value for ALS. *Sci Rep* **2018**, *8* (1), 1–14. <https://doi.org/10.1038/s41598-018-34642-x>.

- (22) Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Golar, O.; Sanchez, J.-C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L. From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. 5.
- (23) Wilkins, M. R.; Sanchez, J.-C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Williams, K. L. Progress with Proteome Projects: Why All Proteins Expressed by a Genome Should Be Identified and How To Do It. *Biotechnology and Genetic Engineering Reviews* **1996**, *13* (1), 19–50. <https://doi.org/10.1080/02648725.1996.10647923>.
- (24) Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humphery-Smith, I. Progress with Gene-Product Mapping of the Mollicutes: *Mycoplasma Genitalium*. *ELECTROPHORESIS* **1995**, *16* (1), 1090–1094. <https://doi.org/10.1002/elps.11501601185>.
- (25) Kenyon, G. L.; DeMarini, D. M.; Fuchs, E.; Galas, D. J.; Kirsch, J. F.; Leyh, T. S.; Moos, W. H.; Petsko, G. A.; Ringe, D.; Rubin, G. M.; Sheahan, L. C. Defining the Mandate of Proteomics in the Post-Genomics Era: Workshop Report: ©2002 National Academy of Sciences, Washington, D.C., USA. Reprinted with Permission from the National Academies Press for the National Academy of Sciences. All Rights Reserved. The Original Report May Be Viewed Online at [Http://Www.Nap.Edu/Catalog/10209.Html](http://www.nap.edu/catalog/10209.html). *Molecular & Cellular Proteomics* **2002**, *1* (10), 763–780.
- (26) Gstaiger, M.; Aebersold, R. Applying Mass Spectrometry-Based Proteomics to Genetics, Genomics and Network Biology. *Nat Rev Genet* **2009**, *10* (9), 617–627. <https://doi.org/10.1038/nrg2633>.

- (27) Omenn, G. S. PRESIDENTIAL ADDRESS: Grand Challenges and Great Opportunities in Science, Technology, and Public Policy. *Science* **2006**, *314* (5806), 1696–1704. <https://doi.org/10.1126/science.1135003>.
- (28) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; Maccoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschläger, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B. How Many Human Proteoforms Are There? *Nature Chemical Biology; Cambridge* **2018**, *14* (3), 206–214. <http://dx.doi.org/prox.lib.ncsu.edu/10.1038/nchembio.2576>.
- (29) Bludau, I.; Aebersold, R. Proteomic and Interactomic Insights into the Molecular Basis of Cell Functional Diversity. *Nat Rev Mol Cell Biol* **2020**, *21* (6), 327–340. <https://doi.org/10.1038/s41580-020-0231-2>.
- (30) International Human Genome Sequencing Consortium. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409* (6822), 860–921. <https://doi.org/10.1038/35057062>.
- (31) International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431* (7011), 931–945. <https://doi.org/10.1038/nature03001>.

- (32) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Miklos, G. L. G.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Francesco, V. D.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R.-R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z. Y.; Wang, A.; Wang, X.; Wang, J.; Wei, M.-H.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S. C.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri,

- V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y.-H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigó, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y.-H.; Coyne, M.; Dahlke, C.; Mays, A. D.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. The Sequence of the Human Genome. *Science* **2001**, *291* (5507), 1304–1351. <https://doi.org/10.1126/science.1058040>.
- (33) Yates, J. R. Mass Spectrometry and the Age of the Proteome. *Journal of Mass Spectrometry* **1998**, *33* (1), 1–19. [https://doi.org/10.1002/\(SICI\)1096-9888\(199801\)33:1<1::AID-JMS624>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1096-9888(199801)33:1<1::AID-JMS624>3.0.CO;2-9).
- (34) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394. <https://doi.org/10.1021/cr3003533>.

- (35) Macklin, A.; Khan, S.; Kislinger, T. Recent Advances in Mass Spectrometry Based Clinical Proteomics: Applications to Cancer Research. *Clin Proteom* **2020**, *17* (1), 17. <https://doi.org/10.1186/s12014-020-09283-w>.
- (36) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat Methods* **2009**, *6* (5), 359–362. <https://doi.org/10.1038/nmeth.1322>.
- (37) Sandra, K.; Moshir, M.; D'hondt, F.; Verleysen, K.; Kas, K.; Sandra, P. Highly Efficient Peptide Separations in Proteomics: Part 1. Unidimensional High Performance Liquid Chromatography. *Journal of Chromatography B* **2008**, *866* (1), 48–63. <https://doi.org/10.1016/j.jchromb.2007.10.034>.
- (38) Zhou, F.; Lu, Y.; Ficarro, S. B.; Webber, J. T.; Marto, J. A. Nanoflow Low Pressure High Peak Capacity Single Dimension LC-MS/MS Platform for High-Throughput, In-Depth Analysis of Mammalian Proteomes. *Anal. Chem.* **2012**, *84* (11), 5133–5139. <https://doi.org/10.1021/ac2031404>.
- (39) Šesták, J.; Moravcová, D.; Kahle, V. Instrument Platforms for Nano Liquid Chromatography. *Journal of Chromatography A* **2015**, *1421*, 2–17. <https://doi.org/10.1016/j.chroma.2015.07.090>.
- (40) Horváth, C.; Melander, W.; Molnár, I. Solvophobic Interactions in Liquid Chromatography with Nonpolar Stationary Phases. *Journal of Chromatography A* **1976**, *125* (1), 129–156. [https://doi.org/10.1016/S0021-9673\(00\)93816-0](https://doi.org/10.1016/S0021-9673(00)93816-0).
- (41) Hancock, W. S.; Bishop, C. A.; Prestidge, R. L.; Harding, D. R. K.; Hearn, M. T. W. Reversed-Phase, High-Pressure Liquid Chromatography of Peptides and Proteins with Ion-Pairing Reagents. *Science* **1978**, *200* (4346), 1168–1170.

- (42) Aguilar, M.-I.; Hearn, M. T. W. [1] High-Resolution Reversed-Phase High-Performance Liquid Chromatography of Peptides and Proteins. In *Methods in Enzymology*; High Resolution Separation and Analysis of Biological Macromolecules Part A: Fundamentals; Academic Press, 1996; Vol. 270, pp 3–26. [https://doi.org/10.1016/S0076-6879\(96\)70003-4](https://doi.org/10.1016/S0076-6879(96)70003-4).
- (43) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9. <https://doi.org/10.1021/ac302789c>.
- (44) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64–71.
- (45) Taylor, Sir. G. Disintegration of Water Drops in an Electric Field. *Proc. R. Soc. Lond. A* **1964**, *280* (1382), 383–397.
- (46) Rayleigh, Lord. XX. *On the Equilibrium of Liquid Conducting Masses Charged with Electricity. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1882**, *14* (87), 184–186. <https://doi.org/10.1080/14786448208628425>.
- (47) Iribarne, J. V. On the Evaporation of Small Ions from Charged Droplets. *J. Chem. Phys.* **1976**, *64* (6), 2287. <https://doi.org/10.1063/1.432536>.
- (48) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular Beams of Macroions. *The Journal of Chemical Physics* **1968**, *49* (5), 2240–2249. <https://doi.org/10.1063/1.1670391>.

- (49) Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. Automated Approach for Quantitative Analysis of Complex Peptide Mixtures from Tandem Mass Spectra. *Nat Methods* **2004**, *1* (1), 39–45. <https://doi.org/10.1038/nmeth705>.
- (50) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989. <https://doi.org/10.1021/jasms.8b00502>.
- (51) Tsou, C.-C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.-C.; Nesvizhskii, A. I. DIA-Umpire: Comprehensive Computational Framework for Data-Independent Acquisition Proteomics. *Nature Methods* **2015**, *12* (3), 258–264. <https://doi.org/10.1038/nmeth.3255>.
- (52) Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry. *Nat Commun* **2018**, *9* (1), 1–12. <https://doi.org/10.1038/s41467-018-07454-w>.
- (53) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nature Biotechnology* **2014**, *32* (3), 219–223. <https://doi.org/10.1038/nbt.2841>.
- (54) Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.; Smith, R. D.; Noble, W. S.; MacCoss, M. J. Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol Cell Proteomics* **2015**, *14* (9), 2301–2307. <https://doi.org/10.1074/mcp.O114.047035>.

- (55) Käll, L.; Vitek, O. Computational Mass Spectrometry–Based Proteomics. *PLoS Comput Biol* **2011**, *7* (12). <https://doi.org/10.1371/journal.pcbi.1002277>.
- (56) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *J. Proteome Res.* **2008**, *7* (1), 40–44. <https://doi.org/10.1021/pr700739d>.
- (57) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *J. Proteome Res.* **2008**, *7* (1), 29–34. <https://doi.org/10.1021/pr700600n>.
- (58) Rosenberger, G.; Bludau, I.; Schmitt, U.; Heusel, M.; Hunter, C. L.; Liu, Y.; MacCoss, M. J.; MacLean, B. X.; Nesvizhskii, A. I.; Pedrioli, P. G. A.; Reiter, L.; Röst, H. L.; Tate, S.; Ting, Y. S.; Collins, B. C.; Aebersold, R. Statistical Control of Peptide and Protein Error Rates in Large-Scale Targeted Data-Independent Acquisition Analyses. *Nat Methods* **2017**, *14* (9), 921–927. <https://doi.org/10.1038/nmeth.4398>.
- (59) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially Modified Protein Abundance Index (EmPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell Proteomics* **2005**, *4* (9), 1265–1272. <https://doi.org/10.1074/mcp.M500061-MCP200>.
- (60) Ong, S.-E.; Mann, M. Mass Spectrometry–Based Proteomics Turns Quantitative. *Nat Chem Biol* **2005**, *1* (5), 252–262. <https://doi.org/10.1038/nchembio736>.
- (61) Nahnsen, S.; Bielow, C.; Reinert, K.; Kohlbacher, O. Tools for Label-Free Peptide Quantification. *Mol Cell Proteomics* **2013**, *12* (3), 549–556. <https://doi.org/10.1074/mcp.R112.025163>.

- (62) *Cancer Statistics - National Cancer Institute*. <https://www.cancer.gov/about-cancer/understanding/statistics> (accessed 2020-11-02).
- (63) Mariotto, A. B.; Robin Yabroff, K.; Shao, Y.; Feuer, E. J.; Brown, M. L. Projections of the Cost of Cancer Care in the United States: 2010–2020. *J Natl Cancer Inst* **2011**, *103* (2), 117–128. <https://doi.org/10.1093/jnci/djq495>.
- (64) Balendiran, G. K.; Dabur, R.; Fraser, D. The Role of Glutathione in Cancer. *Cell Biochemistry and Function* **2004**, *22* (6), 343–352. <https://doi.org/10.1002/cbf.1149>.
- (65) Buchberger, A. R.; DeLaney, K.; Johnson, J.; Li, L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal. Chem.* **2018**, *90* (1), 240–265. <https://doi.org/10.1021/acs.analchem.7b04733>.
- (66) Karas, Michael.; Hillenkamp, Franz. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, *60* (20), 2299–2301. <https://doi.org/10.1021/ac00171a028>.
- (67) Caprioli, R. M.; Farmer, T. B.; Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.* **1997**, *69* (23), 4751–4760. <https://doi.org/10.1021/ac970888i>.
- (68) Laiko, V. V.; Baldwin, M. A.; Burlingame, A. L. Atmospheric Pressure Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Chem.* **2000**, *72* (4), 652–657. <https://doi.org/10.1021/ac990998k>.
- (69) Zenobi, R.; Knochenmuss, R. Ion Formation in MALDI Mass Spectrometry. *Mass Spectrometry Reviews* **1998**, *17* (5), 337–366. [https://doi.org/10.1002/\(SICI\)1098-2787\(1998\)17:5<337::AID-MAS2>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1098-2787(1998)17:5<337::AID-MAS2>3.0.CO;2-S).

- (70) Wang, B. H.; Dreisewerd, K.; Bahr, U.; Karas, M.; Hillenkamp, F. Gas-Phase Cationization and Protonation of Neutrals Generated by Matrix-Assisted Laser Desorption. *J. Am. Soc. Mass Spectrom.* **1993**, *4* (5), 393–398. [https://doi.org/10.1016/1044-0305\(93\)85004-H](https://doi.org/10.1016/1044-0305(93)85004-H).
- (71) Karas, M.; Glückmann, M.; Schäfer, J. Ionization in Matrix-Assisted Laser Desorption/Ionization: Singly Charged Molecular Ions Are the Lucky Survivors. *Journal of Mass Spectrometry* **2000**, *35* (1), 1–12. [https://doi.org/10.1002/\(SICI\)1096-9888\(200001\)35:1<1::AID-JMS904>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1096-9888(200001)35:1<1::AID-JMS904>3.0.CO;2-0).
- (72) Tu, A.; Muddiman, D. C. Internal Energy Deposition in Infrared Matrix-Assisted Laser Desorption Electrospray Ionization with and without the Use of Ice as a Matrix. *J Am Soc Mass Spectrom* **2019**, *30* (11), 2380–2391. <https://doi.org/10.1007/s13361-019-02323-2>.
- (73) Caleb Bagley, M.; Garrard, K. P.; Muddiman, D. C. The Development and Application of Matrix Assisted Laser Desorption Electrospray Ionization: The Teenage Years. *Mass Spectrom Rev* **2021**. <https://doi.org/10.1002/mas.21696>.
- (74) Robichaud, G.; Barry, J. A.; Muddiman, D. C. IR-MALDESI Mass Spectrometry Imaging of Biological Tissue Sections Using Ice as a Matrix. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (3), 319–328. <https://doi.org/10.1007/s13361-013-0787-6>.

## CHAPTER 2: Elucidation of Effective Data Acquisition Parameters for Shotgun

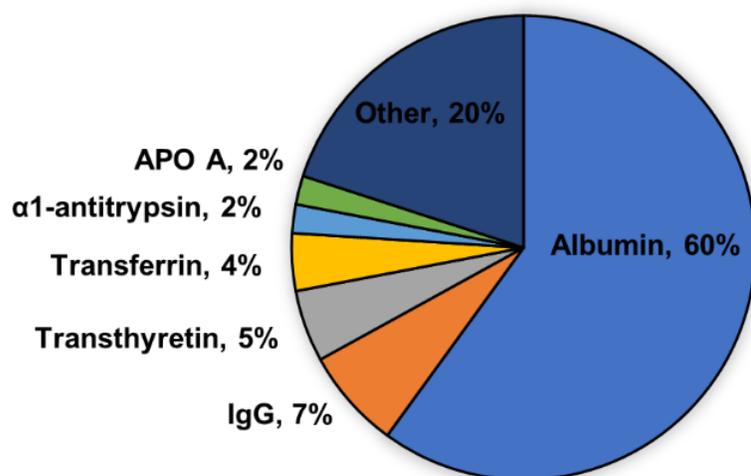
### Proteomics to Assay CSF

Reused with permission from: Mellinger, A.L., Griffith, E.H., Bereman, M.S. *Analytical and Bioanalytical Chemistry*, 2020, 412, 5465-5475. Copyright © 2020, Springer-Verlag GmbH

Germany, part of Springer Nature

#### 2.1 Introduction

Biofluids such as plasma or cerebrospinal are powerful in biomarker studies using clinical proteomics. However, the removal of highly abundant proteins (e.g., albumin) is often required as they dominate the total concentration of proteins. The relative fraction of total protein concentrations of these highly abundant proteins that make up CSF is shown in **Figure 2.1**.<sup>1</sup> Wide spanning dynamic ranges pose a challenge for detecting lower abundant proteins which tend to be important to answering biological questions at stake.<sup>2,3</sup> Depletion methods effectively mitigate this



**Figure 2.1.** Distribution of highly abundant proteins in human CSF. Albumin dominates a large portion of the total protein concentration (60%). Values contained in the “other” category make up a fraction of <1% of the total protein concentration. Adapted from Roche, et al. <sup>1</sup>

problem by removing the abundant proteins from the sample before further preparation, usually using affinity antibody resins for specific protein removal.<sup>4</sup> However, this method faces caveats that make it risky to utilize in clinical studies. These caveats include: 1)

increased expense for depletion kits 2) the large amount of starting material needed as antibodies available become quickly saturated with protein targets and 3) the risk of co-depleting proteins bound to these highly abundant proteins.<sup>5,6</sup> The large number of samples required for robust clinical studies, low starting volumes of precious biological samples, and the risk of co-depleting important proteins make depletion an unfeasible option. For this reason, we were interested in finding the best shotgun proteomics approach to survey the non-depleted CSF proteome.

While original studies noted improvements in the performance of DIA compared to DDA in complex mixtures (e. g., cells and tissues)<sup>7-9</sup>, more recent studies have utilized DIA for the analysis of biological fluids.<sup>10-12</sup> The requirement of a spectral library for appropriate sensitivity in DIA measurements can be a limitation to utilizing the method. However, Searle and coworkers showed that by generating an experimental chromatogram library, sensitive and reproducible measurements of HeLa cell peptides could be made with the use of a public spectral library at the cost of only a few extra injections in an experiment.<sup>7</sup> In this work, we sought to determine the best acquisition method (e.g., DDA vs. DIA) in terms of sensitivity, reproducibility, and quantitative precision for peptide measurements in non-depleted CSF. We also investigated the efficacy of using a publicly available organism-specific spectral library vs a sample specific CSF spectral library built from data contained in repositories.<sup>13</sup>

## **2.2 Materials and Methods**

### *2.2.1 Materials*

Ammonium bicarbonate (ABC), sodium deoxycholate (SDC), urea, dithiothreitol (DTT), iodoacetamide (IAM) and formic acid were all purchased from Sigma Aldrich (St. Louis, MO). Hydrochloric acid was purchased from Fisher Scientific (Pittsburgh, PA) and 0.01% acetic acid was purchased from RICCA chemical company (Arlington, TX). A Pierce BCA Protein Assay

Kit, Vivacon500 30 kDa MWCO spin filters, and Pierce HeLa protein digest and BSA protein digest standards were purchased from ThermoFisher Scientific (Waltham, MA). HPLC grade acetonitrile and water were purchased from Burdick & Jackson (Muskegon, MI) and high purity nitrogen gas from Machine & Welding Supply (Raleigh, NC). Sequencing grade porcine trypsin was obtained from Promega (Madison, WI) and pooled human CSF from BioIVT (Westbury, NY). ReproSil-Pur 120 C-18 AQ 3  $\mu\text{m}$  stationary phase was purchased from Dr. Maisch (Ammerbuch-Entringen, Germany) to pack Pico-Frit columns from New Objective (Woburn, MA).

### *2.2.2 Sample Preparation*

All CSF samples were prepared for shotgun proteomics via a modified filter-aided sample preparation method.<sup>14</sup> Prior to digestion, protein concentrations were measured using a BCA assay. Protein concentrations were normalized by addition of appropriate amounts of 1% SDC solution in 50 mM ABC. Disulfide bonds were reduced via incubation at 60 °C for 30 minutes with 5 mM DTT. Free cysteines were alkylated with 15 mM IAM in the dark for 20 mins. 30 kDa MWCO filters were prepared with the addition of 20  $\mu\text{L}$  of 50 mM ABC and 1% SDC solution for passivation. Protein samples were added to spin filters and centrifuged at  $12,000 \times g$  for 15 minutes. The samples were then washed twice with 200  $\mu\text{L}$  50 mM ABC and 8 M urea solution and twice with 200  $\mu\text{L}$  50 mM ABC. The addition of every wash volume was followed by a 15-minute centrifugation of the samples at  $12,000 \times g$ .

Trypsin was reconstituted to a concentration of 0.1  $\mu\text{g}/\mu\text{L}$  in 0.01% acetic acid and added to each sample at a 1:50 trypsin to protein ratio. Samples were shaken at 600 rpm at 37 °C for four hours. Following digestion, the samples were spun for 15 minutes at  $12,000 \times g$ . They were then washed with a small amount of ABC solution and spun for a final 15 minutes at  $12,000 \times g$ .

Digestion was quenched with addition of HCl to a final concentration of 250 mM. Samples were stored at -80° C until analysis.

### 2.2.3 LC-MS/MS

500 ng of pooled CSF digest were loaded onto a self-packed 2 cm trapping column prior to separation on a 20 cm self-packed reversed-phase analytical column. A 90-minute gradient was used, ramping from 100% mobile phase A (2% ACN in water + 0.1% formic acid) to 40% mobile phase B (80% ACN in water + 0.1% formic acid). This was followed by a 10-minute 100% mobile phase B wash and a 10-minute column re-equilibration to 100% mobile phase A. Eluting peptides were ionized by electrospray and mass analyzed by a hybrid quadrupole-orbitrap mass spectrometer (QE-HF, ThermoFisher Scientific, Bremen, Germany). For comparison of data acquisition methods, 4 replicates of the pooled CSF sample were randomly acquired by both data dependent acquisition DDA and DIA.

For DDA, a “Top 20” method was used in which, each cycle consisted of an MS1 scan from 400 to 1400  $m/z$  at a resolving power of 120,000 at 200  $m/z$ , automatic gain control (AGC) target of  $1e6$ , and maximum injection time (IT) of 30 ms. MS2 scans were acquired with resolving power of 15,000 at  $m/z$  200, an isolation window of 2.0  $m/z$ , 60 ms maximum IT,  $1e5$  AGC target, and normalized collision energy (NCE) of 27. For data dependent settings, the top twenty most abundant multiply charged species were selected for sequencing with an exclusion list of 20s to avoid repeated interrogation of abundant species.

Four replicate injections were analyzed via a DIA “wide-window” method. Each cycle in this method consisted of an MS1 scan followed by 24 MS2 scans covering 400-1000  $m/z$  space, and fragmenting precursors within a 25  $m/z$  isolation window in each scan. MS1 scans of the 400-1000  $m/z$  precursor space were acquired using a resolving power of 30,000 at 200  $m/z$  with a

maximum IT of 55 ms and an AGC target of 3e6. MS2 scans were acquired using a resolving power of 30,000 at 200  $m/z$  with an AGC target of 1e6 and a maximum IT of 55 ms. A NCE of 27 was used and all data was collected in centroid mode.<sup>7</sup> All raw data is publicly available on Panorama. This experiment can be found on the Panorama repository on the Bereman Lab – NCSU project page under “Longitudinal Biomarkers in ALS CSF”, specifically under the DIA vs DDA folder on the Raw Data page.

#### 2.2.4 Chromatogram Library Generation

In efforts to create the most sensitive library for peptide quantification, gas-phase fractionated DIA-MS was performed in the middle of data collection for each cycle. Six different injections of a pooled CSF sample were analyzed with each run covering 100  $m/z$  from 400 to 1000  $m/z$ . Each gas-phase fractionated method consisted of an overlapping scheme as described by Amodei and coworkers.<sup>15</sup> MS1 scans were acquired once per cycle with settings noted *vide supra*. MS2 scans of 4.0  $m/z$  isolation width were overlapped between cycles, resulting in 2.0  $m/z$  windows after demultiplexing. A resolving power of 30,000 at 200  $m/z$  was used with an AGC target of 1e6 and a 55 ms maximum IT. An NCE of 27 was also used in these methods.

#### 2.2.5 Spectral Library Generation

RAW files generated both in-house from previous experiments performed in the lab<sup>16</sup> and downloaded from Proteome Xchange were used to create a CSF specific spectral library. RAW files were chosen from studies of CSF performed on Q Exactive instruments. Descriptions of these RAW files and their studies are provided in **Table A.1**. The RAW files were first analyzed using Proteome Discoverer 1.4. All RAW files were searched against the Uniprot Human FASTA database (modified March 16, 2019 containing 73,911 entries).

Result files were then imported into Skyline (version 19.0.9.190) to create a DLIB spectral library containing 43,716 peptides. This spectral library is publicly available on the Panorama repository on the Bereman Lab- NCSU Raw Data dashboard of the “Longitudinal Biomarkers in ALS CSF” project.

#### *2.2.6 DIA Database Searching*

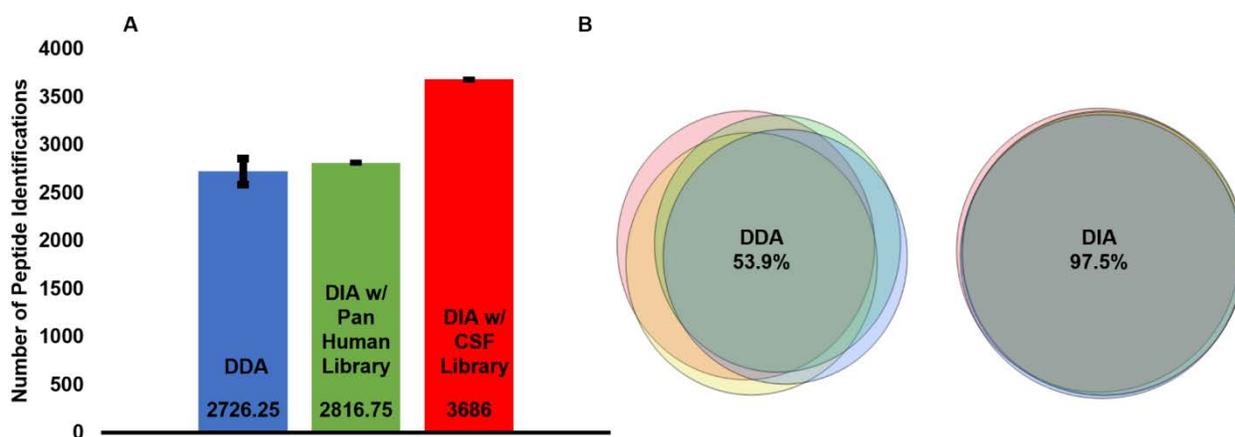
RAW files were converted to mzML format and DIA files were demultiplexed using the MSConvert tool of the ProteoWizard package (Version 3.011417.) Narrow-window mzML files were used to generate an ELIB format chromatogram library in EncyclopeDIA (Version 0.8.2)<sup>7</sup> by searching them against a DLIB format spectral library. Results using the Pan-Human spectral library<sup>17</sup> and the CSF specific library (see previous) were compared. The CSF specific library was used to generate the longitudinal study chromatogram library. EncyclopeDIA was configured to searching the DIA RAW files with default settings to search for tryptic peptides and fragments.<sup>7</sup> For initial comparison between DIA and DDA, data contained in the DIA “quant reports” were used.

#### *2.2.7 DDA Database Searching*

RAW files were batch searched using the SEQUEST<sup>18</sup> algorithm within Proteome Discoverer 2.2 software. Tryptic peptides at least 6 amino acids in length were searched for with a precursor mass tolerance of 5 ppm. Fragment b and y ions were searched with a mass tolerance of 0.02 Da. Methionine oxidation was searched as a variable modification while cysteine carbamidomethylation was searched as a static modification, and up to two missed cleavages were allowed. The data was then filtered using the Percolator<sup>19</sup> algorithm to increase sensitivity and enforce FDR of 1%.

## 2.3 Results and Discussion

While original studies noted improvements in the performance of DIA compared to DDA in complex mixtures (e. g., cells and tissues)<sup>7-9</sup>, more recent studies have utilized DIA for the analysis of biological fluids<sup>10-12</sup>. Thus prior to analysis of these precious biological samples, an experiment was performed to determine the best acquisition method (e.g., DDA vs. DIA) in terms of sensitivity, reproducibility, and quantitative precision for peptide measurements in CSF. In agreement with previous studies<sup>7,11</sup>, we found across eight replicate injections of pooled CSF digest that in general DIA (n=4) outperformed conventional DDA (n=4) across all metrics

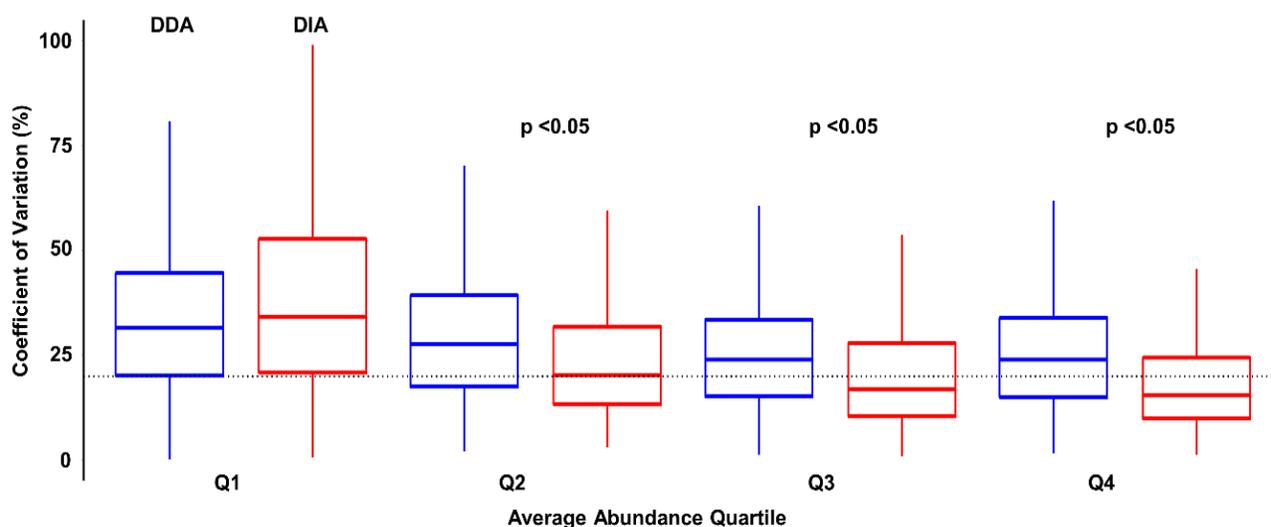


**Figure 2.2.** Comparison of DIA and DDA for CSF shotgun proteomics. **A)** Bar graph of average number and standard deviation of identified peptides in four replicates of each method. DDA (blue), DIA with a Pan-Human spectral library (green), and DIA with a CSF specific spectral library (red). **B)** Proportional Venn Diagrams comparing the number of peptide IDs shared amongst four DDA and four DIA replicates.

investigated. DIA achieved 35% more peptide identifications than DDA, with an average of 3686 peptides using the experiment specific chromatogram library generated from a CSF specific spectral library. On average, 2726 peptide were identified by DDA “Top 20” and 2817 identifications were made with DIA using the Pan-Human spectral library (**Figure 2.2A**).<sup>17</sup>

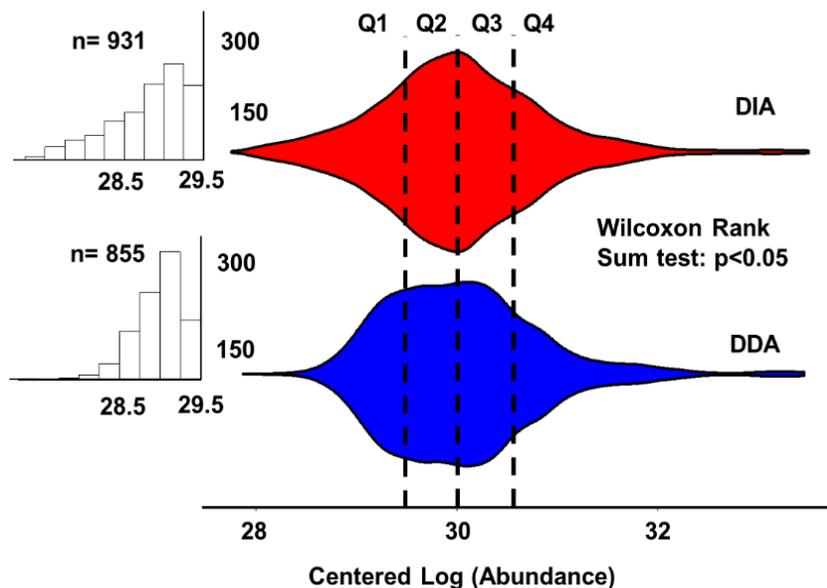
The consistency of the  $m/z$  isolation windows in DIA, compared to the stochastic sampling of DDA, provides higher reproducibility in peptide identification. Only 54% of the peptides identified via DDA were shared in all four replicates, whereas nearly all peptides (98%) were

identified across the DIA procedure (**Figure 2.2B**). In addition, we investigated the precision of quantitation as a function of dynamic range for both acquisition methods. DIA achieved a higher precision of quantitation ( $p < 0.05$ ) across 3 quartiles of peptide abundances (**Figure 2.3**). Interestingly, DDA out-performed DIA in the quantitation of peptides across the lowest quartile of abundance.



**Figure 2.3.** Boxplots of peptide CVs across replicates abundance quartile. A dotted line marks 20% CV. DIA outperformed DDA in terms of quantitative precision with lower and smaller distributions of % CV, except for in the lowest quartile comparisons.

We investigated this observation further by median centering both distributions to 30. As shown in the violin plots of these data (**Figure 2.4**), DIA identified a larger number of peptides at lower abundance. A Wilcoxon Rank Sum test rejected the null that the distributions are the same, adding statistical significance evidence to support this finding. Thus, we conclude that due to the increased dynamic range and the difficulties associated with quantitation of low abundant signals, increased noise, and a low number of points across peptide elution profiles are possible reasons for this result.



**Figure 2.4.** Violin plots comparing distributions of median centered peptide abundances. A Wilcoxon ranked sum test confirmed the difference in the distribution of abundances in the lowest quartile. Histograms of the lowest quartile for each distribution is inset for further visualization of the distributions.

## 2.4 Conclusions

We found that the DIA method outperformed DDA in terms of peptide identifications, reproducibility of those identifications, and quantitative precision in CSF. These findings are important to note for future clinical proteomic analyses of the biofluid, including those detailed in the following Chapter.

## 2.5 Acknowledgements

The authors gratefully acknowledge the ALS Association (grant #19-SI-458) for the funding of this project. All mass spectrometry measurements were made in the Molecular Education, Technology, and Research Innovation Center (METRIC) at NC State University.

## 2.6 References

- (1) Roche, S.; Gabelle, A.; Lehmann, S. Clinical Proteomics of the Cerebrospinal Fluid: Towards the Discovery of New Biomarkers. *PROTEOMICS – Clinical Applications* **2008**, *2* (3), 428–436. <https://doi.org/10.1002/prca.200780040>.
- (2) Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol Cell Proteomics* **2002**, *1* (11), 845–867. <https://doi.org/10.1074/mcp.R200007-MCP200>.
- (3) Spreafico, F.; Bongarzone, I.; Pizzamiglio, S.; Magni, R.; Taverna, E.; De Bortoli, M.; Ciniselli, C. M.; Barzanò, E.; Biassoni, V.; Luchini, A.; Liotta, L. A.; Zhou, W.; Signore, M.; Verderio, P.; Massimino, M. Proteomic Analysis of Cerebrospinal Fluid from Children with Central Nervous System Tumors Identifies Candidate Proteins Relating to Tumor Metastatic Spread. *Oncotarget* **2017**, *8* (28), 46177–46190. <https://doi.org/10.18632/oncotarget.17579>.
- (4) Steel, L. F.; Trotter, M. G.; Nakajima, P. B.; Mattu, T. S.; Gonye, G.; Block, T. Efficient and Specific Removal of Albumin from Human Serum Samples. *Molecular & Cellular Proteomics* **2003**, *2* (4), 262–270. <https://doi.org/10.1074/mcp.M300026-MCP200>.
- (5) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394. <https://doi.org/10.1021/cr3003533>.
- (6) Bellei, E.; Bergamini, S.; Monari, E.; Fantoni, L. I.; Cuoghi, A.; Ozben, T.; Tomasi, A. High-Abundance Proteins Depletion for Serum Proteomic Analysis: Concomitant Removal of Non-Targeted Proteins. *Amino Acids* **2011**, *40* (1), 145–156. <https://doi.org/10.1007/s00726-010-0628-x>.

- (7) Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry. *Nat Commun* **2018**, *9* (1), 1–12. <https://doi.org/10.1038/s41467-018-07454-w>.
- (8) Canterbury, J. D.; Merrihew, G. E.; MacCoss, M. J.; Goodlett, D. R.; Shaffer, S. A. Comparison of Data Acquisition Strategies on Quadrupole Ion Trap Instrumentation for Shotgun Proteomics. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (12), 2048–2059. <https://doi.org/10.1007/s13361-014-0981-1>.
- (9) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Xuan, Y.; Sondermann, J.; Schmidt, M.; Gomez-Varela, D.; Reiter, L. Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol Cell Proteomics* **2017**, *16* (12), 2296–2309. <https://doi.org/10.1074/mcp.RA117.000314>.
- (10) Lin, L.; Zheng, J.; Zheng, F.; Cai, Z.; Yu, Q. Advancing Serum Peptidomic Profiling by Data-Independent Acquisition for Clear-Cell Renal Cell Carcinoma Detection and Biomarker Discovery. *Journal of Proteomics* **2020**, *215*, 103671. <https://doi.org/10.1016/j.jprot.2020.103671>.
- (11) Barkovits, K.; Linden, A.; Galozzi, S.; Schilde, L.; Pacharra, S.; Mollenhauer, B.; Stoepel, N.; Steinbach, S.; May, C.; Uszkoreit, J.; Eisenacher, M.; Marcus, K. Characterization of Cerebrospinal Fluid via Data-Independent Acquisition Mass Spectrometry. *J. Proteome Res.* **2018**, *17* (10), 3418–3430. <https://doi.org/10.1021/acs.jproteome.8b00308>.
- (12) Muntel, J.; Xuan, Y.; Berger, S. T.; Reiter, L.; Bachur, R.; Kentsis, A.; Steen, H. Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a

- Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **2015**, *14* (11), 4752–4762.  
<https://doi.org/10.1021/acs.jproteome.5b00826>.
- (13) Barkovits, K.; Linden, A.; Galozzi, S.; Schilde, L.; Pacharra, S.; Mollenhauer, B.; Stoepel, N.; Steinbach, S.; May, C.; Uszkoreit, J.; Eisenacher, M.; Marcus, K. Characterization of Cerebrospinal Fluid via Data-Independent Acquisition Mass Spectrometry. *J. Proteome Res.* **2018**, *17* (10), 3418–3430. <https://doi.org/10.1021/acs.jproteome.8b00308>.
- (14) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat Methods* **2009**, *6* (5), 359–362.  
<https://doi.org/10.1038/nmeth.1322>.
- (15) Amodei, D.; Egertson, J.; MacLean, B. X.; Johnson, R.; Merrihew, G. E.; Keller, A.; Marsh, D.; Vitek, O.; Mallick, P.; MacCoss, M. J. Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (4), 669–684. <https://doi.org/10.1007/s13361-018-2122-8>.
- (16) Bereman, M. S.; Beri, J.; Enders, J. R.; Nash, T. Machine Learning Reveals Protein Signatures in CSF and Plasma Fluids of Clinical Value for ALS. *Sci Rep* **2018**, *8* (1), 1–14. <https://doi.org/10.1038/s41598-018-34642-x>.
- (17) PXD000954 - The Pan-Human Library: A repository of assays to quantify 10 000 proteins by SWATH-MS / SWATH-MS validation data - OmicsDI  
<https://www.omicsdi.org/dataset/pride/PXD000954> (accessed 2020 -02 -04).
- (18) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989. <https://doi.org/10.1021/jasms.8b00502>.

- (19) Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-Supervised Learning for Peptide Identification from Shotgun Proteomics Datasets. *Nature Methods* **2007**, *4* (11), 923–925. <https://doi.org/10.1038/nmeth1113>.

## **CHAPTER 3: Analysis of the Proteome of CSF Derived from a Longitudinal ALS Cohort to Understand Disease Progression**

Reused with permission from: Mellinger, A.L., Griffith, E.H., Bereman, M.S. *Analytical and Bioanalytical Chemistry*, 2020, 412, 5465-5475. Copyright © 2020, Springer-Verlag GmbH Germany, part of Springer Nature

### **3.1 Introduction**

Single timepoint studies are a straightforward and robust way of studying diseased populations, they are typically unable to resolve disease defining biochemistry from confounding variables within individuals. However, the biochemical influences of factors such as stress, diet, exercise, circadian rhythm, or even drug usage can be accounted for when the same patient is sampled repeatedly.<sup>1,2</sup> In a heterogenous disease such as ALS, this type of experimental design is critical for both identification and validation of biomarkers. Herein, we use shotgun proteomics and DIA, based on results presented in Specific Aim I, to identify peptide biomarkers in CSF longitudinally collected from ALS patients. We first investigate the technical, inter-, and intra-variability of peptide measurements made in this experiment and their contribution to the overall variance. We also examine the role of global inflammation in the progression of the disease. Finally, we develop a repeated measures linear model using peptides that minimize Akaike's Information Criterion (AIC), a commonly used metric for model fit assessment. Using least absolute shrinkage and selection operator (LASSO) for variable selection. This regression analysis method provides interpretable models by minimizing the residual sum of squares subject to the sum of the absolute value of the variable coefficients being less than a specified constant. This constraint produces a subset of coefficients as 0, which means they can be removed for a simpler model. This selection method is amenable to a variety of statistical models, such as the linear

regression used here.<sup>3</sup> The final model created from selected variables was a mixed-effects longitudinal model that serves as a quantitative model of disease progression.

## 3.2 Materials and Methods

### 3.2.1 LC-MS/MS

De-identified longitudinally collected CSF samples from 15 ALS patients at 3-7 time points per patient (63 total samples) were obtained from the Northeastern Amyotrophic Lateral Sclerosis Consortium (NEALS) biorepository. Samples were prepared and analyzed following the agreement between the Bereman Laboratory and NEALS which focused on identification of protein biomarkers CSF. Smoking status of patients was unknown.

Samples were digested in three cycles, the first two containing 23 ALS samples and the final cycle containing 17. Patient were blocked by day in efforts to minimize variability associated with sample preparation within each patient (**Table 3.1**). A pooled CSF sample containing both healthy and other neurological diseases was digested (as described previously in 2.2.2) with each cycle as a digestion quality control. Aliquots of each of the 63 biological samples were pooled together for chromatogram library generation.<sup>4</sup> Digested samples were analyzed randomly by

**Table 3.1.** Description of ALS patient cohort and digestion cycles.

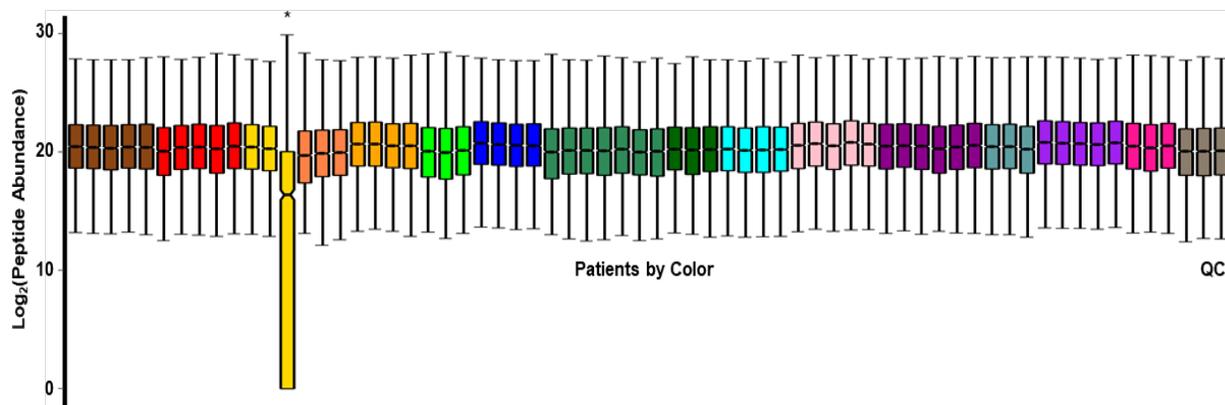
Patient	Digestion Cycle	Visits	Gender	Starting Age	Ending Age	Starting ALSFRS	Ending ALSFRS	Difference ALSFRS	Onset	Years from Diagnosis (Initial Visit)	Years from Diagnosis (Final Visit)	Years Progressed
1	1	5	F	65.3	67.0	45	36	9	Bulbar	1.5	3.2	1.7
2		5	F	70.8	72.1	36	26	10	Limb	3.1	4.3	1.3
3		3	F	40.7	43.8	43	36	7	Limb	1.6	4.6	3.0
4		3	F	50.1	51.0	41	38	3	Limb	1.0	1.9	0.9
5		4	M	47.5	49.0	44	35	9	Limb	1.6	3.1	1.5
6		3	M	56.8	57.4	41	42	-1	Limb	2.5	3.1	0.6
7	2	4	F	61.6	62.8	31	25	6	Bulbar	1.7	2.8	1.2
8		7	M	64.8	66.9	41	36	5	Limb	2.0	4.1	2.2
9		6	M	53.6	56.0	31	23	8	Limb	9.4	11.8	2.4
10		4	M	56.8	58.0	45	45	0	Limb	1.5	2.6	1.1
11		3	M	66.4	67.4	22	25	-3	Limb	9.6	10.6	1.0
12	3	5	F	58.9	60.9	42	19	23	Bulbar	1.4	3.4	2.0
13		3	F	68.1	68.8	45	42	3	Limb	1.5	2.2	0.7
14		6	M	69.0	70.9	38	22	16	Limb	5.1	6.9	1.8
15		3	M	47.0	49.0	45	29	16	Limb	2.4	4.4	2.0

DIA-MS within each cycle. 500 ng of each digested sample was loaded for separation via reversed phase nanoflow liquid chromatography. Peptides were loaded onto a 2 cm trapping column (Acclaim PepMap 100 C18 HPLC, 0.075 mm diameter with 3  $\mu\text{m}$  particle size, ThermoFisher Scientific) prior to separation on a 25 cm reversed phase analytical column (EASY-Spray HPLC, 0.075 mm diameter with 3  $\mu\text{m}$  particle size, ThermoFisher Scientific). A 90-minute gradient was used as described in 2.2.3 within **Chapter 2**. Eluting peptides were ionized by electrospray and mass analyzed by a hybrid quadrupole-orbitrap mass spectrometer (QE-HFX, ThermoFisher Scientific, Bremen, Germany).

Samples were analyzed using DIA-MS with a set isolation window of 12  $m/z$ . Each cycle consisted of an MS1 scan followed by 50 MS2 scans covering 400-1000  $m/z$ , and fragmenting precursors within a 12  $m/z$  isolation window in each scan. Precursor MS1 scans of 400-1000  $m/z$  were acquired using a resolving power of 60,000 at 200  $m/z$  with a maximum IT of 55 ms and an AGC target of  $3e6$ . MS2 scans were acquired using a resolving power of 15,000 at 200  $m/z$  with an AGC target of  $1e6$  and a maximum IT of 25 ms. A NCE of 27 was used and all data was collected in centroid mode. All raw data is publicly available on Panorama. This experiment can be found on the Bereman Lab- NCSU project page under “Longitudinal Biomarkers in ALS CSF”, specifically under the Longitudinal ALS CSF folder on the Raw Data page.

### 3.2.2 *Quality Control*

**Figure 3.1** shows the distribution of logged peptide abundances for every sample. One time point was a clear outlier based on these data and samples from this patient were removed from further analysis. Aside from this single outlier we did not observe any qualitative bias between digestion cycles, as evidenced by the similar distributions of abundances across samples and by the lower variability among the QC samples (median CV=16.3%).



**Figure 3.1.** Boxplot of the  $\log_2$  peptide abundances of each sample. Each patient is indicated by a new color. One outlier was discovered. The reason for the low abundances is unknown, but could be attributed to poor digestion efficiency, a bad injection, or human error in the preparation process. This patient was excluded from subsequent analyses.

### 3.2.3 Skyline

Strict filtering of peptides found in the longitudinal data set was performed in Skyline (version 20.1.0.31). Quantitative wide window mzML files were searched against the library generated in EncyclopeDIA. Peptides with  $>10$  ppm mass error,  $< 3$  coeluting transitions, and unstable retention times, were filtered from the analysis data set. The Skyline files containing the full and filtered data are available on Panorama under the Panorama Dashboard of the Bereman Lab- NCSU, specifically within the “Longitudinal Biomarkers in ALS CSF” project.

### 3.2.4 Multivariate Analysis

Peptide abundances, after manual inspection in Skyline, were imported into Rstudio (Version 3.6.1)<sup>5</sup> for further analysis. Packages including nlme, caret, ggplot2, MASS, glmnet, tidyr, and dplyr were utilized. All statistical analyses were performed after  $\log_2$  transformation of the abundances. Highly correlated peptides were removed in efforts to limit the error in predictions and maximize the interpretability of the mathematical model. This was done using a filtering function that compared highly correlated peptides (absolute value pair-wise correlation coefficient

> 0.6) and removed the peptide with the largest mean absolute correlation. This process left 96 peptides for model selection. A LASSO regression was utilized and cross-validated to select 55 peptides that explained 99% of the deviation in a linear model of ALS-FRS score. Any peptides shared by the same proteins were filtered, with the single best fitting peptide selected based on a lower model AIC and p-value. Using forward selection to minimize AIC for overall model performance<sup>6</sup>, the remaining 44 peptides were investigated as fixed effects to model ALS-FRS score, along with the years from disease onset and the site of disease onset. A mixed-effects model was generated that accounted for each patient with a random intercept. R script detailing this procedure is available within **Appendix B**.

The best-fitting model contained 28 peptides as fixed effects along with time and disease onset site. Diagnostic plots were used to assess model fit. A negative control model was tested by permuting the peak areas within the data set. The new random variables were added to a model in a similar forward selective manner. The proteins represented by each peptide were investigated for possible biological significance using Ingenuity Pathway Analysis (IPA) (Version 49932394), StringDB (Version 11.0), and Cytoscape (Version 3.7.2).

### *3.2.5 Global Inflammation Study*

In order to investigate the change in global inflammation within patients, 2-3 peptides from 17 proteins associated with inflammatory response (DAVID, Version 6.8) were filtered through Skyline (file available on Panorama with Skyline files mentioned previously) and then imported into Rstudio (filtering described above.) The inflammatory proteins and representative peptides used are listed in **Table B.1**. To calculate an inflammatory index, the median peak area of these peptides was taken for each sample. Each patient's first visit was defined as the baseline. The fold change of the inflammatory index from baseline was calculated at each time point and log<sub>2</sub>

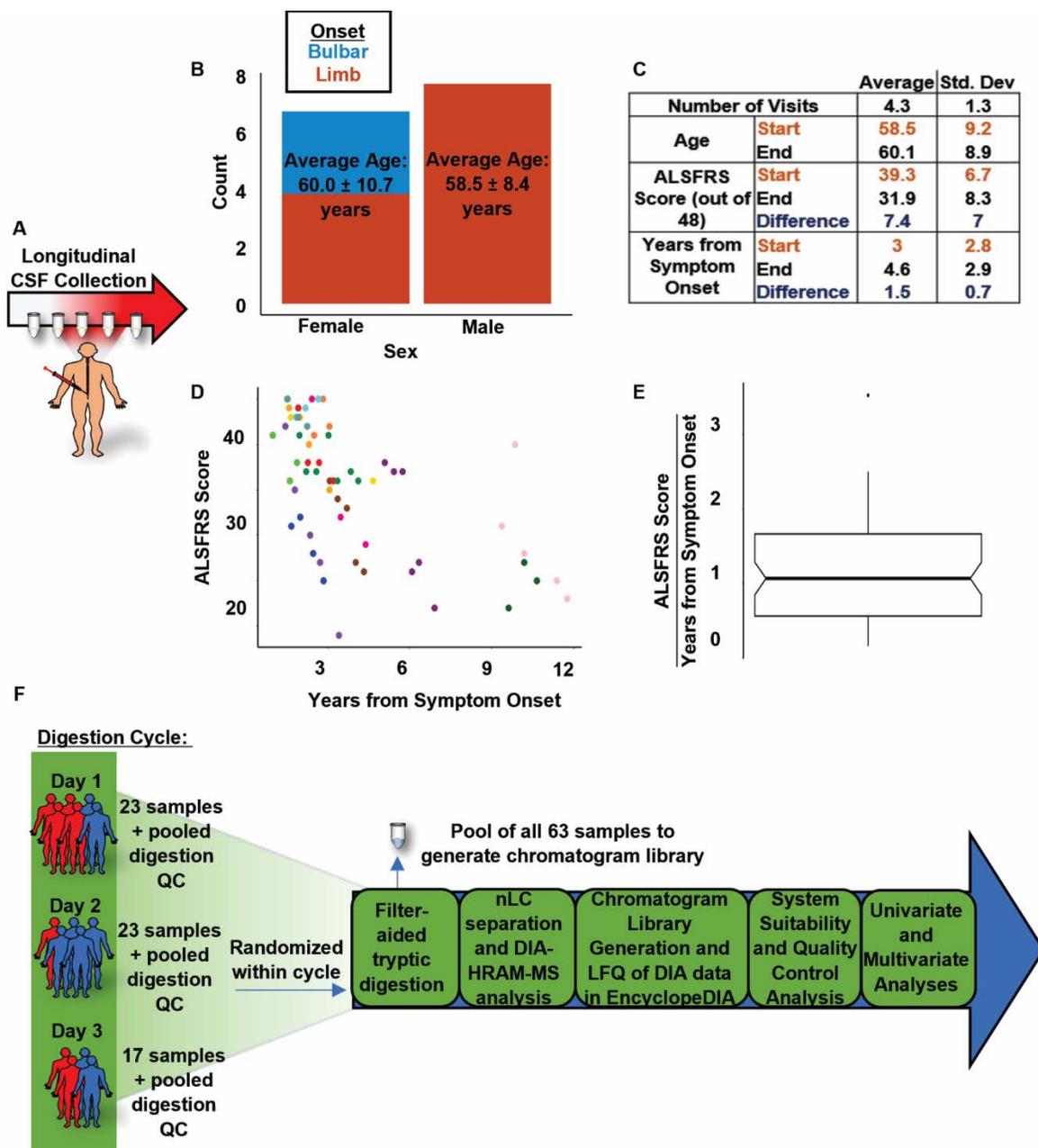
transformed. The rate of disease progression was calculated by dividing the difference in FRS by the change in months between each visit. For each patient, we plotted the rate of disease progression against inflammatory index fold change and investigated the linear regression between these two metrics<sup>7</sup>.

### 3.3 Results and Discussion

The patient cohort analyzed in this study is summarized in **Figure 3.2**. The sample set consisted of male and female along with bulbar and limb onset patients. For these 15 patients, 3-7 samples per patient were collected over the course of a couple years (1.5 +/- 0.7 years on average, approximately 3-5 years from diagnosis of the disease) with a few months in between collection times. The average age of patients at their first visit was 58.5 years (+/- 9.2 years) (**Figure 3.2B/C**). ALS-FRS scores ranged from 19-45 out of the 48-point scale, with an average decrease of 7.4 points among patients from first to last visit (**Figure 3.2D**). The ratio of ALS-FRS score to months from symptom onset was used to visualize disease progression<sup>8</sup>. This cohort had a median progression rate of 0.256 points/month (interquartile range of 0.192 points/month) (**Figure 3.2E**). Samples were digested in 3 cycles (1 cycle per day) and randomized before sequencing for analysis (**Figure 3.2F**).

#### 3.3.1 Peptide Variability in ALS

Several studies have investigated the intra-individual protein variability in plasma;<sup>9-11</sup> However, little is known about the protein dynamics in CSF especially in neurodegenerative diseases. CSF not only serves as a protective fluid for the central nervous system, but also functions as a bridge between the central nervous system and other biological symptoms in the body. Therefore, most proteins in CSF are produced in blood plasma, and diffuse into CSF through the



**Figure 3.2.** Summary of patient cohort analyzed. **A)** CSF samples were collected via lumbar puncture over the course of patients' disease. **B)** Bar-graph describing distribution of sex, disease onset, and age by sex. **C)** Averages of factors describing disease progression of patients. **D)** Scatterplot of ALSFRS scores over the years from symptom onset to collection date. Color by patient. **E)** Distribution of "progression index" describing disease progression at each visit. Progression index is defined as the ratio of the difference of 48 and the patient's measured score at collection to the months from symptom onset at that measured point. **F)** Summary of experimental design. Figure was partially created using images purchased in the PPT Drawing Toolkits-BIOLOGY Bundle from Motifolio, Inc.

blood-brain barrier at the choroid plexus.<sup>12</sup> Approximately 20% of the proteins in CSF are specifically brain derived.<sup>13,14</sup> In our longitudinal study of CSF, the reproducibility of peptide identification on average within patients was 96.5%, where peptides with three or more fragment ions were detected in at least two samples from an individual patient. **Figure 3.3A** displays the variability (%CV) of peptide abundances across the sample set. The total median variability of peptide abundances quantified across the sample set was 37 %CV with an interquartile range (IQR) of 26%. The median of the within patient peptide abundance variabilities was 18 %CV while the technical variability from the digestion replicates was 16 %CV (**Figure 3.3B**). We can use these medians to estimate the contribution of biological and technical variability to the total experimental variability.<sup>15</sup>

$$(3.1) CV^2_{Intra BioV \& TechV} = CV^2_{Intra BioV} + CV^2_{TechV}$$

$$17.8^2 = CV^2_{Intra BioV} + 16.2^2$$

$$CV^2_{Intra BioV} = 316.8 - 262.4 = 54.4$$

$$CV_{Intra BioV} = 7.4$$

$$(3.2) CV^2_{Total} = CV^2_{Inter BioV} + CV^2_{Intra BioV} + CV^2_{TechV}$$

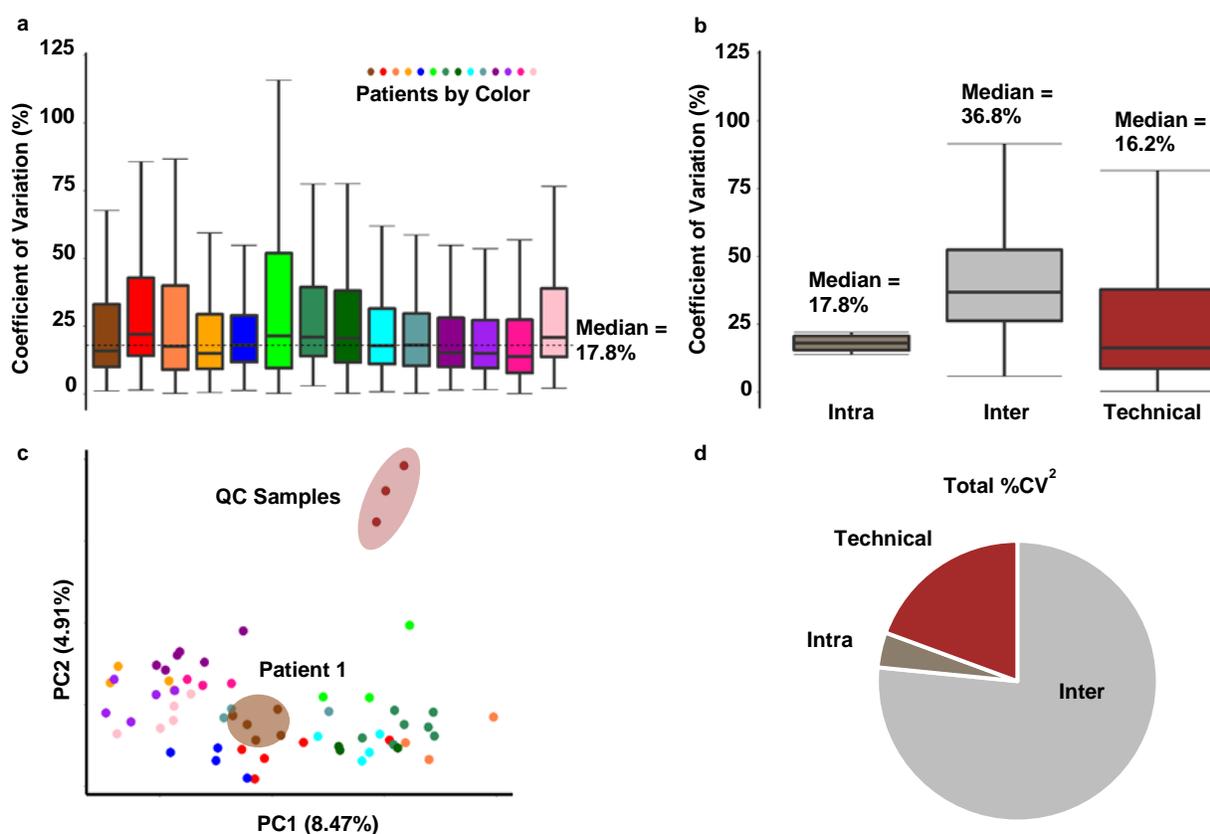
$$36.8^2 = CV^2_{Inter BioV} + 7.4^2 + 16.2^2$$

$$CV^2_{Inter BioV} = 1354.2 - 54.4 - 262.4 = 1037.0$$

$$CV_{Inter BioV} = 32.2$$

First, we solve **Equation 3.1** to estimate exclusively the contribution of the biochemical intra-variability ( $CV_{Intra BioV}$ , i.e., without technical variability) to measurements within patient time points. The estimated contribution of the within peptide biological variability is 7.4 %CV which suggests that the CSF proteome is tightly regulated within an individual. This high reproducibility is further exemplified in PCA space as individuals qualitatively show clustering closer than

distances between individuals (**Figure 3.3C**). In solving **Equation 3.1**, we find that most variability in peptide abundances is attributed to between individual variability (32.2 %CV). Consequently, the overall estimated contributions to variance of the experiment ( $CV^2_{Total}$  in **Equation 3.2**) were from inter-biological variance (76.6%) > technical variance (19.4%) > intra-biological variance (4.0%) (**Figure 3.3D**). These data agree with a small two time-point study of six healthy patients that the between individual variance is much greater than within individual variance.<sup>16</sup> These data will help focus efforts to improve the precision of CSF peptide



**Figure 3.3.** **A)** Boxplot of peptide coefficients of variation within patient specific samples. **B)** Boxplots showing distribution of patient median %CVs (intra), %CVs of median peptide abundances between patients (inter), and %CVs across QC samples (technical). **C)** Principal Component Analysis of peptides detected in samples. QC samples and samples from patient 1 are circled, points are colored by patient. **D)** Pie chart of contributions to total variance of technical, intra, and inter variability.

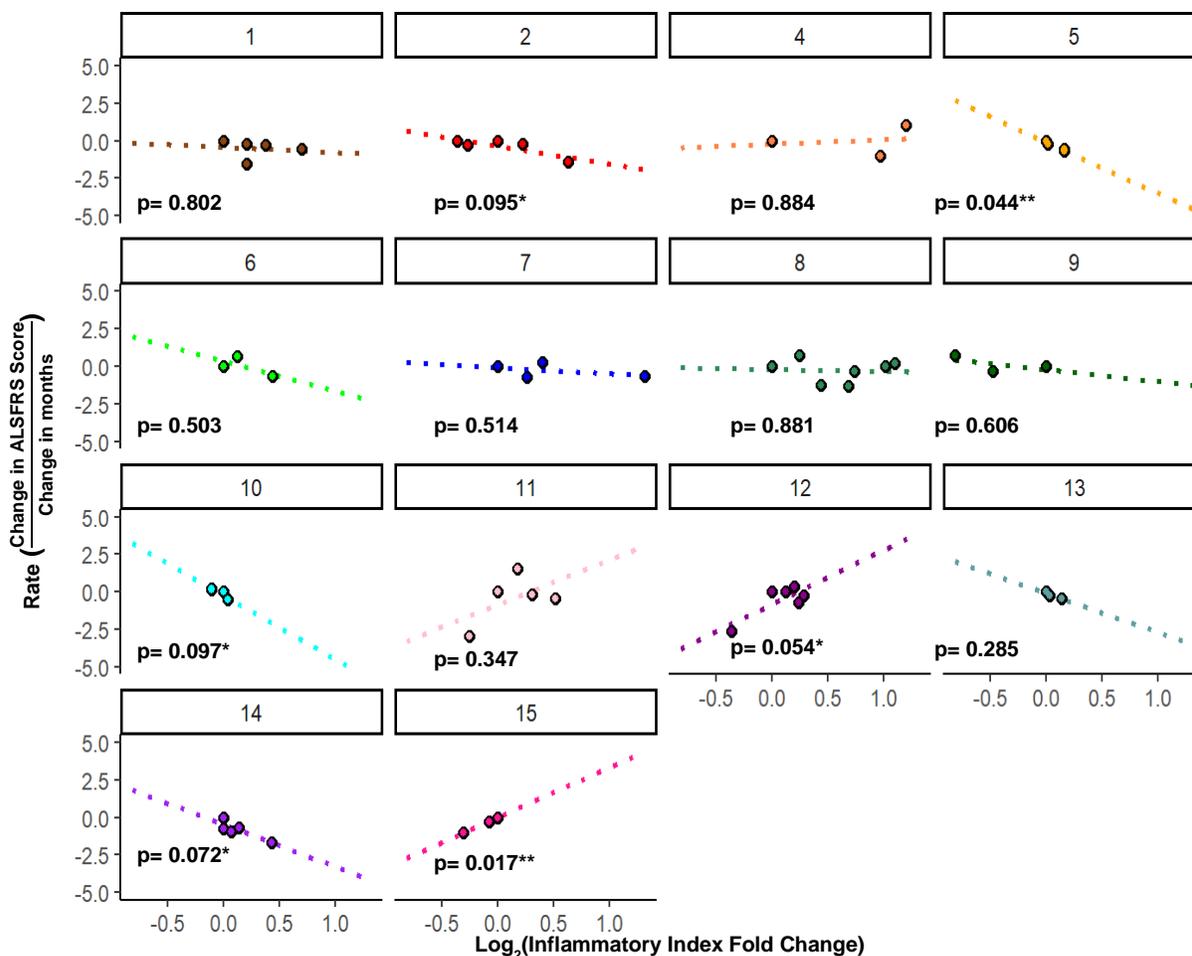
measurements (i.e., technical variability) and will be important for the design of future experiments (e.g., power calculations).

Next, we investigated proteins and respective metabolic pathways that were highly regulated versus more variable within individuals. By taking the intersection of all peptide abundances with a %CV of less than 20% within individuals, proteins found in both complement activation and coagulation cascade were reproducible. Complement activation and coagulation pathways are the front lines of defense upon cellular injury and more recently have been found to communicate with each other. Interestingly both pathways have been linked to ALS and proteins from these pathways are elevated in ALS CSF compared to controls<sup>17</sup>. These data suggest a sustained activation of both pathways in ALS. After removal of likely contaminant proteins (the top 10% most variable proteins), we did not observe any clear enrichment of pathways related to these proteins that were deemed more variable (%CV greater than 20%).

### *3.3.2 Longitudinal Biomarkers*

Although the specific mechanisms of disease pathogenesis remain unresolved, the association of neuroinflammation in the progression of ALS and its contributions to motor neuron death have been reported. However, inflammatory response has been found to have both protective and toxic effects on the system, depending on the stage of the disease. Most evidence of this dual nature of immune response is limited to the mouse model of the familial form of ALS. This hinders success in translating preclinical progress targeting neuroinflammation into proper therapeutic trials for the more common sporadic form of the disease in humans.<sup>18–20</sup> We were therefore interested in testing the hypothesis that inflammation is involved in disease progression within our human cohort (**Figure 3.4**). This is indicated by a negative association, or a faster loss in points

on the ALS-FRS with increased inflammation (i.e., neurotoxicity). A positive association indicates that increased inflammation leads to a slower decline on the scale (i.e., neuroprotective effects).



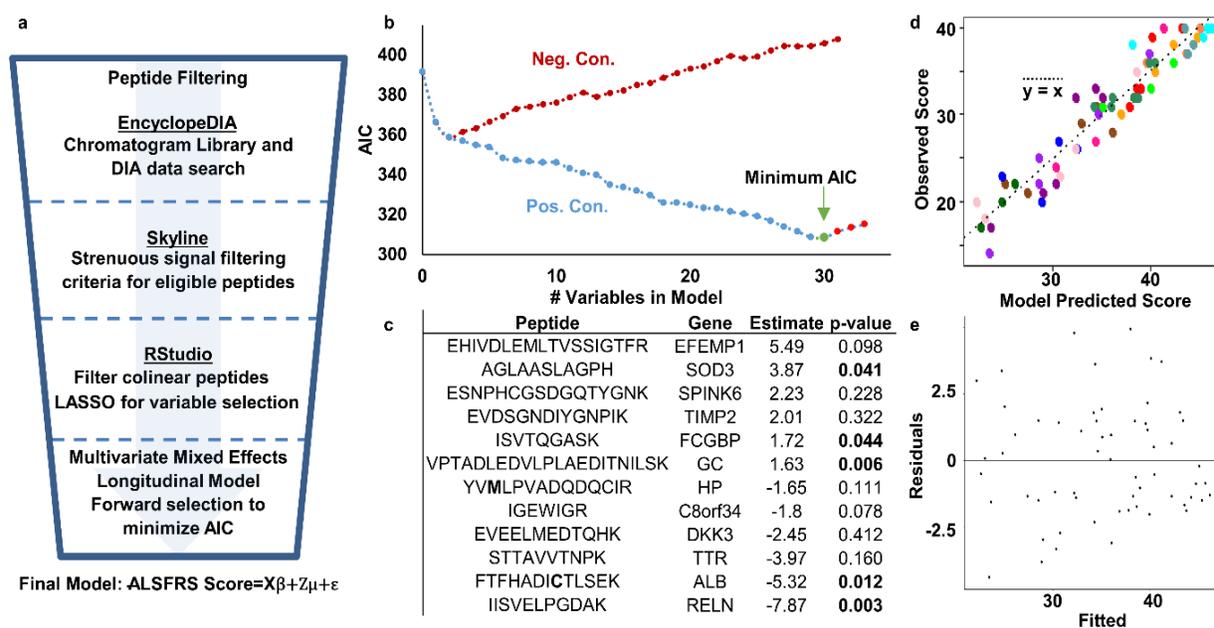
**Figure 3.4.** Rate of disease progression as a function of inflammatory response. Neurotoxic effects are represented by a loss of points on the ALS-FRS with an increase in inflammation, or a negative slope. Neuroprotective effects are therefore represented by a positive slope. Significant evidence of both negative and positive linear relationships is displayed (Patients 5 and 15, respectively) within this cohort.

In these patients, the log<sub>2</sub> fold changes from baseline of the inflammatory proteins ranged from -0.81 to 1.32. In most patients in this study, we observed a neurotoxic effect where marginal increases in fold change of these proteins resulted in faster decline on the ALS-FRS per month. Estimates ranged from -0.13 to -4.26 points per month with increases in fold change. The

relationship in patient “5” reached high statistical significance ( $p < 0.05$ ) while patients “2”, “10” and “14” showed moderate evidence of such a relationship ( $p < 0.1$ ). However, several patients also exemplified positive linear relationships, patient “15” reaching high statistical significance and patient “12” reaching moderate significance of a neuroprotective effect. These positive estimates ranged from 0.30 to 3.63 points per month with increases in fold change. These data lead to the possible conclusions: 1) That inflammation is both protective and toxic in ALS and its role may depend on the stage of disease, subtype, and or individual; 2) Due to the marginal significance observed, this group of proteins may be a poor gauge of neuroinflammation in ALS or 3) The FRS is a poor indicator of disease progression driven by neuroinflammation in ALS. Regardless, more research is needed of inflammation’s role in ALS with the view that a dynamic modulation of the immune system could be most therapeutic when approaching new treatments for the disease.<sup>18,21</sup>

Next, we took a global approach to investigating the efficacy of peptide measurements in describing disease progression in a quantitative manner. A summary of the approach used for peptide filtering and selection for mathematical modeling of ALS-FRS score is described in **Figure 3.5A**. 6396 peptides were identified in EncyclopeDIA software (corresponding to 951 proteins). The large number of identifications were filtered in Skyline to ensure signals were reproducible and quantifiable. Subsequently, a set of 96 non-collinear high confidence peptides were narrowed down using LASSO variable selection followed by filtering of protein sharing peptides. There does not yet exist a LASSO algorithm for mixed-effects models. However, we found using this approach for a general linear regression to be effective in selecting an appropriate number of peptides that could be manually tested using forward selection in a mixed-effects linear model of ALS-FRS. Finally, these 44 potential peptides were tested as fixed effects using forward selection in efforts to minimize the AIC <sup>6,22</sup> in the linear mixed-effects model. A random effect to account

for individual patients was included in the final model. In minimizing AIC (**Figure 3.5B**), we found that a minimum of 308 was reached with 30 variables as fixed effects. Time (years from disease symptom onset) and location of disease onset (limb vs bulbar) were included as fixed effects within the model along with 28 peptides to generate a well-fitting model of the ALS-FRS scores of this sample cohort. The top 12 peptides with the largest slope estimates of the final model are shown in **Figure 3.5C**. (All estimates are included in **Table B.2**.) The proteins for which these peptides provide the most evidence according to the EncyclopeDIA search algorithm is also included. The peptide slope estimates ranged from -7.87 points per increase in peptide log<sub>2</sub> peak area to 5.49 points per increase log<sub>2</sub> peak area, with several peptides having high evidence ( $p < 0.05$ ) of a linear slope not equal to zero. After accounting for other fixed effects, time was found to



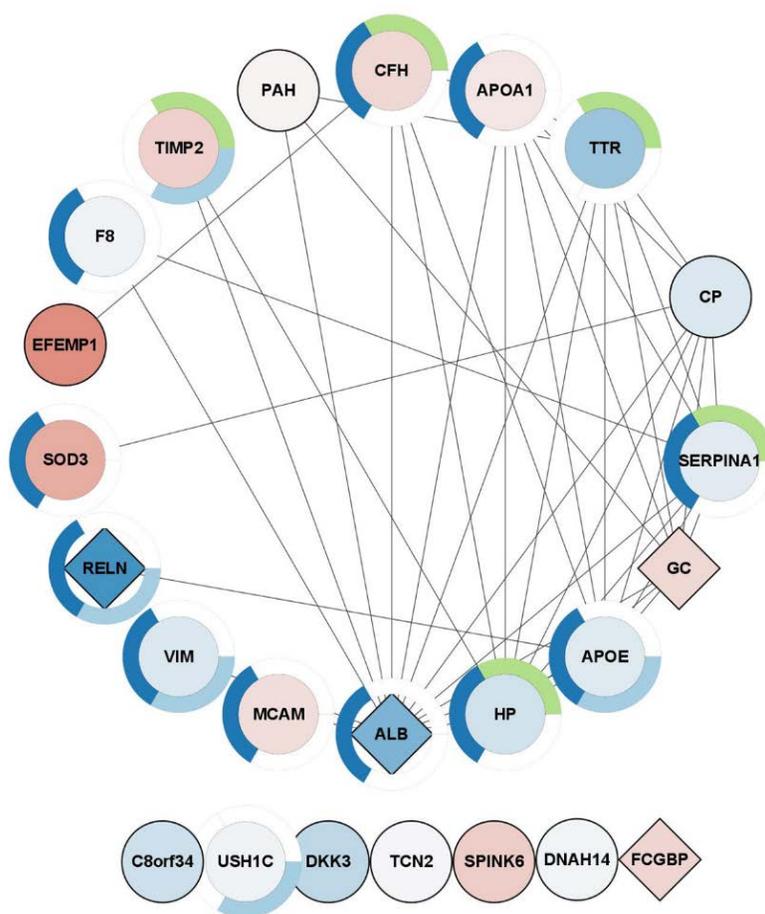
**Figure 3.5.** Mixed-effects model generation to predict ALS-FRS score via peptide measurements. **A)** Data analysis workflow and generation of a longitudinal model using peptide data to fit ALS-FRS scores. **B)** Forward selection of peptides as fixed effects after initial filtering and LASSO selection to minimize AIC. Negative control model was also created with permuted data. **C)** Summary of top contributing peptides to model. **D)** Plot comparing the patient scores predicted with model vs actual patient scores. Points colored by patient. Dotted refers to where the model predicted score is equal to observed score. **E)** Diagnostic residual plot.

linearly decrease score by 2.74 points per year from symptom onset and limb onset patients were found to have a score 11.3 points higher on average than patients with bulbar onset. The final model had an AIC of 308, compared to a null model AIC of 391.

Model fit was investigated by comparing the scores predicted using the model with the actual scores measured in the data set. Predicted and observed scores were nearly linear (**Figure 3.5D**). Residuals were randomly scattered around zero (**Figure 3.5E**) indicating constant variance.<sup>23</sup> A negative control model was also fit to predict ALS-FRS scores. Adding randomly selected variables did not decrease AIC values beyond the minimum found with forward selection (**Figure 3.5B**). To our knowledge, this is the first global proteomics study resulting in an effective and strictly quantitative longitudinal model of ALS disease progression. The variables within the model and their linear relationships with ALS-FRS are another interesting result. Several proteins selected for the model have already been implicated with ALS in cross-sectional studies. We found that alpha-1 antitrypsin (SERPINA1) decreased scores by 0.94 points per increase in  $\log_2$  peak area indicating that it increases with disease progression. Brettschneider and coworkers found this protein elevated in ALS patients. However, they found reduced levels ceruloplasmin precursor, whereas we found that it also increases with disease progression (1.18 decrease in points with increase in logged peak area) after accounting for other fixed effects.<sup>24</sup> Similarly, we found that transthyretin decreased score by -3.97 points per increase in logged peak area while it has been previously shown to be decreased in ALS patients vs healthy controls.<sup>25</sup> Other proteins included in our model have been previously implicated in neuromuscular degenerative diseases, and these are marked in **Table B.1**.

In addition, we wanted to explore the biological relationships that might have led to the simultaneous selection of these proteins for this model. An investigation of the interaction network

of the proteins included in this model is shown in **Figure 3.6**. Seventeen of the proteins create a significant interaction network. Within the interaction network, 11 proteins were associated with response to stress, and seven of these proteins had a negative estimate in the linear model. Three immune response related proteins also had a negative estimate (out of five identified), along with four out of five proteins identified as components of neuron projection. These findings further emphasize the important role of inflammation and stress response in the tracking of disease progression.



**Figure 3.6.** Network interactions of proteins with peptide representation in model. Node color represents the estimate value (blue represents a negative estimate, i.e., increase with disease progression; red represents a positive estimate, i.e., a decrease with disease progression.) Diamond shaped nodes have a p-value <0.05. Nodes with a green ring are associated with the Reactome pathway “innate immune system.” Dark blue ringed nodes are associated with the “response to stress” GO Process while light blue rings are associated with the “neuron projection” GO Component.

### **3.4 Conclusions**

A quantitative mathematical model of ALS patient disease progression was created by measuring changes in CSF peptide abundance. This model was successful in relating objective proteomic measurements in CSF to ALS-FRS, the current gold standard in monitoring disease progression. Further targeted studies and validation of these results should be undertaken in an independent cohort. However, this is a promising preclinical step supporting the efficacy of proteomic measurements in quantitative and objective modeling of ALS patient disease progression.

### **3.5 Acknowledgements**

We gratefully acknowledge the NEALS Biorepository for providing all the biofluids from ALS patients used in this study. We also acknowledge the ALS Association (grant #19-SI-458) for the funding of this project. All mass spectrometry measurements were made in the Molecular Education, Technology, and Research Innovation Center (METRIC) at NC State University.

### 3.6 References

- (1) Hawkrigde, A. M.; Muddiman, D. C. Mass Spectrometry–Based Biomarker Discovery: Toward a Global Proteome Index of Individuality. *Annu Rev Anal Chem (Palo Alto Calif)* **2009**, *2*, 265–277. <https://doi.org/10.1146/annurev.anchem.1.031207.112942>.
- (2) Enroth, S.; Johansson, Å.; Enroth, S. B.; Gyllensten, U. Strong Effects of Genetic and Lifestyle Factors on Biomarker Variation and Use of Personalized Cutoffs. *Nature Communications* **2014**, *5* (1), 1–11. <https://doi.org/10.1038/ncomms5684>.
- (3) Tibshirani, R.; Bien, J.; Friedman, J.; Hastie, T.; Simon, N.; Taylor, J.; Tibshirani, R. J. Strong Rules for Discarding Predictors in Lasso-Type Problems: Strong Rules for Discarding Predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2012**, *74* (2), 245–266. <https://doi.org/10.1111/j.1467-9868.2011.01004.x>.
- (4) Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry. *Nat Commun* **2018**, *9* (1), 1–12. <https://doi.org/10.1038/s41467-018-07454-w>.
- (5) R Core Team (2019). *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria.
- (6) Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* **1974**, *19* (6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>.
- (7) Wewer Albrechtsen, N. J.; Geyer, P. E.; Doll, S.; Treit, P. V.; Bojsen-Møller, K. N.; Martinussen, C.; Jørgensen, N. B.; Torekov, S. S.; Meier, F.; Niu, L.; Santos, A.; Keilhauer, E. C.; Holst, J. J.; Madsbad, S.; Mann, M. Plasma Proteome Profiling Reveals Dynamics

- of Inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery. *Cell Systems* **2018**, 7 (6), 601-612.e3. <https://doi.org/10.1016/j.cels.2018.10.012>.
- (8) Gille, B.; De Schaepdryver, M.; Goossens, J.; Dedeene, L.; De Vocht, J.; Oldoni, E.; Goris, A.; Van Den Bosch, L.; Depreitere, B.; Claeys, K. G.; Tournoy, J.; Van Damme, P.; Poesen, K. Serum Neurofilament Light Chain Levels as a Marker of Upper Motor Neuron Degeneration in Patients with Amyotrophic Lateral Sclerosis. *Neuropathology and Applied Neurobiology* **2019**, 45 (3), 291–304. <https://doi.org/10.1111/nan.12511>.
- (9) Corzett, T. H.; Fodor, I. K.; Choi, M. W.; Walsworth, V. L.; Turteltaub, K. W.; McCutchen-Maloney, S. L.; Chromy, B. A. Statistical Analysis of Variation in the Human Plasma Proteome. *J Biomed Biotechnol* **2010**, 2010. <https://doi.org/10.1155/2010/258494>.
- (10) Geyer, P. E.; Wewer Albrechtsen, N. J.; Tyanova, S.; Grassl, N.; Iepsen, E. W.; Lundgren, J.; Madsbad, S.; Holst, J. J.; Torekov, S. S.; Mann, M. Proteomics Reveals the Effects of Sustained Weight Loss on the Human Plasma Proteome. *Molecular Systems Biology* **2016**, 12 (12), 901. <https://doi.org/10.15252/msb.20167357>.
- (11) Harney, D. J.; Hutchison, A. T.; Hatchwell, L.; Humphrey, S. J.; James, D. E.; Hocking, S.; Heilbronn, L. K.; Larance, M. Proteomic Analysis of Human Plasma during Intermittent Fasting. *J. Proteome Res.* **2019**, 18 (5), 2228–2240. <https://doi.org/10.1021/acs.jproteome.9b00090>.
- (12) Hühmer, A. F.; Biringer, R. G.; Amato, H.; Fonteh, A. N.; Harrington, M. G. Protein Analysis in Human Cerebrospinal Fluid: Physiological Aspects, Current Progress and Future Challenges. *Dis. Markers* **2006**, 22 (1–2), 3–26. <https://doi.org/10.1155/2006/158797>.

- (13) Reiber, H. Proteins in Cerebrospinal Fluid and Blood: Barriers, CSF Flow Rate and Source-Related Dynamics. *Restorative Neurology & Neuroscience* **2003**, *21* (3/4), 79–96.
- (14) Reiber, H. Dynamics of Brain-Derived Proteins in Cerebrospinal Fluid. *Clinica Chimica Acta* **2001**, *310* (2), 173–186. [https://doi.org/10.1016/S0009-8981\(01\)00573-3](https://doi.org/10.1016/S0009-8981(01)00573-3).
- (15) Piehowski, P. D.; Petyuk, V. A.; Orton, D. J.; Xie, F.; Moore, R. J.; Ramirez-Restrepo, M.; Engel, A.; Lieberman, A. P.; Albin, R. L.; Camp, D. G.; Smith, R. D.; Myers, A. J. Sources of Technical Variability in Quantitative LC–MS Proteomics: Human Brain Tissue Sample Analysis. *J. Proteome Res.* **2013**, *12* (5), 2128–2137. <https://doi.org/10.1021/pr301146m>.
- (16) Stoop, M. P.; Coulier, L.; Rosenling, T.; Shi, S.; Smolinska, A. M.; Buydens, L.; Ampt, K.; Stingl, C.; Dane, A.; Muilwijk, B.; Luitwieler, R. L.; Sillevius Smitt, P. A. E.; Hintzen, R. Q.; Bischoff, R.; Wijmenga, S. S.; Hankemeier, T.; van Gool, A. J.; Luider, T. M. Quantitative Proteomics and Metabolomics Analysis of Normal Human Cerebrospinal Fluid Samples\*. *Mol Cell Proteomics* **2010**, *9* (9), 2063–2075. <https://doi.org/10.1074/mcp.M110.000877>.
- (17) Bereman, M. S.; Beri, J.; Enders, J. R.; Nash, T. Machine Learning Reveals Protein Signatures in CSF and Plasma Fluids of Clinical Value for ALS. *Sci Rep* **2018**, *8* (1), 1–14. <https://doi.org/10.1038/s41598-018-34642-x>.
- (18) Liu, J.; Wang, F. Role of Neuroinflammation in Amyotrophic Lateral Sclerosis: Cellular Mechanisms and Therapeutic Implications. *Front Immunol* **2017**, *8*. <https://doi.org/10.3389/fimmu.2017.01005>.
- (19) Komine, O.; Yamanaka, K. Neuroinflammation in Motor Neuron Disease. *Nagoya J Med Sci* **2015**, *77* (4), 537–549.

- (20) Hooten, K. G.; Beers, D. R.; Zhao, W.; Appel, S. H. Protective and Toxic Neuroinflammation in Amyotrophic Lateral Sclerosis. *Neurotherapeutics* **2015**, *12* (2), 364–375. <https://doi.org/10.1007/s13311-014-0329-3>.
- (21) Huang, F.; Zhu, Y.; Hsiao-Nakamoto, J.; Tang, X.; Dugas, J. C.; Moscovitch-Lopatin, M.; Glass, J. D.; Brown, R. H.; Ladha, S. S.; Lacomis, D.; Harris, J. M.; Scarce-Levie, K.; Ho, C.; Bowser, R.; Berry, J. D. Longitudinal Biomarkers in Amyotrophic Lateral Sclerosis. *Annals of Clinical and Translational Neurology* *n/a* (n/a). <https://doi.org/10.1002/acn3.51078>.
- (22) Shibata, R. An Optimal Selection of Regression Variables. *Biometrika*. 1981;68:45–54.
- (23) Altman, N.; Krzywinski, M. Regression Diagnostics. *Nat Methods* **2016**, *13* (5), 385–386. <https://doi.org/10.1038/nmeth.3854>.
- (24) Brettschneider, J.; Mogel, H.; Lehmensiek, V.; Ahlert, T.; Süßmuth, S.; Ludolph, A. C.; Tumani, H. Proteome Analysis of Cerebrospinal Fluid in Amyotrophic Lateral Sclerosis (ALS). *Neurochem Res* **2008**, *33* (11), 2358–2363. <https://doi.org/10.1007/s11064-008-9742-5>.
- (25) Ranganathan, S.; Williams, E.; Ganchev, P.; Gopalakrishnan, V.; Lacomis, D.; Urbinelli, L.; Newhall, K.; Cudkowicz, M. E.; Brown, R. H.; Bowser, R. Proteomic Profiling of Cerebrospinal Fluid Identifies Biomarkers for Amyotrophic Lateral Sclerosis. *J. Neurochem.* **2005**, *95* (5), 1461–1471. <https://doi.org/10.1111/j.1471-4159.2005.03478.x>.

## CHAPTER 4: Detailing a Method for Discovery Proteomics of Human Placental Tissue

Reused with permission from: Mellinger, A.L., McCoy, K., Minior, D.A.T., Williams, T.I. *Rapid Communications in Mass Spectrometry*, 2021, e9189. Copyright © 2021 John Wiley & Sons Ltd

### 4.1 Introduction

A label-free proteomics workflow for placental tissue analysis is described, one which allows modelling and exploration of toxicant exposure during gestation. Placental tissue from women who have engaged in recreational drug use during pregnancy can be investigated to determine proteins affected by such exposure. Understanding the natural detoxification processes of the human placenta could potentially assist in identifying targets for prenatal therapies to reduce the effects of toxicants on developing fetuses. A label-free proteomic method has been developed to confidently identify a robust number of protein targets within each tissue sample. When considering under-studied systems such as the human placenta, and ailments associated with abnormalities in placental function, discovery-based proteomics can serve as an effective platform for future hypothesis-driven investigations of placental function. Nine different tissue sections from each placenta were combined to form one sample for the work described here, to survey information across the entire placenta. This protocol can be applied to a large amount of placental tissue (~1-2 grams).<sup>1-6</sup> Although human tissue is specifically described in this protocol, this procedure may be used to study placental tissue from other mammals. However, the protein database against which the data is searched should be appropriate for the organism that is being studied.

#### 4.1.1 Key Points

- Label-free Discovery Proteomics
- Homogenization of >1 g placenta tissue
- Can be performed in ~3 days' time, post placental tissue collection and sampling.

#### 4.1.2 Acronyms and Abbreviations

- ABC: Ammonium bicarbonate
- AGC: Automatic gain control
- BCA: Bicinchoninic acid
- DDA: Data dependent acquisition
- DTT: Dithiothreitol
- FASP: Filter Aided Sample Preparation
- HCD: Higher-energy collision-induced dissociation
- HCl: Hydrochloric acid
- HCOOH: Formic acid
- IAA: Iodoacetamide
- MPA: Mobile phase A
- MPB: Mobile phase B
- MWCO filter unit: Molecular weight cut off filter unit
- PBS: Phosphate buffered solution
- PD: Proteome Discoverer
- QC: Quality Control
- RBC: Red blood cells
- SDC: Sodium deoxycholate
- *[A]*: signifies an important step in this protocol where alternative methods exist that may be used but were not tested here.

## 4.2 Materials and Methods

### 4.2.1 Materials

- Stainless steel beads (Fisher Scientific Catalog # 50-153-2314)
- Glass scintillation vials (Fisher Scientific Catalog # 03-341-25D)
- Vivacon-500, 30 kDa MWCO filters (Sartorius Catalog # VN01H22ETO)
- 1.5 mL Low Retention Eppendorf Tubes (Fisher Catalog # PI90410)
- 15 mL Falcon Tubes to prepare FASP solutions (Fisher Catalog # 05-527-90)
- 50 mL Falcon Tubes to prepare FASP solutions (Fisher Catalog # 14-959-49A)
- Oven, 37°C and 60°C (Fisher Catalog # 11-475-153)
- Acclaim PepMap™ C<sub>18</sub> trap column (Thermo Scientific Catalog # 164946)
- PepMap™ C<sub>18</sub> column (Thermo Scientific Catalog # ES802A)
- Autosampler vials and caps (Fisher Catalog # C4011-13 and 60180-676)
- Omni handheld homogenizer probe (Fisher Catalog # 15-340-167)
- Omni homogenizer tips (Fisher Catalog # 15-340-100)
- Fisherbrand Isotemp Shaker (Fisher Catalog # 02-217-741)
- IBI Scientific “The Belly Dancer” Orbital Platform Shaker (Fisher Catalog #15-453-211)
- Fisherbrand accuSkan FC Filter-Based Microplate Photometer (14-377-575)
- Thermo Scientific Sorvall Legend Micro 21 Microcentrifuge (Fisher Catalog #75-772-436)
- Eppendorf Microcentrifuge 5430R (Fisher Catalog #05-100-180)
- Fisherbrand Model 505 Sonic Dismembrator (Fisher Catalog FB5050110)
- Analytical balance
- Graduated cylinders, 10-500 mL are preferable
- Glass solvent bottle, 1 L

- Metal spatula to dispense solid reagents
- Weight boats to measure solid reagents

#### 4.2.2 Chemicals

- Optima LC/MS Grade water (Fisher Scientific Catalog # W6-4)
- Optima LC/MS Grade ACN (Fisher Scientific Catalog # A955-4)
- Formic acid (Fluka Catalog # 94318-50ML)
- Ammonium bicarbonate (Millipore Sigma # 09830)
- Sodium deoxycholate (Millipore Sigma # D6750)
- Acetic acid (Fisher Scientific Catalog # 143-16)
- Hydrochloric acid (Fisher Scientific Catalog # 60-007-63)
- Dithiothreitol (BioRad Catalog # 161-0611)
- Iodoacetamide (BioRad Catalog # 163-2109)
- Urea (BioRad Catalog # 161-0730)
- Sequencing Grade Modified Trypsin (Promega Catalog # V5111)
- Pierce Standard LC/MS Grade BSA Protein Digest (Thermo Scientific Catalog # 88341)
- Pierce Standard HeLa Protein Digest (Thermo Scientific Catalog # 88329)
- BCA Protein Concentration Assay (Fisher Scientific Catalog # 23225)
- Roche Protease Inhibitor (Millipore Sigma Catalog # 04693116001)
- Halt Protease Inhibitor (Thermo Scientific Catalog # 78430)
- Phosphate buffer solution (Fisher Scientific Catalog # BP661-10)
- Pierce BCA Protein Assay Kit (Catalog # 23225)

As with any proteomics sample preparation workflow, the samples, workspace, reagents and equipment should be safeguarded from keratin contamination. Keratin originates from skin and hair as well as dust, clothing and even latex gloves. If keratins are present in high concentrations, they can interfere with mass spectrometry analyses, rendering loss in signal and decreased identification of potentially interesting proteins. In the case of placental tissues, samples should also be adequately rinsed with ample quantities of sterile solution in order to minimize presence of maternal red blood cells (RBCs). Sterile water and phosphate buffered solution (PBS) with Halt protease inhibitor are used during various stages to ensure placental tissues are amply rinsed to minimize risk of contamination. Indeed, the issues with presence of haemoglobin from maternal RBCs can be similar to those with unwanted keratin presence in these proteomics samples. Workspaces and equipment should be kept clean, reagents should be prepared fresh (in clean containers) and gloves should be changed frequently. Leaning over the workspace must be avoided. Reading through the protocol in advance and pre-labelling any containers that are needed is recommended. Samples will change containers multiple times throughout the workflow and so keeping track of samples will be particularly important.

#### *4.2.3 Collection of Tissues for Proteomic Analysis<sup>7-9</sup>*

##### *4.2.3.1 Placental Collection and Storage*

Once a study participant infant is delivered and the study team has confirmed with the maternal medical team that the placenta was not intended for pathology review, the study team collects the umbilical cord and placenta. Placentas should be placed in freezer-safe plastic containers or wrapped in aluminum foil and placed into a plastic freezer bag for temporary storage at 4°C. Samples are transferred to -80°C within 24 hours, in keeping with data from Sjaarda *et al.*, 2018 who determined that refrigeration for up to 24 hours prior to

processing/storage does not compromise results. Placentas are kept in secure storage at  $-80^{\circ}\text{C}$  until time of sampling. Given the often-unpredictable timing of placenta availability and the need to ensure timely collection and storage, it is important to designate key personnel and procedures well in advance in order to maximize efficiency and safeguard the integrity of samples. *For the results reported in this study, a total of 64 placental samples were collected between September 2017 and August 2019. Of these, 20 samples (including 7 exposed and 13 controls) met inclusion criteria and were subsequently analyzed using the proteomics workflow detailed here.*

#### 4.2.3.2 Placental Sampling Protocol

##### Processing Time: 30-60 minutes per placenta

Frozen samples are removed from  $-80^{\circ}\text{C}$  and allowed to thaw at  $4^{\circ}\text{C}$  for approximately 18-22 hours. Placentas are placed with umbilical cord side up on a stainless-steel tray placed on top of frozen steel beads to keep tissues cold during processing. Punch biopsies are obtained in a 3x3 grid pattern from the left, middle, and right sides of the placenta per 3 rows, for a total of 9 biopsy samples to represent centrally and peripherally located tissues based on the systematic uniform random approach suggested by Mayhew, 2008. This sufficiently exceeds the minimum 4 samples per placenta to generate representative data suggested by Burton *et al.*, 2014. Each sample is taken with a 6 mm diameter biopsy punch. However, since placentas are different thicknesses, each core varies in mass between approximately 1.2 to 3.8 g. Photograph of this procedure is shown in **Figure 4.1**. Core samples are rinsed three separate times in PBS with Halt protease inhibitor on ice for 5 minutes each to remove maternal blood and leukocytes, possible confounders for proteomic analysis.<sup>9</sup> All nine core samples from each placenta are placed into glass scintillation vials. Samples are stored at  $-80^{\circ}\text{C}$  until they are processed.



**Figure 4.1.** Photograph of placenta tissue collection.

#### 4.2.4 Preparation of Placental Tissues for Proteomic nanoLC-MS/MS Analysis

Time Commitment: 2-3 days

##### 4.2.4.1 Preparation of Buffers Prior to Tissue Processing [A]

Processing Time: 30-45 minutes.

Buffers should be made fresh on the day of the sample preparation step. to prepare in advance, only measure out dry reagents and make appropriate solutions on the morning of sample preparation (**Tables 4.1 and 4.2**).

**Table 4.1.** Preparation of rinsing solution, protease inhibitor in water.

Processing time	Step	Comments / Tips
15-30 minutes	Measure out water scaled to appropriate amount needed. At least 25 mL per sample will be needed, but if tissues appear very bloody, more might be needed. High quantities of haemoglobin from blood can dominate mass spectrometry signal and diminish identifications of other interesting proteins. Add to glass solvent bottle.	Scaling Example: It is a good idea to prepare additional volume for ample supply. For 5 washes of 20 samples: 5 mL × 5 rounds × 20 samples + 50 mL extra volume = 550 mL

**Table 4.1** (continued.)

	Add one Roche Protease Inhibitor Tablet for every 50 mL water per manufacturer's instructions. Shake until dissolved.	If samples are very bloody, additional rinse steps can be applied. For example, it might take 10 wash steps (50 minutes total shaking, 50 mL rinsing solution per sample) to rinse out most of the blood from the sample.
	Rinsing solution should always be kept on ice.	

**Table 4.2.** Preparation of lysis buffer, 50 mM ABC, and 1% SDC (w/v) with protease inhibitor, 500 mL.

Processing time	Step	Comments / Tips
15-20 minutes	Measure out <b>2.0 ± 0.1 g</b> ABC and add to glass solvent bottle.	If necessary, this step can be scaled down to preserve reagents. The following should be considered when scaling: <ul style="list-style-type: none"> <li>□ 10 mL per sample to submerge sample tissue (more or less, depending on amount of sample)</li> <li>□ Volume needed as a blank for protein quantitation method</li> <li>□ Volume need to rinse sonicator probe (~ 3 mL per sample to prepare 3 fresh tubes for rinsing)</li> </ul>
	Measure out <b>5.0 ± 0.1 g</b> SDC and add to same glass solvent bottle.	
	Add <b>500 mL</b> of water to glass solvent bottle. Shake well until reagents are fully dissolved.	
	Add <b>10</b> (one for every 50 mL buffer) Roche Protease Inhibitor Tablets. Shake until dissolved.	
Note that lysis buffer should always be kept on ice.		

#### 4.2.4.2 Preparation of Reducing and Alkylating Agent Aliquots Prior to Proteolytic Digestion [A]

Processing Time: 30-45 minutes (Tables 4.3 and 4.4).

**Table 4.3.** Preparation of reducing agent, dithiothreitol (500 mM).

Processing time	Step	Comments / Tips
15-30 minutes	Weigh out <b>77.1 ± 0.1 mg</b> of dithiothreitol into an Eppendorf tube.	Prepared DTT solution can be stored at -20°C for at least 1 month.
	Reconstitute in <b>1 mL</b> of water.	
	Vortex.	
	Aliquot solution into Eppendorf tubes in 20 µL volumes and store at -20°C until ready for use.	

**Table 4.4.** Preparation of alkylating agent, iodoacetamide (500 mM).

Processing time	Step	Comments / Tips
15-30 minutes	Prepare in as little light as possible. Compound is sensitive to light.	Prepared IAA solution can be stored at -20°C for at least 1 month.
	Weigh out <b>305 ± 0.1 mg</b> of iodoacetamide into a 15 mL Falcon tube.	
	Reconstitute in <b>3.3 mL</b> of water.	
	Vortex.	
	Aliquot into Eppendorf tubes (volumes appropriate for a 6 µL aliquot per sample, between 30 µL and 150 µL for example) and store at -20°C in the dark until ready for use.	

#### 4.2.4.3 Preparation of Protein Solubilization Buffers Prior to Proteolytic Digestion [A]

Processing Time: 45-60 minutes

Buffers should be made fresh on the day of digestion, to maintain buffer strength. To prepare in advance, only measure out dry reagents and add water on the morning of digestion (Tables 4.5-4.7).

**Table 4.5.** Preparation of 50mM ABC with 1% (w/v) SDC.

Processing time	Step	Comments / Tips
15-30 minutes	Measure out <b>79.1 ± 0.1 mg</b> ABC and add this to a 50 mL Falcon Tube.	Carefully tip the Falcon Tube on its side before adding water. This prevents SDC from getting stuck in the bottom of the Falcon Tube and helps it to dissolve completely.
	Measure out <b>200.0 ± 0.1 mg</b> SDC and add to same Falcon Tube containing ABC.	
	Add approximately 10 mL water and vortex until dissolved. Allow bubbles to settle.	If necessary, centrifugation can help minimize bubbles from SDC.
	Add water until solution reaches the <b>20 mL</b> mark on the Falcon Tube. Vortex and allow bubbles to settle.	

**Table 4.6.** Preparation of 50 mM ABC (only).

Processing time	Step	Comments / Tips
15-30 minutes	Measure out <b>79.1 ± 0.1 mg</b> ABC into a 50 mL Falcon Tube.	The pH of ABC buffers in this protocol should be about 7.8.
	Add approximately 5 mL water and vortex until fully dissolved.	
	Add water until solution reaches the <b>20 mL</b> mark on the Falcon Tube.	

**Table 4.7.** Preparation of 50 mM ABC and 8 M urea.

Processing time	Step	Comments / Tips
15-20 minutes	Measure out <b>79.1 ± 0.1 mg</b> ABC into 50 mL Falcon Tube.	This will seem like a large volume of urea and will take a lot of time to fully dissolve. Add water carefully and slowly to avoid adding over the 20 mL point. Warming in your hands will help with dissolution. Do not heat the solution.
	Measure out <b>9.8 ± 0.1 g</b> urea into same Falcon Tube.	
	Add 10 mL water and vortex until mostly dissolved.	
	Bring up to 15 mL mark with water. Vortex for a few minutes until mostly dissolved.	

**Table 4.7** (continued.)

	Add water until solution reaches the <b>20 mL</b> mark on the Falcon Tube.	
--	--	--

#### 4.2.4.4 Safety Precautions

BSL2 facilities were used, and all contaminated materials and waste were appropriately autoclaved before disposal. Standard laboratory Personal Protective Equipment (PPE) should be worn during these experiments; this includes safety glasses, a clean lab coat or disposable gown, and disposable gloves. Since human tissues were involved, we also recommend masks and hair covers as part of standard PPE. This also minimizes risk of keratin contamination of samples. Homogenization/sonication of tissues should take place under a hood. Upon completion of sampling, excess tissue samples should be appropriately disposed of via approved human tissue waste disposal routes only. (**Table 4.8**)

#### 4.2.4.5 Tissue Preparation (1-2 days) [A]

##### Processing Time: 9-10 hours

For a 3-day preparation, consider completing this step separately. This is recommended for larger quantities of samples (>10). It can be done right after tissues are collected/received. For a 2-day preparation, consider completing as a first step and following with homogenization. All buffers and tissues should be kept on ice to prevent degradation of the tissue and indiscriminate cleavage of proteins. (**Tables 4.8-4.11**)

**Table 4.8.** Rinsing blood from samples.

Processing time	Step	Comments / Tips
2-3 hours	Prepare rinsing solution fresh on day of use (see above, Table 1). Solution should always be scaled based on the number of samples and kept on ice after preparation.	This can be done on a separate day. Depending on how many samples are being processed or how many rinses are required to remove as much visible blood as possible, this step may require longer net processing time.
	Add <b>5 mL</b> protease inhibitor solution to each sample. Samples should be kept on ice throughout entire washing procedure.	
	Shake samples on BellyDancer shaker at the highest speed possible while on ice for 5 minutes.	
	Replace inhibitor solution with fresh 5 mL, and repeat process five times for a total of 25 minutes of shaking (or more as needed, see tip).	

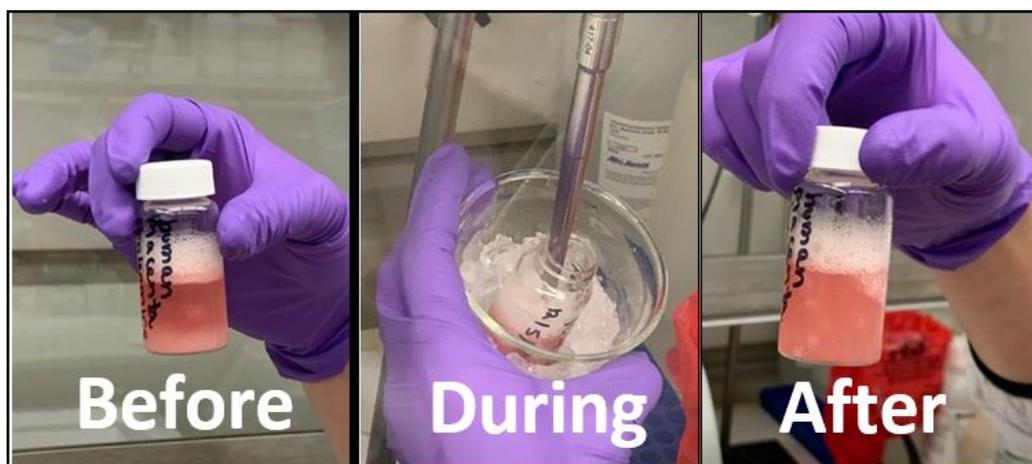
**\*End of Day 1 out of 3\***

**Table 4.9.** Tissue homogenization.

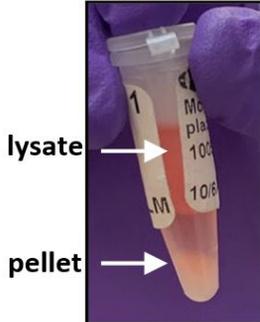
Processing time	Step	Comments / Tips
3-4 hours	Prepare lysis buffer (see above, Table 2). Lysis buffer should be kept on ice at all times.	<b>This step should take place under a hood.</b>  This can be done on a separate day. Depending on how many samples are being processed, this step may require longer net processing time.
	Add <b>10 mL</b> of cold lysis buffer to each tissue container. Tissue should be submerged in buffer. Tissues should be kept on ice.	
	Using the Omni handheld homogenizer under a hood, homogenize each sample individually on ice for 4-5 minutes until sample appears homogeneous in solution. Highest speed (35,000 rpm) is recommended for large amounts of tissue.	
	Replace with a fresh homogenizer probe before the next sample is processed. Place the samples on ice after homogenization.	

**Table 4.10.** Tissue sonication.

Processing time	Step	Comments / Tips
2-3 hours	While on ice, probe sonicate each tissue at 20% power for 2 minutes using the Fisherbrand Model 505 Sonic Dismembrator. Sample should appear more homogenous ( <b>Figure 4.2.</b> )	<b>This step should take place under a hood.</b>
	Aliquot approximately 1.0-1.5 mL sonicated sample into an Eppendorf tube. Immediately place on ice.	Homogenized samples tend to settle and separate while sitting on ice. Sonication is important to ensure a homogeneous aliquot of tissue homogenate is removed from the sample for analysis.
	In between samples, clean probe with methanol, followed by three clean tubes of fresh lysis buffer.	

**Figure 4.2.** Photograph of tissue homogenization procedure.

**Table 4.11.** Centrifugation and removal of lysate.

Processing time	Step	Comments / Tips
2-3 hours	Centrifuge all samples at 16,000 x g for 30 minutes at 4°C to form pellets.	With high quantities of tissue in 10 mL lysis buffer, concentration risks being higher than the calibration curve of the Pierce BCA Protein Assay Kit. It is therefore recommended to create a 1:10 diluted sample in lysis buffer prior to the total protein quantification assay analysis.
	Remove lysate to a fresh tube and label. <b><u>Care should be taken not to disturb the pellet.</u></b> Lysates should be kept on ice, and pellets can be stored at -80°C for potential future analysis, if needed. Lysates of samples are now ready for protein quantitation and digestion.	
	Quantitation of protein can be performed using the Pierce BCA Protein Assay Kit or using the Nanodrop A280 protein assay. Lysis buffer should be used as a blank. Following protein quantitation, lysates should be stored at -80°C until the beginning of the digestion procedure.	

**\* End of Day 2 out of 3 (or Day 1 out of 2 if tissue washing/homogenization steps are combined) \***

#### 2.2.4.6 Filter-Aided Sample Preparation (FASP) and Proteolytic Digestion<sup>5,6</sup> [A]

Processing Time: 9-10 hours

The remaining procedure should be completed within the same day. The procedure takes a minimum of 7 hours. If using one oven for the entire workflow, make sure that it is set to 60°C prior to beginning. Gloves should always be used and replaced often throughout procedure to avoid contamination. Long hair should be pulled back to minimize keratin contamination. Leaning over the samples and workspace should be avoided to further help with minimizing keratin impurities.

**(Table 4.12-4.17)**

**Table 4.12.** Aliquoting and diluting.

Processing time	Step	Comments / Tips
30-45 minutes	Remove samples from freezer. Once thawed, aliquot volume containing <b>50 µg</b> of sample into separate Eppendorf Tubes based off measured protein concentrations. This is important to normalize sample protein for the purpose of performing relative quantification of specific proteins detected in nanoLC-MS/MS analysis.	With large numbers of samples, it is very important to keep track of labelling. Samples will change containers a few times in this protocol.
	Bring samples <b>up to 200 µL</b> final volume with buffer containing 50 mM ABC and 1% (w/v) SDC to solubilize protein. Vortex.	

**Table 4.13.** Denaturing proteins by breaking disulfide bonds between cysteine residues.

Processing time	Step	Comments / Tips
45-60 minutes	Add <b>180 µL</b> buffer containing 500 mM ABC and 1% (w/v) SDC to thawed, 20 µL aliquot of DTT. If using more than one aliquot, combine into one Eppendorf tube and vortex. The DTT solution has now been diluted to 50 mM.	Prepare one DTT aliquot per 9 samples to account for any volume lost in pipetting steps.
	Add <b>20 µL</b> of 50 mM DTT to each sample to yield a final concentration of 4.5 mM DTT. Vortex each sample.	
	Incubate samples at 60°C for 30 minutes.	
	Remove from oven and allow to cool to room temperature before continuing.	

**Table 4.14.** Alkylating reactive thiols to prevent reformation of disulfide bonds.

Processing time	Step	Comments / Tips
30-45 minutes	If using more than one 500 mM IAA aliquot, combine into one Eppendorf tube and vortex.	<p><b>Perform alkylation in as low of light as possible. IAA is light sensitive.</b></p> <p>While samples are incubating with DTT, remove IAA aliquots from freezer and allow to thaw in the dark. Thaw enough aliquots for 6 <math>\mu\text{L}</math> per sample while accounting for any sample lost during pipetting steps (for example, thaw 6 <math>\mu\text{L}</math> <math>\times</math> # samples + 5-10 <math>\mu\text{L}</math>).</p>
	Add 6 $\mu\text{L}$ IAA to each sample to yield a final IAA concentration of 15 mM. Vortex.	
	Incubate samples at room temperature in the dark for 20 minutes.	

**Table 4.15.** Transferring samples to MWCO filtration units.

Processing time	Step	Comments / Tips
15-30 minutes	Label a Vivacon-500 30 kDa MWCO filtration unit (filter and tube) for each sample and passivate by adding 20 $\mu$ L 50 mM ABC + 1% SDC to the top of each filter.	Prepare (label and passivate) filtration units while samples are undergoing 20-minute alkylation reaction. The first collection tubes should be labelled “waste” as these will contain small molecules being washed from protein samples. A second set of collection tubes should be labelled “peptides” and set aside for post-digestion collection.  Note that passivating MWCO filters before use is important to prevent protein loss in the FASP workflow.
	Transfer entire volume of each sample to appropriate filtration unit.	
	Centrifuge sample-containing filtration units at 12,000 x <i>g</i> for 15 minutes.	

**Table 4.16.** Washing twice with buffer containing 50 mM ABC and 8 M urea.

Processing time	Step	Comments / Tips
1-2 hours	Add 200 $\mu$ L 50 mM ABC + 8 M urea to each sample filtration unit. Centrifuge units for 15 minutes at 12,000 x g.	<p>There are many steps requiring centrifugation. It is helpful to check off which steps have already been completed to keep track. Keeping an eye on the level of eluted solution can also be helpful since waste should be emptied every two spin cycles (so waste should be emptied a total of 2 times before switching to final collection tubes.). Sometimes, due to filter clogging, the filtration rate may be rather slow. In such a situation, it may be ok to increase the centrifugation speed (for any of the steps requiring centrifugation) to 12,500 x g or 13,000 x g, but no more.</p> <p>A number of washing steps are encountered in this workflow. Based on the buffer being used, these steps help to clean and denature the protein, getting the sample ready for proteolytic digestion. Note that some buffer components, like urea, are great for protein denaturation but must be removed from the samples prior to proteolytic digestion to prevent inhibition of the protease.</p>
	Empty eluent into waste so that it does not reach a volume that touches the filter. Handle the filter very carefully.	
	Repeat the wash with another 200 $\mu$ L 50 mM ABC + 8 M urea buffer. Centrifuge units for 15 minutes at 12,000 x g.	

**Table 4.17.** Washing twice with buffer containing 50 mM ABC.

Processing time	Step	Comments / Tips
1-2 hours	Add <b>200 <math>\mu\text{L}</math></b> 50 mM ABC ( <b>only ABC</b> ) to each sample filtration unit. Centrifuge units for 15 minutes at 12,000 x g.	Reconstitute trypsin during last wash step.
	Empty eluent into waste for the second time so that it does not reach a volume that touches the filter. Handle the filter very carefully.	Particularly for the last wash step, increase the centrifugation speed to 12,500 x g or 13,000 x g if need be to pass all the wash solution through the filter.
	Repeat the wash with another 200 $\mu\text{L}$ 50 mM ABC ( <b>only</b> ) buffer. Centrifuge units for 15 minutes at 12,000 x g.	

\*\*Change collection tubes under filter\*\*

All eluents should be kept from this point forward in collection tubes labelled “peptides.”

Waste collection tubes can be discarded. (**Tables 4.18-4.20**)

**Table 4.18.** Proteolytic digestion.

Processing time	Step	Comments / Tips
15-30 minutes + 4 hours shaking time	Reconstitute trypsin in <b>200 <math>\mu\text{L}</math></b> 0.01% acetic acid to stabilize protease and yield a 0.1 $\mu\text{g}/\mu\text{L}$ protease concentration. Vortex.	Wrap a small amount of parafilm around tops of filtration units to prevent them from popping open during the 4-hour digestion period. Remove the parafilm before centrifugation (following digestion).
	Add <b>70 <math>\mu\text{L}</math></b> 50 mM ABC buffer to each filtration unit.	
	Add <b>10 <math>\mu\text{L}</math></b> trypsin solution directly to ABC volume on top of filter for a 1:50 enzyme to protein ratio. Do not pipette on to the sides of the filter.	
	Shake filtration units at 600 rpm at 37°C for 4 hours on Isotemp Shaker.	

**Table 4.19.** Elution of tryptic peptides.

Processing time	Step	Comments / Tips
45-60 minutes	Centrifuge filtration units for 15 minutes at 12,000 x g to elute peptides.	To maximize tryptic peptide recovery, increase the centrifugation speed to 12,500 x g or 13,000 x g if need be to pass all the digest solution through the filter.
	Add <b>15.8 µL</b> 50 mM ABC to elute any remaining peptides on filter. Centrifuge for 15 min at 12,000 x g.	

**Table 4.20.** Quenching of enzyme.

Processing time	Step	Comments / Tips
15-30 minutes	Discard filters.	Acidifying the peptide solution quenches the trypsin protease.
	Add <b>4.2 µL</b> of 6 M HCl to each sample.	

**\*End of Day 3 out of 3 (or Day 2 out of 2) \***

Samples are now ready for nanoLC-MS/MS analysis. An aliquot (~20 µL) of each sample can be added to an appropriately labelled LC-MS autosampler vial and placed in the nanoLC autosampler for analysis. The remaining sample can be stored at -20°C. Two microliter injections are performed during nanoLC-MS/MS experiments. (**Tables 4.21-4.22**)

#### 4.2.5 Separation and Analysis by nanoLC-MS/MS [A]

Sample order should be randomized prior to sequencing to avoid bias from sequencing order. See section on “Quality Control” for recommendations on interspersed quality control and blank sample sequencing. Each sample should be subjected to online desalting and reversed-phase nano-LC separation (in a “trap and elute” configuration.) Separated peptides are ionized via electrospray for mass spectrometry analysis. A data dependent acquisition method for untargeted or discovery analysis should be used. The following tables (**Tables 4.21 and 4.22**) provide recommended settings for a two-hour analytical separation and DDA analysis, with a 140-minute

total method length, using an EASY nano-LC system coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Scientific).

Tip: It is helpful to find a systematic way of loading the samples into the autosampler during longer experiments. This prevents degradation of samples that are analyzed later from sitting in the autosampler for more than a few days.

**Table 4.21.** nanoLC Parameters for two-hour separation.

<b>Sample pickup and Loading</b>			
<b>Sample Pickup</b>			
Volume	2	μL	
Flow	10	μL/min	
<b>Sample Loading</b>			
Volume	6	μL	
Flow	1	μL/min	
Max. Pressure	700	Bar	
<b>Solvents</b>	A: water/acetonitrile + formic acid (98/2 + 0.1 % by volume)		B: water/acetonitrile + formic acid (80/20 + 0.1 % by volume)
<b>Gradient</b>			
Time	Duration	Flow (nL/min)	%B
0:00	0:00	300	5
2:00	2:00	300	5
107:00	105:00	300	25
122:00	15:00	300	40
123:00	1:00	300	95
124:00	1:00	300	95
140:00	16:00	300	95

**Table 4.21** (continued.)

<b>Pre-column and Analytical Column</b>		
<b>Pre-column Equilibration</b>		
Volume	3	μL
Flow	3.00	μL/min
Max. Pressure	700	Bar
<b>Analytical Column Equilibration</b>		
Volume	6	μL
Flow	-	μL/min
Max. Pressure	700	Bar
*Custom wash used for <b>Autosampler</b> and column heating controlled externally on Easy-Spray Ion Source. Column is kept at 45 °C throughout analysis.		

**Table 4.22.** Mass spectrometric parameters (Data Dependent Acquisition).

<b>Ion Source Properties</b>	
Ion Source Type	NSI
Spray Voltage	Static
Positive Ion (V)	1800
Gas Mode	Static
Sweep Gas (Arb)	0
Ion Transfer Tube Temp (°C)	275
FAIMS Mode	Not Installed
<b>Tune Settings</b>	
Application Mode	Peptide
Method Duration	140
Infusion Mode	Liquid Chromatography
Expected LC Peak Width (s)	15
Advanced Peak Determination	Checked
Default Charge State	2
Internal Mass Calibration	User-defined Lock Mass
Current Lock Mass for Positive Ion Mode	445.12006
<b>Method</b>	
Time Range (min)	0-140
Full Scan	2 sec
MIPS	
Intensity	
Charge State	
Dynamic Exclusion	
ddMS <sup>2</sup>	

Table 4.22 (continued.)

<b>Full Scan Properties</b>	
Orbitrap Resolution	120000
Scan Range (m/z)	375-1600
RF Lens (%)	40
AGC Target	Custom
Normalized AGC Target (%)	300
Maximum Injection Time Mode	Custom
Maximum Injection Time (ms)	100
Microscans	1
Data Type	Profile
Polarity	Positive
Source Fragmentation	Unchecked
<b>MIPS Properties</b>	
Monoisotopic peak determination	Peptide
Relax restrictions when too few precursors are found	Checked
<b>Intensity Properties</b>	
Filter Type	Intensity Threshold
Intensity Threshold	5.00E+03
<b>Charge State Properties</b>	
Include charge state(s)	2-6
Include undetermined charge states	unchecked
<b>Dynamic Exclusion Properties</b>	
Dynamic Exclusion Mode	Custom
Exclude after n times	1
Exclusion duration (s)	20
Mass Tolerance	ppm
Low	10
High	10
Exclude isotopes	checked
Perform dependent scan on single charge state per precursor only	checked

**Table 4.22** (continued.)

<b>Data Dependent MS<sup>2</sup> Scan Properties</b>	
Multiplex ions	unchecked
Isolation Window (m/z)	1.4
Isolation Offset	Off
Collision Energy Mode	Fixed
Collision Energy Type	Normalized
HCD Collision Energy (%)	30
Orbitrap Resolution	7500
TurboTMT	Off
Scan Range Mode	Define First Mass
First Mass (m/z)	100
AGC Target	Custom
Normalized AGC Target (%)	100
Maximum Injection Time Mode	Custom
Maximum Injection Time (ms)	18
Microscans	1
Data Type	Centroid

### 4.3 Quality Control (QC)

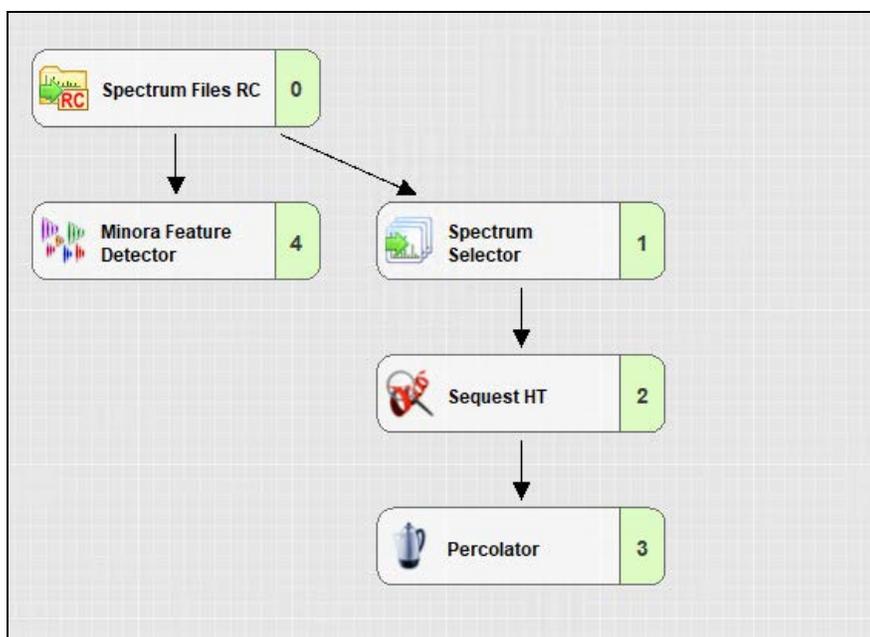
Randomized samples, BSA QC (std. BSA digest) and blanks (MPA) are run as follows: blank, BSA QC, blank, Sample-01, Sample-02, Sample-03, Sample 04, blank, BSA QC, blank, Sample-05, Sample-06, Sample-07, Sample-08, blank, BSA QC, blank... etc. Through the AutoQC software on PanoramaWeb (<https://panoramaweb.org/home/>), BSA QCs enable the monitoring of such metrics as retention time reproducibility, peak area and FWHM in real time.<sup>7</sup> A guide set should first be created (prior to any experiments) from five- seven independent BSA runs with each trap column, analytical column, or solvent change to account for shifts in retention times from these factors. Metrics are expected stay within three standard deviations from the guide set mean value. If these metrics fall outside of three standard deviations or are not reproducible, this could indicate an issue with instrument performance. For example, if major shifts in peptide retention times are visible, this may indicate an issue with the nLC system or columns. Users should troubleshoot the system as needed and consider re-analyzing biological replicates if metrics

indicate system deterioration during the experiment. The importance of monitoring system suitability in bottom-up proteomics experiments is discussed further by Bereman and coworkers.<sup>10</sup>

Several standard HeLa digest runs (HeLa QC) are conducted at the beginning and end of experiments to check on proteome coverage. Significant drops in coverage between runs may indicate a deterioration in system suitability.

#### **4.4 Label-Free Data Analysis**

Data files are now ready to be analyzed using software such as Proteome Discoverer for identification and quantification of proteins. The attached analysis workflow files for use with Proteome Discoverer 2.4 software are summarized in the following figures and tables. The basic parameters listed can be used with alternative search software. The identification of haemoglobin protein may be checked to evaluate degree of interference. While the data shared here is in raw abundance format (*i.e.*, unnormalized) quantification methods should be carefully considered before drawing statistical and biological conclusions from such a dataset. Normalization methods are available in Proteome Discoverer 2.4 (*e.g.*, to the total peptide amount in the Precursor Ions Quantifier node.) Open-source tools such as NormalizerDe are available to assist in normalization method selection and dataset analysis.<sup>11</sup> (**Figures 4.3 and 4.4**) (**Tables 4.23-4.26**)



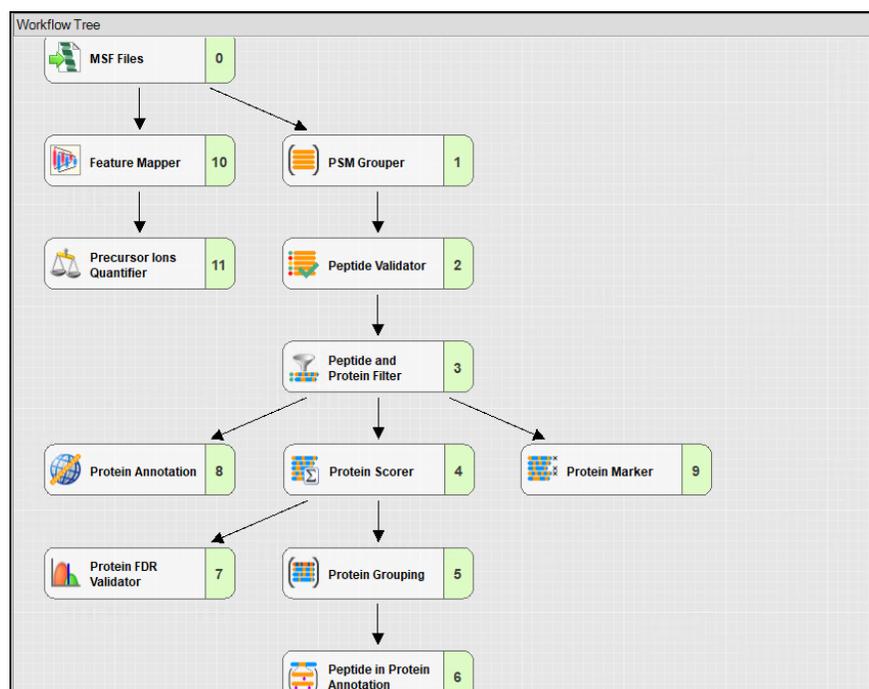
**Figure 4.3.** Processing workflow tree for Proteome Discoverer.

**Table 4.23.** Processing workflow nodes and basic parameters.

<b>Spectrum Files RC</b>	
<b>Search Settings</b>	
Protein Database	Contaminants.fasta; Homo Sapiens (Swiss Prot TaxID=9606)
Enzyme Name	Trypsin (Full)
1. Static Modification	Carbamidomethyl/ +57.021 Da (C)
<b>Spectrum Selector</b>	
<b>General Settings</b>	
Precursor Selection	Use MS1 Precursor
Provide Profile Spectra	TRUE
<b>Spectrum Properties Filter</b>	
Min. Precursor Mass	350 Da
Max. Precursor Mass	5000 Da
<b>Peak Filters</b>	
S/N Threshold (FT-only)	1.5

Table 4.23 (continued.)

<b>Sequest HT</b>	
<b>Input Data</b>	
Protein Database	Contaminants.fasta; Homo Sapiens (Swiss Prot TaxID=9606)
Enzyme Name	Trypsin (Full)
Max. Missed Cleavage Sites	2
Min. Peptide Length	6
Max. Peptide Length	144
<b>Tolerances</b>	
Precursor Mass Tolerance	5 ppm
Fragment Mass Tolerance	0.02 Da
<b>Spectrum Matching</b>	
Weight of a Ions	0.2
Weight of b Ions	1
Weight of x ions	0.2
Weight of y ions	1
<b>Dynamic Modifications</b>	
Max. Equal Modifications Per Peptide	2
Max. Dynamic Modifications Per Peptide	3
1. Dynamic Modification	Oxidation/ +15.995 Da (M)
2. Dynamic Modification	Deamidated/ +0.984 Da (N, Q)
<b>Dynamic Modifications (protein terminus)</b>	
1. N-Terminal Modification	Acetyl/ +42.011 Da (N-terminus)
2. N-Terminal Modification	Met-loss + Acetyl/ -89.030 Da (M)
<b>Static Modifications</b>	
1. Static Modification	Carbamidomethyl/ +57.021 Da (C)
<b>Percolator</b>	
<b>Target/Decoy Strategy</b>	
Target/Decoy Selection	Concatenated
Validation based on	q-Value
<b>Input Data</b>	
Maximum Delta CN	0.05
<b>FDR Targets</b>	
Target FDR (Strict)	0.01
Target FDR (Relaxed)	0.05



**Figure 4.4.** Consensus workflow tree for Proteome Discoverer.

**Table 4.24.** Consensus workflow nodes and basic parameters.

<b>MSF Files</b>	
<b>Storage Settings</b>	
Spectra to Store	Identified or Quantified
Feature Traces to Store	All
<b>PSM Grouper</b>	
<b>Peptide Group Modifications</b>	
Site Probability Threshold	75
<b>Peptide Validator</b>	
<b>General Validation Settings</b>	
Validation Mode	Automatic (Control peptide level error rate if possible)
Target FDR (Strict) for PSMs	0.01
Target FDR (Strict) for Peptides	0.01
<b>Peptide and Protein Filter</b>	
<b>Peptide Filters</b>	
Peptide Confidence At Least	High
Keep Lower Confident PSMs	FALSE
<b>Protein Filters</b>	
Minimum Number of Peptide Sequences	1

Table 4.24 (continued.)

<b>Protein Grouping</b>	
<b>Protein Grouping</b>	
Apply strict parsimony principle	TRUE
<b>Peptide in Protein Annotation</b>	
<b>Flanking Residues</b>	
Annotate Flanking Residues of the Peptide	TRUE
<b>Modifications in Peptide</b>	
Protein Modifications Reported	Only for Master Proteins
<b>Modifications in Protein</b>	
Modification Sites Reported	All And Specific
Minimum PSM Confidence	High
Report Only PTMs	True
<b>Positions in Protein</b>	
Protein Positions for Peptides	Only for Master Proteins
<b>Feature Mapper</b>	
<b>Chromatographic Alignment</b>	
Perform RT Alignment	TRUE
Maximum RT Shift [min]	2
<b>Feature Linking and Mapping</b>	
Minimum S/N Threshold	3
<b>Precursor Ions Quantifier</b>	
<b>General Quantification Settings</b>	
Peptides to Use	All
<b>Precursor Quantification</b>	
Precursor Abundance Based On	Area
<b>Normalization and Scaling</b>	
Normalization Mode	Total Peptide Amount
<b>Exclude Peptides from Protein Quantification</b>	
For Protein Roll-Up	Use All Peptides
<b>Quan Rollup and Hypothesis Testing</b>	
Protein Abundance Calculation	Summed Abundances
N for Top N	5
<b>Protein Annotation</b>	
<b>Annotation Aspects</b>	
1. Aspect	Biological Process
2. Aspect	Cellular Component
3. Aspect	Molecular Function

**Table 4.24** (continued.)

<b>Protein Marker</b>	
<b>Contaminant Database</b>	
Protein Database	Contaminants.fasta
<b>Protein FDR Validator</b>	
<b>Confidence Thresholds</b>	
Target FDR (Strict)	0.01
Target FDR (Relaxed)	0.05

**Table 4.25.** Summary of protein identification results.

<b>Sample</b>	<b>Proteins</b>	<b>High Confidence Proteins</b>	<b>Quantified Proteins</b>	<b>Peptides</b>	<b>PSMs</b>	<b>MS/MS</b>
1	3,144	2,775	2627	23,375	40,593	222,373
2	3,648	3,178	3021	27,178	44,909	225,954
3	3,310	2,937	2788	24,300	42,360	221,100
4	3,267	2,853	2718	23,891	41,239	220,129
5	3,361	2,872	2757	25,485	43,301	217,922
6	3,726	3,266	2088	27,736	45,731	225,942
7	3,376	2,957	2803	24,598	41,500	222,897
8	3,034	2,651	2520	22,169	39,336	224,590
9	3,649	3,166	2982	27,433	45,999	224,925
10	3,225	2,779	2642	24,077	43,041	219,667
11	3,734	3,333	3150	27,411	44,724	224,688
12	3,544	3,110	2938	25,842	43,169	223,632
13	3,706	3,251	3068	27,647	45,761	225,482
14	3,763	3,302	3110	27,789	45,395	224,921
15	3,372	2,938	2779	24,681	43,730	223,157
16	3,285	2,883	2744	24,323	42,589	226,070
17	3,512	2,992	2842	25,502	43,077	224,773
18	3,788	3,324	3145	27,998	46,399	226,754
19	3,158	2,798	2622	22,823	37,952	219,531
20	3,402	2,966	2837	24,998	43,694	223,508
<b>Average</b>	3,450	3,017	2809	25,463	43,225	223,401
<b>Standard Deviation</b>	233	209	251	1,826	2,280	2,530
<b>%CV</b>	6.8	6.9	8.9	7.2	5.3	1.1

**Table 4.26.** Summary of hemoglobin identification across samples.

	<b>Peptides</b>	<b>PSMs</b>
<b>Hemoglobin Sum (averaged over 20 replicates)</b>	107 +/- 8	1,996 +/- 366
<b>Total Sum (averaged over 20 replicates)</b>	25,463 +/- 1826	43,225 +/- 2,280
<b>% of Total Identifications</b>	0.42%	4.60%

#### 4.5 Acknowledgements

This work was performed in part by the Molecular Education, Technology and Research Innovation Center (METRIC) at NC State University, which is supported by the State of North Carolina. Research reported in this publication was supported in part by NIEHS under award number P30ES025128, and R15ES028436. Seed funding was also supplied by the Division of Neonatal-Perinatal Medicine within the Department of Pediatrics at East Carolina University/The Brody School of Medicine.

#### 4.6 Data Availability Statement

Raw data from placenta and quality control samples along with the BSA sample quality control Skyline file will be available via PanoramaWeb.org under the NCSU-METRIC Project, labelled “20210127- Proteomics of Human Placenta” or via the following link:

<https://panoramaweb.org/NCSU%20-%20METRIC/METRIC%20Public%20Data/20200127->

[%20Proteomics%20of%20Human%20Placenta/project-begin.view?](https://panoramaweb.org/NCSU%20-%20METRIC/METRIC%20Public%20Data/20200127-%20Proteomics%20of%20Human%20Placenta/project-begin.view?)

#### 4.7 References

- (1) Robinson J.M., Ackerman IV W.E., Kniss D.A., Takizawa T., Vandr  D.D. Proteomics of the Human Placenta: Promises and Realities. *Placenta*. **2008**; 29 (2): 135-143.
- (2) Jin X., Xu, Z., Cao, J., *et al.* Proteomics analysis of human placenta reveals glutathione metabolism dysfunction as the underlying pathogenesis for preeclampsia. *BBA- Proteins and Proteomics*. **2017**; 1865: 1207-1214.
- (3) Bian, Y., Zheng, R., Bayer, F., *et al.*, Robust, reproducible, and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. *Nat. Comm.* **2020**; 11 (157): 1-12.
- (4) Luyten, L.J., Dieu, M., Demazy, C., *et al.*, Optimization of label-free nano LC-MS/MS analysis of the placental proteome. *Placenta*. **2020**; 101: 159-162.
- (5) Wi niewski, J., Zougman, A., Nagaraj, N. *et al.* Universal sample preparation method for proteome analysis. *Nat. Methods*. 2009; 6: 359–362.
- (6) Porter, C.J., Bereman, M.S., Comparison of LC-MS/MS Compatible Detergents with Sodium Deoxycholate for Shotgun Proteomics. *J. Proteins and Proteomics*. **2014**; 151-161.
- (7) Sjaarda, L.A., Ahrens, K.A., Kuhr, D.L., *et al.* Pilot study of placental tissue collection, processing, and measurement procedures for large-scale assessment of placental inflammation. *PLOS ONE*, **2018**; 13 (5): 1-10.
- (8) Mayhew, T.M. Taking tissue samples from the placenta: An illustration of principles and strategies. *Placenta*, **2008**; 29 (1): 1-14.
- (9) Burton, G.J., Sebire, N.J., Myatt, L., *et al.* Optimising sample collection for placental research. *Placenta*, **2014**; 35 (1): 9-22.

- (10) Bereman, M.S., Beri, J., Sharma, V., *et al.* An Automated Pipeline to Monitor System Performance in Liquid Chromatography Tandem Mass Spectrometry Proteomic Experiments. *J. Proteome Res.*, **2016**; *15* (12): 4763-4769.
- (11) Willforss, J.W., Chawade, A., Levander, F., NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *J. Proteome Res.*, **2019**, *18* (2), 732-740.

## **CHAPTER 5: Highlighting Functional Mass Spectrometry Imaging (fMSI) Methods in Bioanalysis**

Reprinted with permission from Mellinger, A.L., Muddiman, D.C., Gamsik, M.P. *Journal of Proteome Research*, **2022**, 21, 1800-1807. Copyright © 2022 American Chemical Society.

### **5.1 Introduction**

Mass spectrometry is the primary tool used to provide detailed proteomic and metabolomic data on biological tissue. Often, tissue samples are homogenized prior to analysis and the resulting data represents a volume-averaged level of molecular components. However, both healthy and pathologic tissue are heterogeneous in structure and cell types that function under the influence of varying microenvironments. This results in a widely diverse collection of molecular content across the tissue. Therefore, the molecular data collected depends upon the tissue sampling site. The introduction of mass spectrometry imaging (MSI) methods has largely addressed this issue and molecular content can now be mapped from the nanoscale to entire organs. Combined with histological mapping, the MSI data can also be linked to cellular organelles, anatomical structures or cell types.<sup>1</sup>

### **5.2 Why prioritize functional measurements?**

Even with the mapping capability, molecular methods such as MSI often provide a static picture of the steady-state levels of tissue content. These steady-state levels are the net result of the dynamic processes that import/produce and export/consume a particular species in the tissue. Healthy biosystems often strive to maintain these levels within a narrow range by homeostatic mechanisms even when under stress. For example, exercising muscle consumes ATP 100-times faster than resting muscle yet the steady-state levels decrease by only 25%.<sup>2</sup> Therefore, limiting analyses to steady-state levels will not provide a complete picture of how tissues function. In

addition, many biomarkers used in clinical diagnosis are discovered by detected differences in the molecular profiles (i.e., steady-state levels) of healthy and diseased tissues. Since steady-state levels are a result of two contributing dynamic processes (production/consumption), changes in these levels must be preceded by changes in one or both processes. Earlier detection of disease, therefore, may be possible by examining tissue dynamics rather than limiting the analyses to steady-state levels thereby expanding the range of potential biomarkers. These dynamic processes are best measured by assessing metabolic activity or flux; that is the rate of transport or transformation of molecular species. These flux data are direct measures of tissue function that are the net output of all molecular interactions and regulatory processes under the influences of the microenvironments.<sup>3</sup> Microenvironments that influence flux include cell-matrix, cell-cell interactions, hormones, cytokines, growth factors and nutrient/oxygen status that are affected by perfusion and will vary across all tissues. These environmental effects are difficult to account for in molecular inventories or when constructing models of tissue function. Since function, like molecular content, will vary across tissue, functional imaging methods are necessary.

### **5.3 Functional Imaging**

There is no generally accepted definition for functional imaging, but the term seems to have originated in neuroscience to describe mainly noninvasive imaging methods such as positron-emission tomography (PET), electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) to map brain biological processes such as metabolism, electrical activity, and blood oxygenation changes. The field of functional imaging has since evolved beyond the brain. One of the most well-known examples of functional imaging is the uptake of 2-fluorodeoxyglucose by <sup>18</sup>F-fluorodeoxyglucose PET (FDG-PET). The FDG-PET signal detected in tissues is the result of several biological processes and reflects the rate of glucose delivery, cellular uptake and

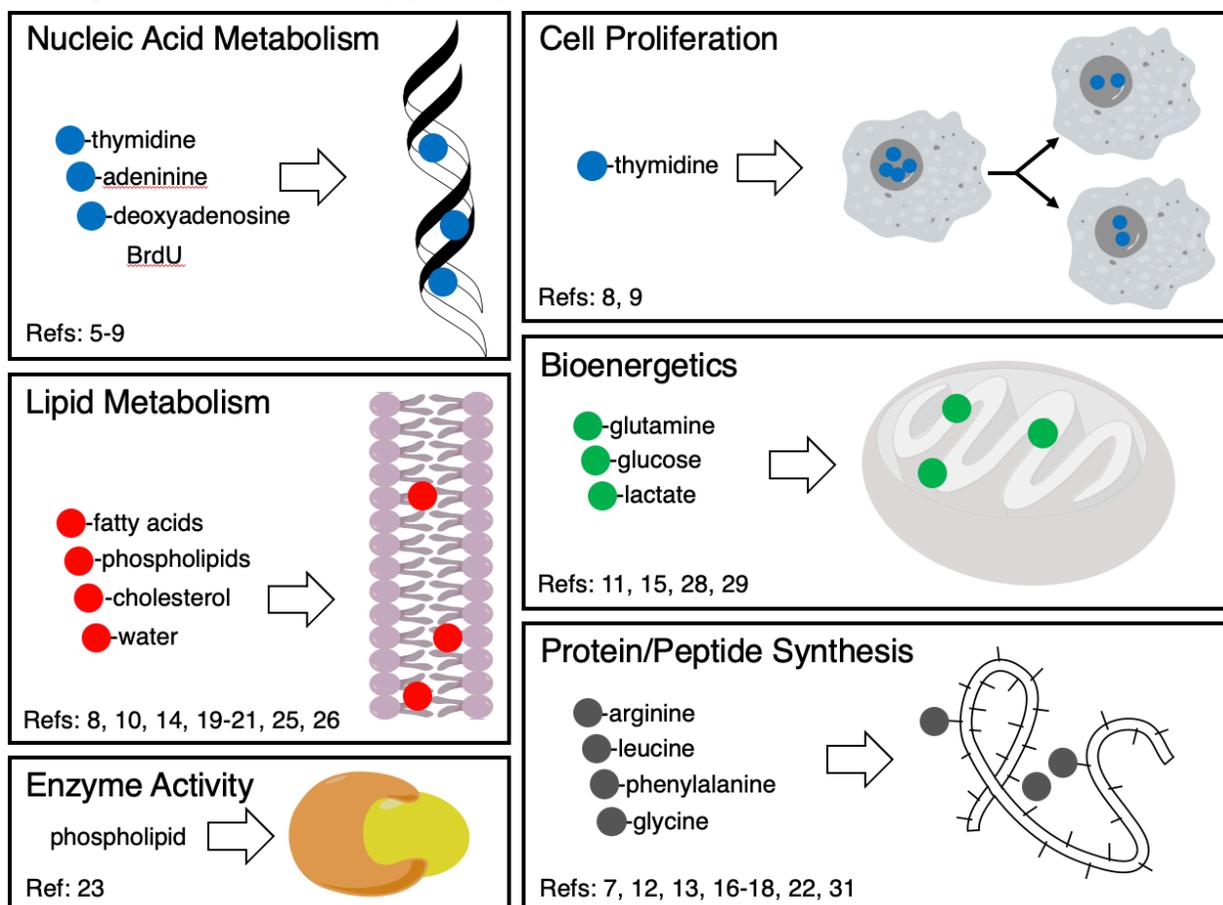
hexokinase activity. FDG-PET has long been cited as a commonly utilized clinical functional imaging method in the brain and other tissues. However, with the advances in technology, functional imaging may be more broadly defined as any modality that yields data describing the spatial distribution of biological activity. This activity would include molecular uptake/efflux, metabolic transformation, changes in electrical potential and/or physical displacement and all add a time dependent factor to the image data generated. By this broadest definition, functional imaging is not restricted to noninvasive modalities. One example of this is autoradiography, a long-standing method used to detect uptake and distribution of radiolabeled substrates in ex vivo tissue samples with the advantage of providing functional data at higher spatial resolutions (microns) than the noninvasive approaches listed above (millimeters).<sup>4</sup> Autoradiography provides functional information using radioactively labeled substrates. Indeed, most forms of functional MSI highlighted in this perspective use both radioactive and stable-isotope labels (SIL) to track biological activity, but in many cases has the advantage of monitoring many molecules within a specified mass range. Depending on the technique used, many product molecules can be monitored simultaneously in a single recorded mass spectrum. Autoradiography and MSI also both primarily analyze ex vivo tissue samples which is an advantage in many cases as adjacent tissue sections can

**Table 5.1** Spatial resolutions for highlighted fMSI methods

<b>MSI Method</b>	<b>Spatial Resolution</b>	<b>Reference(s)</b>
SIMS	35-100 nm	5,6
MIMS	30-35 nm	7,8,9,10,11,12,13
Nano-SIMS	50 nm	14,15,16,17,18
TOF-SIMS	4 $\mu$ m	19
MALDI	10-120 $\mu$ m	18
NIMS	50 $\mu$ m	20
DESI	75 $\mu$ m	21
MALDESI	50-150 $\mu$ m	22

also be examined by classic histopathology methods to correlate function to cell markers or structure. Some ionization methods can be used to collect MSI data at spatial resolutions equivalent to or better than autoradiography (see **Table 5.1**) yielding subcellular measures of function. The aim of this perspective is to highlight examples where MSI experiments have been designed to

### Biological Processes Mapped by fMSI



**Figure 5.1.** Functional processes mapped by MSI methods. The blue, red, green, and gray circles represent isotope labels.

provide functional information mainly on biomedical tissue samples. The biological processes detected by MSI methods, and covered in this communication, are summarized in **Figure 5.1**.

What types of studies can be characterized as functional MSI? The clearest example is the work of Hamilton et al.,<sup>23</sup> who first coined the term functional MSI (fMSI) to describe a method to chart the spatial location of enzyme activity in excised tissue. This study used a matrix-assisted

laser desorption ionization (MALDI) MSI method to detect phospholipase activity after applying unlabeled and SIL substrates to a tissue section for a set time to yield images with 50  $\mu\text{m}$  spatial resolution. They observed that the transformation of substrates to products detected by MSI was unevenly distributed across the tissue during the exposure time demonstrating different rates of enzymatic transformation. This novel functional mapping technique can be used to study any enzyme-catalyzed process and will be an aid to connecting molecular profiles to dynamic processes.

Although this may have been the first study to use the term fMSI, under the general definition above, there have been other relevant studies going back decades that can also be characterized as fMSI. Earliest examples utilize secondary ion mass spectrometry (SIMS). SIMS imaging has exquisite sensitivity and much higher spatial resolution than the MALDI-MSI method cited above, enabling subcellular mapping of biological function (**Table 5.1**). However, the SIMS method bombards samples with a primary ion beam to generate secondary ions which often results in the generation of fragments rather than intact molecular species.<sup>24</sup> Therefore, many SIMS methods rely on detecting unique atomic or diatomic fragments provided by isotopic labeling or introduction of xenobiotics with unique elemental content. Consequently, many SIMS MSI methods track the label and cannot distinguish between the intact substrate and/or downstream metabolites. The examples highlighted in this perspective are generally grouped by the functional pathway targeted in the studies, showing the breadth of capabilities already in place within the field.

#### **5.4 Nucleic Acid Metabolism**

Both PET and autoradiography are prime examples of the use of isotopic labels to provide functional data, and both require the use of radioactive probes but yield images with very different

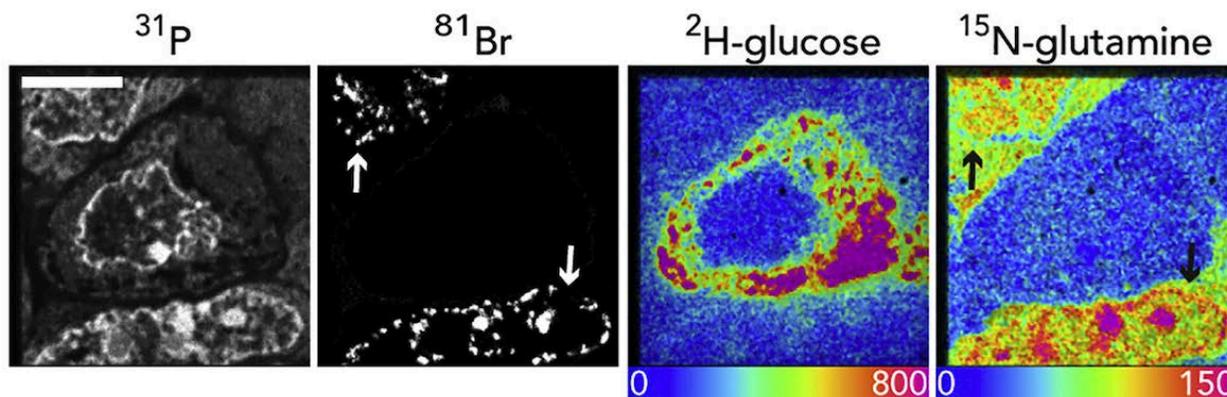
spatial resolution. The spatial resolution of SIMS can exceed that of autoradiography and some of the earliest SIMS-based applications also use radioactive probes. For example, the subcellular distribution of  $^{14}\text{C}$ -deoxyadenosine in cultured fibroblasts was detected with a 35 nm spatial resolution.<sup>5</sup> Due to the ionization method, only  $^{14}\text{C}$ - and  $^{14}\text{C}^{14}\text{N}$ - ion fragments were detected rather than intact  $^{14}\text{C}$ -deoxyadenosine or its metabolites. These ion fragments mapped to within the cytoplasm, nucleus and nucleoli. This fragment distribution reflects the uptake and transit through the cells and possibly the transformation of deoxyadenosine into DNA or other metabolic products that occurred during the 24-hour exposure time to the substrate. This same work also demonstrated SIMS detection of the uptake of stable isotopes in cells exposed to  $^{15}\text{N}$ adenine and  $^2\text{H}$ -arginine.<sup>6</sup> In addition, the demonstrated ability to detect stable isotope labels distinguishes the MSI methods from autoradiography and opens the door to the use of a far greater range of stable isotope labeled substrates and additional pathways that can be targeted for functional studies. Later improvements in technology enabled SIMS-based MSI to track several probes, both radioactive and stable isotopes, in a single sample by a method termed multiple-isotope mass spectrometry (MIMS).<sup>7</sup> The mapping capability of MIMS as a functional imaging method was noted to be "...analogous to positron emission tomography, but at the nano-scale..."<sup>25</sup>

The ability of MIMS to track multiple functional pathways was nicely demonstrated in studies to track both DNA and RNA metabolism in tissue.<sup>7</sup> In this study, a well-known nucleoside analog bromodeoxyuridine (BrdU), was used. BrdU has long been used to study cellular proliferation by tracking its incorporation into the newly synthesized DNA using immunohistochemistry. Antibody binding that is key to immunochemical detection can be plagued by problems with antigen access or variations in binding affinity. SIMS can be used to track the bromine atom in BrdU thereby circumventing these problems. In this study, following exposure

of tissue to BrdU and  $^{15}\text{N}$ -uridine, detection of  $^{81}\text{Br}$ - and  $^{12}\text{C}^{15}\text{N}$ - ions via MIMS MSI likely represent regions of DNA and RNA synthesis, respectively.<sup>7</sup> The results show  $^{81}\text{Br}$  ions in the nuclear envelope and nucleoli while the  $^{13}\text{C}^{15}\text{N}$ - signals were strongest within the nucleoli and along the nuclear envelope.<sup>7</sup> Detection of subcellular accumulation of labels were obtained from data down to 33 nm spatial resolution.

Similarly,  $^{15}\text{N}$ -thymidine administered in a pulse-chase protocol with BrdU was used to track stem cell division and metabolism in mouse and *Drosophila* intestinal tissue and fibroblasts.<sup>8</sup> This same study also emphasized another important advantage of using stable isotopes; they can be safely administered to humans as demonstrated by studies infusing  $^{15}\text{N}$ -thymidine to study label uptake into white blood cells. MIMS imaging using  $^{15}\text{N}$ -thymidine was also used to monitor the differences of DNA synthesis rates in cardiomyocytes in newborn, young and old mice to determine their source of renewal during normal aging and after induced infarction.<sup>9</sup>

The power of MIMS MSI to detect multiple nuclei to link cellular proliferation, indicated by BrdU uptake, to bioenergetics (see below) is well-illustrated in the study of Zhang et al.,<sup>11</sup> (Figure 5.2). The  $^{31}\text{P}$  MIMS MSI data on the left of Figure 5.2 shows high levels of phosphorus



**Figure 5.2.** MIMS imaging data showing the heterogeneity in tumor cells showing regions of enhanced BrdU uptake (arrows) in two nuclei which correlated to cells displaying a preference for  $^{15}\text{N}$ -glutamine uptake rather than  $^2\text{H}$ -glucose. Figure adapted from Zhang et al., (2020) “Imaging mass spectrometry reveals tumor metabolic heterogeneity,” (Licensed under CC BY 4.0).

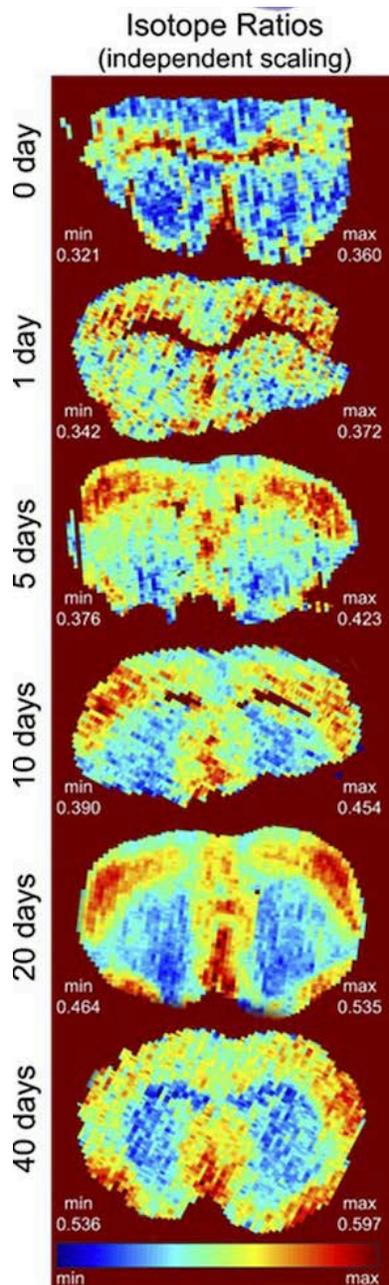
that delineates nuclei of three tumor cells in the field of view. Lower levels of  $^{31}\text{P}$  surrounding the nuclei outline the rest of the cell structures. The  $^{81}\text{Br}$  data shows enhanced BrdU uptake in two of the three nuclei (arrows) likely indicating active DNA synthesis in these cells. Differential uptake of  $^2\text{H}$ -glucose and  $^{15}\text{N}$ -glutamine in these cells shown in the next two panels is discussed below. These data illustrate the power of MIMS and similar ionization methods to detect subcellular functional heterogeneity.

### 5.5 Lipid Metabolism

Fatty acid uptake into adipocytes was tracked by treating cells with  $^{13}\text{C}$ -oleate for 20 minutes and detecting the  $^{13}\text{C}$ - and  $^{13}\text{C}^{14}\text{N}$ - ions outside and inside the cell and within lipid droplets with a spatial resolution of 35-50 nm.<sup>10</sup> Steinhäuser et al.,<sup>8</sup> showed that by tracking the isotope label in  $^{13}\text{C}$ -palmitate into *Drosophila* intestinal cells over multiple timepoints of label administration and washout, a more detailed kinetic curve can highlight differences in label turnover in subcellular locations. Time-dependent data was also collected by treating mice with  $^{13}\text{C}$ - or  $^2\text{H}$ -labeled fatty acids for periods ranging from five minutes to four days to detect label uptake into adipose, heart and brain tissue at 50 nm resolution using a nanoscale SIMS (NanoSIMS) instrument.<sup>14</sup> The ionization and detection technologies used in time-of-flight SIMS (TOF-SIMS) can detect larger molecular fragments (**Table 5.1**). In many cases, this allows for the discrimination between the substrate and downstream metabolites and enables use of different substrates containing the same isotope label. For example, TOF-SIMS can track the fate of both deuterated  $^2\text{H}$ -phosphatidylcholine and  $^2\text{H}$ -phosphatidylethanolamine in PC12 cells after 19 h of incubation with these substrates. The spatial resolution of 4  $\mu\text{m}$ , while not enabling subcellular analysis, still allowed single cell mapping of both species.<sup>19</sup> Similarly, the fate of deuterated  $\alpha$ -linolenic acid and linoleic acid and turnover of PC12 membrane lipids was mapped with TOF-

SIMS imaging with 6  $\mu\text{m}$  resolution.<sup>26</sup> Softer ionization methods used with MSI can detect much larger intact molecules. For example, the dynamics of plaque formation was the subject of a MALDI-MSI study of the uptake and distribution of  $^2\text{H}_6$ -cholesterol in mouse aorta.<sup>27</sup> When used with high resolving power mass spectrometers, MALDI can readily distinguish between the intact labeled-cholesterol substrate and labeled cholesterol ester in the plaques. Mice were maintained for three weeks on daily oral doses of  $^2\text{H}_6$ -cholesterol and tissue from three groups of mice were obtained weekly over an additional three weeks enabling acquisition of kinetic data from tissue with a spatial resolution of 50  $\mu\text{m}$ . Nanostructure-initiator mass spectrometry (NIMS) combined with MSI can map intact isotope-labeled molecules and was used to show the variability in lipid metabolism in tumor tissue with a spatial resolution of 50  $\mu\text{m}$ .<sup>20</sup> In that study, rather than a specifically labeled lipid substrate, mice were administered drinking water enriched with deuterium for five days. This approach provides a measure of the rate of synthesis of many different lipids in the tumor and requires a soft ionization technique to distinguish the various deuterium-labeled phospholipids produced. The authors of that study named this method 'kinetic MSI' (kMSI) which is an apt description of the process and a synonym for fMSI. Of note, the authors emphasized that regions showing enhanced lipid synthesis did not correspond to regions of high steady-state levels, highlighting the ability of kMSI and fMSI methods to monitor dynamic biological processes that are undetectable by conventional MSI. They also detected subpopulations of tumor cells with enhanced lipid synthesis and correlated functional measures with histopathological markers, an advantage to using *ex vivo* tissue sections as previously stated.

In a similar study, lipid dynamics were studied in mice administered deuterated water for 40 days and used desorption electrospray ionization (DESI) imaging to study the resulting deuterated lipids<sup>21</sup> (**Figure 5.3**). The data in **Figure 5.3** show the changes in deuterated



**Figure 5.3.** Isotope ratio images of phosphatidylserine in the brain tissue of mice administered deuterated drinking water over 40 days. Figure adapted from Carson et al., (2017) “Imaging Regiospecific lipid turnover in mouse brain with desorption electrospray ionization mass spectrometry,” (Licensed under CC BY 4.0).

phosphatidylserine in mouse brain over 40 days. In total, the enrichment rate of four different lipid species in different areas of the brain with 75  $\mu\text{m}$  spatial resolution was obtained by sampling anatomically similar tissue over the time course of the study. These two studies illustrate how deuterated water, relatively inexpensive single isotope-labeled substrate that is easy to administer, combined with the capability of ionization methods such as NIMS and DESI, can distinguish and track the kinetics of multiple molecular species.

## 5.6 Bioenergetics

The amino acid glutamine is often used in cancer cells as a fuel and feeds into multiple metabolic pathways in addition to its incorporation into proteins and peptides. Subcellular detection using the NanoSIMS technology at 50 nm resolution of  $^{12}\text{C}^{15}\text{N}$ - ions in MCF7 cells after 16 h of incubation in  $^{15}\text{N}$ -glutamine medium showed variations in the ion distribution to different cellular organelles.<sup>15</sup> Since glutamine can be rapidly metabolized in cancer cells, solely mapping isotope-containing fragments limits the functional information that can be provided. However, even with such rapidly metabolized species such as  $^2\text{H}$ -glucose and  $^{15}\text{N}$ -glutamine, important data on functional heterogeneity at the single cell level is still accessible<sup>11</sup> (**Figure 5.2**). As noted above, the MIMS MSI data in **Figure 5.2** shows two of three tumor cells actively incorporated of BrdU. These two cells also show high levels of  $^{15}\text{N}$ -glutamine uptake whereas the low BrdU cells show prefer  $^2\text{H}$ -glucose as a substrate. The authors of this study noted a correlation between the amount of functional heterogeneity detectable in a tissue sample and therapy resistance.<sup>11</sup> These results are consistent with many FDG-PET studies which indicate that the degree of functional metabolic heterogeneity in tumors provides predictive clinical data.<sup>28</sup> However, the single-cell and subcellular heterogeneity detected by MIMS could not be revealed by methods such as PET with

much poorer spatial resolution demonstrating the significant advantages of multiscale data provided by fMSI.

At an even lower spatial resolution, MALDI analysis was used to detect regional differences in metabolism of  $^{13}\text{C}$ -glucose activity in mouse brain.<sup>29</sup> This study demonstrated the ability of MALDI to detect flux through a range of glycolytic pathways that occurred in 15 min after intraperitoneal injection of labeled glucose. Regional differences in the flux of glucose to downstream metabolites in the brain tissue, measured with a spatial resolution of 30 or 120  $\mu\text{m}$  included highly labile metabolites that were likely detectable due to the novel microwave tissue fixation process used. The same group studied labeled  $^{13}\text{C}$ -glucose and  $^{13}\text{C}$ -lactate metabolism in a mouse model of cardiac ischemia.<sup>30</sup> In that study, tissue sections were obtained from the cardiac base to apex to generate a unique three-dimensional functional map in the tissue and showed the interrelationship of different functional pathways in various regions of the heart.

### **5.7 Amino Acid, Peptide and Protein Metabolism**

Differential rates of protein renewal in a mouse cochlea were inferred from SIMS mapping of  $^{12}\text{C}^{15}\text{N}$ - ions after incorporating  $^{15}\text{N}$ -leucine into the diet for nine days.<sup>7</sup> Varying enrichment of  $^{15}\text{N}$  ions were measured between cell types and within subpopulations of the same cells demonstrating the power of the technique to yield kinetic information on leucine metabolism/protein renewal in different cells/tissue regions.

In experiments designed to provide kinetic information on protein turnover, stable isotope labeling kinetics (SILK) was combined with SIMS mapping (SILK-SIMS) in the study of mouse models of Alzheimer's disease and in human brain tissue.<sup>16</sup> Administration of  $^{13}\text{C}$ -leucine to tissue with or without a period of label washout prior to tissue sampling enabled assessment of  $^{13}\text{C}^{14}\text{N}$ -ion distribution in brain plaques using NanoSIMS. The results showed label incorporation activity

was highly variable even within individual plaques. A similar study also manipulated the length of the label administration and washout periods (pulse-chase) to obtain more detailed kinetics on protein turnover.<sup>12</sup> The authors maintained newborn mice on an <sup>15</sup>N-enriched diet for 21 or 45 days and harvested tissue when mice were 6, 18 or 26 months old. The liver had mixtures of cells with different lifespans, including some showing no replicative activity.

One drawback is that nearly all MSI methods require *ex vivo* tissue sampling and therefore provide only one timepoint of a functional process. The studies noted above, use creative changes of exposure and/or washout periods in pulse-chase experiments to glean more kinetic data from these single tissue samples. In line with classic kinetic isotope tracing studies, multiple tissue samples attained over a time course can yield more detailed kinetic information on protein turnover. This is possible when similar anatomical sites can be harvested over time; for example, the study of lipid dynamics on multiple brain samples shown in **Figure 5.3**. For protein dynamics, one such study of mice fed <sup>15</sup>N-leucine tracked different rates of protein synthesis within hair cells over 150 days via multiple tissue sampling.<sup>13</sup>

In addition, isotope labeling can be used to monitor functional changes induced by therapeutic intervention. For example, a diet containing <sup>15</sup>N-leucine was fed to mice at the start of a 28-day treatment with erythropoietin.<sup>17</sup> Examination of brain tissue detected increased levels of <sup>12</sup>C<sup>15</sup>N- ions in non-dividing pyramidal cells indicative of EPO-induced protein synthesis. This demonstrated timed isotope delivery combined with MSI analysis can yield unique kinetic information on molecular and cellular turnover in response to treatment.

Rather than starting at the amino acid level, <sup>15</sup>N-enriched proteins can be delivered to cells and uptake can be observed over time.<sup>31</sup> This study noted the uptake of two different proteins into the cell nucleus after 4 h and then transit of the proteins over the next few hours into the cytoplasm.

Clearly, the advantage of using SIMS with high spatial resolution (50 nm) was necessary to detect intracellular transit.

In order to obtain more molecular data from amino acid feeding studies, albeit with lower spatial resolution, MALDI MSI was used to track the fate of intravenously injected  $^{13}\text{C}$ -phenylalanine in mouse liver over a time course of 10, 30 and 60 min following injection.<sup>32</sup> The authors noted that pretreatment of the tissue sections with an amino acid derivatizing agent was required to increase ionization efficiency in order to detect uptake of  $^{13}\text{C}$ -phenylalanine and its transformation into  $^{13}\text{C}$ -tyrosine. This was performed at a 25  $\mu\text{m}$  spatial resolution. A variation on acquiring tissue samples over a time course was to change the time for tissue sampling in an isotope pulse-chase experiment by a method called 'imaging stable isotope labeling kinetics' (iSILK)<sup>18</sup> in a similar way to the SILK-SIMS study mentioned earlier. In fact, these authors used both MALDI and NanoSIMS alongside several other imaging methods to detect plaque formation in brain tissue. In this study, knock-in mice susceptible to Alzheimer's-like plaque formation, were fed a  $^{15}\text{N}$ -enriched diet over 4 or 11 weeks followed by variable chase (washout) periods. The MALDI MSI data showed distinct labeling dynamics in intact proteins in amyloid plaques at 10  $\mu\text{m}$  spatial resolution while NanoSIMS detailed the heterogeneous core structure within each plaque with a 50 nm spatial resolution.

Time-dependent or serial sampling is advantageous to detecting the trends in the kinetic data that can be affected by differential rates of substrate delivery to the tissue (i.e., perfusion). Serial sampling of tissue may not always be possible, especially in clinical studies. In addition, each tissue sample is unique so detecting functional activity in an identical spatial location by MSI in multiple samples is not possible. For some studies, serial sampling and examination of

anatomically similar sites may suffice<sup>8, 13, 32</sup> but in many other cases, it is desirable to examine each unique spatial location for metabolic activity.

To address this issue, we have adapted a timed substrate infusion method, first described by Dudley et al.,<sup>33</sup> and later coined Multiple Infusion Start Times (MIST) by Paolini et al.,<sup>34</sup> that has been used in mass spectrometry analysis to provide detailed labeling kinetic data from a single tissue sample.<sup>33-36</sup> Although not previously used with MSI, the MIST protocol is ideal for use in fMSI studies as a single tissue sample provides labeling data from multiple time points. Our study of mouse liver (detailed in subsequent **Chapter 6**) demonstrated that matrix-assisted laser desorption electrospray ionization (MALDESI) MSI can distinguish both the glycine isotopologue substrates and the glutathione products of metabolism.<sup>22</sup> This method may likely be adapted to any system so long as at least three isotopologues of precursor can be synthesized if not already commercially available. This work was performed with an orbitrap instrument allowing for high resolving power measurements, key to measuring separate isotopologues. For example, in this work two infused glycine labels are separated by 0.0063 Da mass shift, requiring a resolving power of approximately 50,000 (at 308.0769  $m/z$ ) for separation after enrichment of the glutathione molecules. The introduction of the orbitrap benchtop mass spectrometers is important to note in this context as it provided accessibility to such high resolving power measurements.

## **5.8 Drug Uptake and Metabolism**

Along with the use of SIMS, MALDI and other MSI technologies to track biological activity using metabolic substrates, studies using MSI to map uptake of xenobiotics or therapeutics can be classified as a form of fMSI. This is an active area of research and can obtain both pharmacokinetic and pharmacodynamic data and has been covered in recent reviews.<sup>37-39</sup>

## 5.9 Conclusions

The advances in analytical technologies in biomedical research have generated reams of data on molecular content in healthy and pathologic tissue. However, every tissue specimen is comprised of multiple cell types each with a unique molecular profile that can be catalogued using recent advances in single-cell analytical methods. Importantly, tissue function relies on the organized dynamic interactions between these cell types and tissue microenvironments and has spurred the development of spatial molecular profiling methods.<sup>40</sup> Many of these newly developed methods supply location-specific molecular data from length scales ranging from subcellular structures to multicellular units to whole tissues. MSI is a major contributor to providing spatially-resolved data on these scales with an unparalleled ability to provide highly detailed molecular profiles but still lack of information relating this content to function. This perspective has demonstrated that MSI methods have already been adapted to provide multi-scale functional data that can assist in connecting tissue organization to phenotype and can be a key player in interpreting spatial biology data. With strategic selection of isotope labels, combined with clever infusion protocols, these fMSI technologies can report on biological processes that occur on timeframes of minutes to months at these different length scales. There is no other technique that can provide equivalent data.

Isotope labeling is a key component to studying metabolic and cellular dynamics and, as noted above, stable isotope labeling can safely be used in human studies.<sup>41-43</sup> The multiscale capability of fMSI technologies can provide measures of dynamic heterogeneity in single cell or cell aggregates in biopsy specimens that may enhance diagnostic information gleaned from a single sample.<sup>28,44</sup> In addition, increased metabolic flux, rather than changes in steady-state levels can be indicative of tissue pathologies such as breast cancer.<sup>45</sup> The fMSI technique is positioned to

provide a map of flux heterogeneity in clinical specimens that can detect high-flux subpopulations of cells that would likely be missed with standard tissue homogenization approaches. Therefore, another contribution of fMSI can be the addition of functional information to clinical samples.

The cited examples above just scratch the surface of the biochemical and cellular networks amenable to study by fMSI methods and possible translational uses of this method. We hope that by highlighting some of the studies, more investigators develop these technologies to study a wider array of biological processes.

## 5.10 References

- (1) Buchberger, A. R.; DeLaney, K.; Johnson, J.; Li, L. Mass spectrometry imaging: A review of emerging advancements and future insights. *Analytical Chemistry* **2018**, *90* (1), 240-265.
- (2) Hochachka, P. W.; Matheson, G. O. Regulating atp turnover rates over broad dynamic work ranges in skeletal muscles. *J Appl Physiol* (1985) **1992**, *73* (5), 1697-1703.
- (3) Niklas, J.; Heinzle, E. Metabolic flux analysis in systems biology of mammalian cells. *Adv. Biochem. Engin. Biotechnol.* **2012**, *127*, 109-132.
- (4) Alitalo, O.; Rantamäki, T.; Huhtala, T. Digital autoradiography for efficient functional imaging without anesthesia in experimental animals: Reversing phencyclidine-induced functional alterations using clozapine. *Prog Neuropsychopharmacol Biol Psychiatry* **2020**, *100*, 109887.
- (5) Hindie, E.; Hallégot, P.; Chabala, J. M.; Thorne, N. A.; Coulomb, B.; Levi-Setti, R.; Galle, P. Ion microscopy: A new approach for subcellular localization of labelled molecules. *Scanning Microsc* **1988**, *2* (4), 1821-1829.
- (6) Hindie, E.; Coulomb, B.; Beaupain, R.; Galle, P. Mapping the cellular distribution of labelled molecules by sims microscopy. *Biol Cell* **1992**, *74* (1), 81-88.
- (7) Lechene, C.; Hillion, F.; McMahon, G.; Benson, D.; Kleinfeld, A. M.; Kampf, J. P.; Distel, D.; Luyten, Y.; Bonventre, J.; Hentschel, D.; et al. High-resolution quantitative imaging of mammalian and bacterial cells using stable isotope mass spectrometry. *J. Biology* **2006**, *5*, 20.21-20.30.

- (8) Steinhauser, M. L.; Bailey, A. P.; Senyo, S. E.; Guillermier, C.; Perlstein, T. S.; Gould, A. P.; Lee, R. T.; Lechene, C. Multi-isotope imaging mass spectrometry quantifies stem cell division and metabolism. *Nature* **2012**, *481*, 516-520.
- (9) Senyo, S. E.; Steinhauser, M. L.; Pizzimenti, C. L.; Yang, V. K.; Cai, L.; Wang, M.; Wu, T. D.; Guerin-Kern, J. L.; Lechene, C. P.; Lee, R. T. Mammalian heart renewal by pre-existing cardiomyocytes. *Nature* **2013**, *493* (7432), 433-436.
- (10) Kleinfeld, A. M.; Kampf, J. P.; Lechene, C. Transport of <sup>13</sup>C-oleate in adipocytes measured using multi imaging mass spectrometry. *J Am Soc Mass Spectrom* **2004**, *15* (11), 1572-1580.
- (11) Zhang, Y.; Guillermier, C.; De Raedt, T.; Cox, A. G.; Maertens, O.; Yimlamai, D.; Lun, M.; Whitney, A.; Maas, R. L.; Goessling, W.; et al. Imaging mass spectrometry reveals tumor metabolic heterogeneity. *iScience* **2020**, *23* (8), 101355.
- (12) Arrojo e Drigo, R.; Lev-Ram, V.; Tyagi, S.; Ramachandra, R.; Deerinck, T.; Bushong, E.; Phan, S.; Orphan, V.; Lechene, C.; Ellisman, M. H.; et al. Age mosaicism across multiple scales in adult tissues. *Cell Metab* **2019**, *30* (2), 343-351.e343.
- (13) Zhang, D. S.; Piazza, V.; Perrin, B. J.; Rzdzińska, A. K.; Poczatek, J. C.; Wang, M.; Prosser, H. M.; Ervasti, J. M.; Corey, D. P.; Lechene, C. Multi-isotope imaging mass spectrometry reveals slow protein turnover in hair-cell stereocilia. *Nature* **2012**, *481*, 520-525.
- (14) Jiang, H.; Goulbourne, C. N.; Tatar, A.; Turlo, K.; Wu, D.; Beigneux, A. P.; Grovenor, C. R.; Fong, L. G.; Young, S. G. High-resolution imaging of dietary lipids in cells and tissues by nanosims analysis. *J Lipid Res* **2014**, *55* (10), 2156-2166.

- (15) Jiang, H.; Favaro, E.; Goulbourne, C. N.; Rakowska, P. D.; Hughes, G. M.; Ryadnov, M. G.; Young, S. G.; Ferguson, D. J. P.; Harris, A. L.; Grovenor, C. R. M. Stable isotope imaging of biological samples with high resolution secondary ion mass spectrometry and complementary techniques. *Methods* **2014**, *68* (2), 317-324.
- (16) Wildburger, N. C.; Gyngard, F.; Guillermier, C.; Patterson, B. W.; Elbert, D.; Mawuenyega, K. G.; Schneider, T.; Green, K.; Roth, R.; Schmidt, R. E.; et al. Amyloid- $\beta$  plaques in clinical alzheimer's disease brain incorporate stable isotope tracer in vivo and exhibit nanoscale heterogeneity. *Front Neurol* **2018**, *9*, 169.
- (17) Hassouna, I.; Ott, C.; Wüstefeld, L.; Offen, N.; Neher, R. A.; Mitkovski, M.; Winkler, D.; Sperling, S.; Fries, L.; Goebbels, S.; et al. Revisiting adult neurogenesis and the role of erythropoietin for neuronal and oligodendroglial differentiation in the hippocampus. *Mol Psychiatry* **2016**, *21* (12), 1752-1767.
- (18) Michno, W.; Stringer, K. M.; Enzlein, T.; Passarelli, M. K.; Escrig, S.; Vitanova, K.; Wood, J.; Blennow, K.; Zetterberg, H.; Meibom, A.; et al. Following spatial a $\beta$  aggregation dynamics in evolving alzheimer's disease pathology by imaging stable isotope labeling kinetics. *Sci Adv* **2021**, *7* (25).
- (19) Lanekoff, I.; Sjövall, P.; Ewing, A. G. Relative quantification of phospholipid accumulation in the pc12 cell plasma membrane following phospholipid incubation using tof-sims imaging. *Analytical Chemistry* **2011**, *83* (13), 5337-5343.
- (20) Louie, K. B.; Bowen, B. P.; McAlhany, S.; Huang, Y.; Price, J. C.; Mao, J. H.; Hellerstein, M.; Northen, T. R. Mass spectrometry imaging for in situ kinetic histochemistry. *Sci Rep* **2013**, *3*, 1656.

- (21) Carson, R. H.; Lewis, C. R.; Erickson, M. N.; Zagieboylo, A. P.; Naylor, B. C.; Li, K. W.; Farnsworth, P. B.; Price, J. C. Imaging regiospecific lipid turnover in mouse brain with desorption electrospray ionization mass spectrometry. *J Lipid Res* **2017**, *58* (9), 1884-1892.
- (22) Mellinger, A. L.; Garrard, K. P.; Khodjanizyazova, S.; Rabbani, Z. N.; Gamcsik, M. P.; Muddiman, D. C. Multiple infusion start time mass spectrometry imaging of dynamic sil-glutathione biosynthesis using infrared matrix-assisted laser desorption electrospray ionization. *J Proteome Res* **2022**, *21* (3), 747-757.
- (23) Hamilton, B. R.; Marshall, D. L.; Casewell, N. R.; Harrison, R. A.; Blanksby, S. J.; Undheim, E. A. B. Mapping enzyme activity on tissue by functional mass spectrometry imaging. *Angew Chem Int Ed Engl* **2020**, *59* (10), 3855-3858.
- (24) Muddiman, D. C.; Gusev, A. I.; Hercules, D. M. Application of secondary ion and matrix-assisted laser desorption-ionization time-of-flight mass spectrometry for the quantitative analysis of biological molecules. *Mass Spec. Rev.* **1995**, *14*, 383-429.
- (25) Narendra, D. P.; Steinhauser, M. L. Metabolic analysis at the nanoscale with multi-isotope imaging mass spectrometry (mims). *Curr Protoc Cell Biol* **2020**, *88* (1), e111.
- (26) Philipsen, M. H.; Sämfors, S.; Malmberg, P.; Ewing, A. G. Relative quantification of deuterated omega-3 and -6 fatty acids and their lipid turnover in pc12 cell membranes using tof-sims. *J Lipid Res* **2018**, *59* (11), 2098-2107.
- (27) Castro-Perez, J.; Hatcher, N.; Kofi Karikari, N.; Wang, S. P.; Mendoza, V.; Shion, H.; Millar, A.; Shockcor, J.; Towers, M.; McLaren, D.; et al. In vivo isotopically labeled atherosclerotic aorta plaques in apoe ko mice and molecular profiling by matrix-assisted laser desorption/ionization mass spectrometric imaging. *Rapid Commun Mass Spectrom* **2014**, *28* (22), 2471-2479.

- (28) Soussan, M.; Orhac, F.; Boubaya, M.; Zelek, L.; Ziol, M.; Eder, V.; Buvat, I. Relationship between tumor heterogeneity measured on fdg-pet/ct and pathological prognostic factors in invasive breast cancer. *PLoS ONE* **2014**, *9* (4), e94017.
- (29) Sugiura, Y.; Honda, K.; Kajimura, M.; Suematsu, M. Visualization and quantification of cerebral metabolic fluxes of glucose in awake mice. *Proteomics* **2014**, *14* (7-8), 829-838.
- (30) Sugiura, Y.; Katsumata, Y.; Sano, M.; Honda, K.; Kajimura, M.; Fukuda, K.; Suematsu, M. Visualization of in vivo metabolic flows reveals accelerated utilization of glucose and lactate in penumbra of ischemic heart. *Sci Rep* **2016**, *6*, 32361.
- (31) Delaune, A.; Cabin-Flaman, A.; Legent, G.; Gibouin, D.; Smet-Nocca, C.; Lefebvre, F.; Benecke, A.; Vasse, M.; Ripoll, C. 50nm-scale localization of single unmodified, isotopically enriched, proteins in cells. *PLoS ONE* **2013**, *8* (2), e56559.
- (32) Arts, M.; Soons, Z.; Ellis, S. R.; Pierzchalski, K. A.; Balluff, B.; Eijkel, G. B.; Dubois, L. J.; Lieuwes, N. G.; Agten, S. M.; Hackeng, T. M.; et al. Detection of localized hepatocellular amino acid kinetics by using mass spectrometry imaging of stable isotopes. *Angew Chem Int Ed Engl* **2017**, *56* (25), 7146-7150.
- (33) Dudley, M. A.; Burrin, D. G.; Wykes, L. J.; Toffolo, G.; Cobelli, C.; Nichols, B. L.; Rosenberger, J.; Jahoor, F.; Reeds, P. J. Protein kinetics determined in vivo with a multiple-tracer single-sample protocol: Application to lactase synthesis. *Am. J. Physiol.* **1998**, *274* (Gastrointest. Liver Physiol.), G591-G598.
- (34) Paolini, C. L.; Marconi, A. M.; Pike, A. W.; Fennessey, P. V.; Pardi, G.; Battaglia, F. C. A multiple infusion start time (mist) protocol for stable isotope studies of fetal blood. *Placenta* **2001**, *22*, 171-176.

- (35) Agostini, F.; Dalla Libera, L.; Rittweger, J.; Mazzucco, S.; Jurdana, M.; Mekjavic, I. B.; Pisot, R.; Gorza, L.; Narici, M.; Biolo, G. Effects of inactivity on human muscle glutathione synthesis by doubletracer and single-biopsy approach. *J. Physiol.* **2010**, *588*, 5089-5104.
- (36) Zhang, X. J.; Chinkes, D. L.; Wolfe, R. R. Measurement of muscle protein fractional synthesis and breakdown rates from a pulse tracer injection. *Am J Physiol Endocrinol Metab* **2002**, *283* (4), E753- 764.
- (37) Goodwin, R. J. A.; Takats, Z.; Bunch, J. A critical and concise review of mass spectrometry applied to imaging in drug discovery. *SLAS Discov* **2020**, *25* (9), 963-976.
- (38) Noun, M.; Akoumeh, R.; Abbas, I. Cell and tissue imaging by tof-sims and maldi-tof: An overview for biological and pharmaceutical analysis. *Microsc Microanal* **2021**, 1-26.
- (39) Nishidate, M.; Hayashi, M.; Aikawa, H.; Tanaka, K.; Nakada, N.; Miura, S. I.; Ryu, S.; Higashi, T.; Ikarashi, Y.; Fujiwara, Y.; et al. Applications of maldi mass spectrometry imaging for pharmacokinetic studies during drug development. *Drug Metab Pharmacokinet* **2019**, *34* (4), 209-216.
- (40) Palla, G.; Fischer, D. S.; Regev, A.; Theis, F. J. Spatial components of molecular tissue biology. *Nat Biotechnol* **2022**, *40* (3), 308-318.
- (41) Fan, T. W.; Lane, A. N.; Higashi, R. M.; Farag, M. A.; Gao, H.; Bousamra, M.; Miller, D. M. Altered regulation of metabolic pathways in human lung cancer discerned by (13)c stable isotoperesolved metabolomics (sirm). *Mol Cancer* **2009**, *8*, 41.
- (42) Paterson, R. W.; Gabelle, A.; Lucey, B. P.; Barthélemy, N. R.; Leckey, C. A.; Hirtz, C.; Lehmann, S.; Sato, C.; Patterson, B. W.; West, T.; et al. Silk studies - capturing the turnover of proteins linked to neurodegenerative diseases. *Nat Rev Neurol* **2019**, *15* (7), 419-427.

- (43) Pichumani, K.; Mashimo, T.; Vemireddy, V.; Ijare, O. B.; Mickey, B. E.; Malloy, C. R.; MarinValencia, I.; Baskin, D. S.; Bachoo, R. M.; Maher, E. A. Measurement of (13) c turnover into glutamate and glutamine pools in brain tumor patients. *FEBS Lett* **2017**, *591* (21), 3548-3554.
- (44) Alobaidli, S.; McQuaid, S.; South, C.; Prakash, V.; Evans, P.; Nisbet, A. The role of texture analysis in imaging as an outcome predictor and potential tool in radiotherapy treatment planning. *Br J Radiol* **2014**, *87* (1042).
- (45) Possemato, R.; Marks, K. M.; Shaul, Y. D.; Pacold, M. E.; Kim, D.; Birsoy, K.; Sethumadhavan, S.; Woo, H. K.; Jang, H. G.; Jha, A. K.; et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **2011**, *476*, 346-350.

## **CHAPTER 6: Multiple Infusion Start Time Mass Spectrometry Imaging of Dynamic SIL-Glutathione Biosynthesis Using Infrared Matrix-Assisted Laser Desorption Electrospray Ionization**

Reprinted with permission from Mellinger, A.L., Garrard, K.P., Khodjanizyazova, S., Rabbani, Z.N., Gamcsik, M.P., Muddiman, D.C., *Journal of Proteome Research*, **2022**, *21*, 3, 747-757.

Copyright © 2022 American Chemical Society.

### **6.1 Introduction**

Glutathione is a tripeptide, synthesized within mammals from glutamate, cysteine, and glycine amino acids. The reduced form of glutathione (GSH) is up to 100-fold more concentrated in cells at physiological conditions than the oxidized disulfide form (GSSG).<sup>1</sup> The reduced form is highly abundant in most mammalian cells and tissue types and its functional importance is due to its characteristic thiol group within the cysteine residue.<sup>1-3</sup> GSH plays critical roles in maintaining redox homeostasis within cells, performing as an essential antioxidant against reactive oxygen species, and managing the thiol status of proteins.<sup>3</sup> The dysregulation of GSH has been implicated in the progression of many human diseases, including neurodegenerative diseases and cancer.<sup>2,4-7</sup> For example, its role in oxidative stress mitigation adds particular interest to age-related neurodegenerative diseases such as Alzheimer's disease, the most common of these diseases, due to the proposed "free radical theory of aging."<sup>8-10</sup> Another area of interest is the observed elevation of GSH in proliferating tissues that often parallels increased tumor progression in many cancer types<sup>2,11</sup> as well the increased therapy resistance.<sup>1,2,7,12</sup> A better understanding of the distinct role of GSH in tumor progression and the effects of tumor microenvironments on GSH metabolism would be helpful in furthering related therapeutic approaches.<sup>1,12-14</sup> Tissue, whether healthy or pathologic, is heterogeneous and consists of various cell types and microenvironments. Most

glutathione assays utilize homogenized tissue extracts that obscure tissue subpopulations or environments that may dictate tissue function and therapy response. Therefore, there is a need for methods to study spatial heterogeneity of GSH across tissues to further define its roles in healthy tissue and different disease states.

Mass spectrometry imaging (MSI) is an excellent candidate analytical platform to measure GSH biosynthesis heterogeneity because the approach combines advantages of mass spectrometry for biomolecule characterization and relative quantification with the ability to visualize spatial relationships.<sup>15,16</sup> While a plethora of ionization sources for MSI exist, infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI) offers many key advantages to biospecimen imaging.<sup>17,18</sup> Cryosections of sample tissues are thaw-mounted and immediately sampled (*i.e.*, without homogenization), thus preserving the spatial integrity of the biological system.<sup>19</sup> The technique operates at atmospheric pressure and currently utilizes a mid-IR laser wavelength of 2970 nm which excites water at its O-H stretching bands. Therefore, a controlled ice-matrix, in addition to endogenous water, provides ample energy-absorbing matrix to facilitate the desorption of analytes.<sup>20,21</sup> Desorbed biomolecules are then post-ionized in an orthogonal electrospray plume for mass spectrometry analysis. Nazari and coworkers successfully distinguished healthy and cancerous hen ovarian tissue using both relative and absolute quantification of GSH using IR-MALDESI-MSI, adding to the attractiveness of this technique for the method described here.<sup>11,22</sup>

Metabolism is a dynamic process and measuring steady-state levels of metabolites such as glutathione reflects only a small part of the complex molecular interaction that maintain these levels. Metabolic rates reflect the summed total of molecular (genes, proteins, metabolites) and environmental contributions to tissue function. However, kinetic analysis of metabolites remains

a challenge in MSI. Unlike most biofluids which are easily sampled and regenerated by organisms, the study of solid tissue requires harvesting the tissue from an animal. Therefore, only a limited single time-point in the organism's metabolism is preserved and represented by the measured data in each tissue section, limiting the scope of conclusions that may be made.<sup>23</sup> Stable isotope labeling (SIL) technology has emerged as a useful way of time stamping metabolites to measure metabolic dynamics. With the incredible improvements in resolving power and mass accuracy of benchtop mass spectrometers, the utility of various SIL atoms within molecules has increased substantially. Commonly used examples include  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{H}$ , and  $^{18}\text{O}$ , all of which can be utilized to chemically synthesize various isotopologues of the same molecule.<sup>24,25</sup> Many organic molecules can therefore be labeled in more than one way, including GSH in this case, which contains carbon, hydrogen, nitrogen, and oxygen as potential labeling sites. One way to study kinetics is to infuse isotopically labeled precursors that can be metabolically incorporated into a product by a biological system, such as a cell culture or a living organism. This results in an "enrichment" of the metabolites with these SILs. Enrichments of the targeted products are measured by the mass spectrometer as a mass shift of the monoisotopic peak from its lighter and unlabeled native species.

To time stamp these enrichments, most often, multiple animals are injected with a SIL at the same time, referred to as a simultaneous or single infusion start time (SIST) protocol. The tissues are then harvested at separate time points.<sup>26,27</sup> These experiments often require a high number of tissue samples (animals) to achieve statistical significance,<sup>28</sup> resulting in significant inter-sample variability and considerable expense. Large quantities of instrument time are also required for the analysis of many tissues. To combat these disadvantages and to reduce the additional biological variability inherent in the SIST protocol, we adapted a multiple infusion start time (MIST) protocol that varied the time at which three different glycine isotopologues were

administered and was followed by a single tissue harvest. This produced GSH with three distinct mass labels that reflected the time course of incorporation of the isotopologues. This method allows for the monitoring of enrichment reflecting three different time points within a single tissue sample, minimizing not only the number of animals, but also reduces the analysis time and biological variability between replicates. Although the MIST protocol have been previously shown to work in tissue homogenates,<sup>29–31</sup> to our knowledge this is the first use of this method in a MSI study.

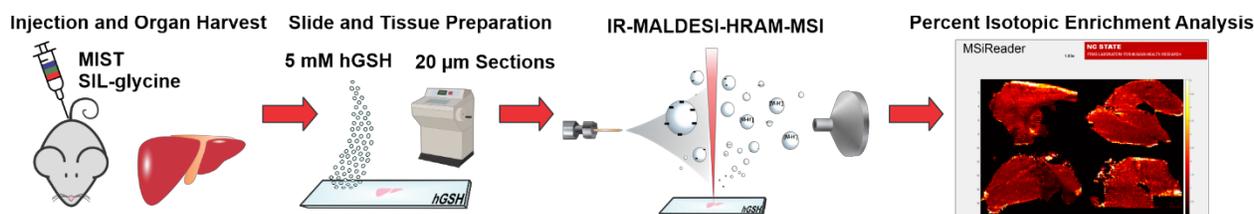
In this chapter, we describe a protocol combining MIST with IR-MALDESI-MSI. This protocol was used to successfully image heterogeneity in both steady-state (*i.e.*, single time point) GSH abundances and kinetic GSH levels across mouse liver tissue sections. We also developed the percent isotope enrichment (PIE) tool in the open-source software MSiReader<sup>19,32</sup> for spatial analysis of isotope enrichment in tissue imaged by mass spectrometry. This new tool generates enrichment heatmaps for SIL across tissues after accurate correction for overlapping isotopologues, allowing for visualization of changes in metabolic flux across tissues. This protocol may be applied for the study of spatial changes in GSH metabolism in mammalian tissue for deeper comparisons between biological conditions. The PIE tool will be included in the next release of MSiReader.<sup>33</sup>

## 6.2 Materials and Methods

### 6.2.1 Mouse Tissue Labeling and Collection

Female Balb/c mice (6-8 weeks old) were obtained from Charles River (Wilmington, MA). Mice were anesthetized (isoflurane/oxygen) and the tail veins were cannulated. The three isotopologues of glycine, [2-<sup>13</sup>C,<sup>15</sup>N]-glycine, [1,2-<sup>13</sup>C<sub>2</sub>]-glycine, and [1,2-<sup>13</sup>C<sub>2</sub>,<sup>15</sup>N]-glycine, were obtained from Cambridge Isotope Laboratories (Tewksbury, MA). Unlabeled glycine (Sigma

Chemical Co.) and the isotopologues were dissolved in phosphate-buffered saline. The isotopologues were administered to the anesthetized mice by bolus injection (0.05 mL) at doses of 300 moles/kg *via* the tail vein cannula at 0-, 30- and 50-minute timepoints. Control experiments administered equivalent doses of unlabeled glycine at the three time points. In all experiments, at the 60-minute timepoint, mice were euthanized, and tissues were harvested and immediately frozen in isopentane cooled in liquid nitrogen. Tissues were stored at -80 °C until sectioning for MSI studies. The use of animals and all associated procedures were reviewed and approved by the NC State University Institutional Animal Care and Use Committee. Data shown here are from two mice: one SIL-glycine labeled mouse and one control/unlabeled mouse. Replicate sections from each liver were analyzed (n=2).



**Figure 6.1.** Summary of workflow employing MIST protocol for mapping of functional heterogeneity of GSH across mouse liver tissue via IR-MALDESI-MSI.

### 6.2.2 Sample Preparation

The workflow for preparing and analyzing samples is summarized in **Figure 6.1**. Microscope slides (1 mm height, plain, Fisher Scientific, Pittsburgh, PA) were evenly sprayed with a solution of 1 mg/mL homogluthathione (hGSH) (BaChem, Torrance, CA, USA) in 50% methanol using a pneumatic sprayer (TM Sprayer, HT-X Technologies, Chapel Hill, NC, USA) and previously optimized settings.<sup>34</sup> The tissues were equilibrated to -15°C and sliced to 20 µm thickness using a Leica CM1950 cryostat (Buffalo Grove, IL, USA) prior to being thaw-mounted onto the prepared hGSH-coated slides. Prepared slides that were not immediately analyzed were stored at -80 °C until IR-MALDESI analysis.

### 6.2.3 Mass Spectrometry Imaging

All tissue analyses were performed using the home-built IR-MALDESI source<sup>35</sup> coupled with an Exploris 240 Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany) for MSI. Mounted tissue sections were placed on a Peltier-cooled XY stage within the source for analysis. To apply a desorption facilitating ice matrix to sample surfaces<sup>20</sup> the source enclosure was purged with nitrogen gas until relative humidity within the enclosure reached 10%. The sample stage was then cooled to -8 °C and exposed to humidity until an even ice matrix was formed. The enclosure was then resealed and purged again to maintain the 10% relative humidity throughout analysis of the mounted tissues.

A 2970-nm wavelength laser (JGMA, Burlington, MA, USA) was used to ablate targeted tissue regions with a single burst of five pulses (1 mJ/burst of energy) at a rate of 10 kHz. An orthogonal electrospray plume was generated by applying a voltage of 3.2 kV to electrospray solvent containing 1 mM acetic acid in 50:50 (v/v) water: acetonitrile and flowing at 1.5  $\mu\text{L}/\text{min}$  in negative mode. Ionized molecules then entered the mass spectrometer for analysis. The automatic gain control function (AGC) remained off while the injection time (IT) was fixed at 15 ms to synchronize timing of the ablation plume with the collection of ions in the C-trap. Ions were analyzed in negative mode at a mass resolving power of 240,000<sub>FWHM</sub> at 200  $m/z$  with a  $m/z$  range of 286-346. This  $m/z$  range was chosen to target the reduced form GSH and homogluthathione molecules. While no lock mass was utilized, the instrument was mass calibrated daily to maintain sub-parts per million (ppm) mass measurement accuracy. Pixel size was 150  $\mu\text{m}^2$ .

### 6.2.4 Data Analysis and Processing

Recorded mass spectra and isotopic distributions were directly viewed and analyzed in XCalibur. Raw data files were also converted from the XCalibur .RAW format to mzML files via

the ProteoWizard tool, MSConvert,<sup>36</sup> and then to imzML format using imzMLConverter<sup>37</sup>. Images were generated from the imzML files in MSiReader v1.03c. A new software tool was developed for the analysis of these data and was used to calculate abundances including correction for the overlap between GSH isotopologues and SIL monoisotopic peaks. Corrected abundances were used to calculate the percent isotope enrichment of each SIL in the GSH product pool, along with the proportion of remaining native GSH measured in each voxel. Specifics of the calculations made by the tool are discussed below.

### 6.2.5 Percent Isotopic Enrichment (PIE) Tool in MSiReader

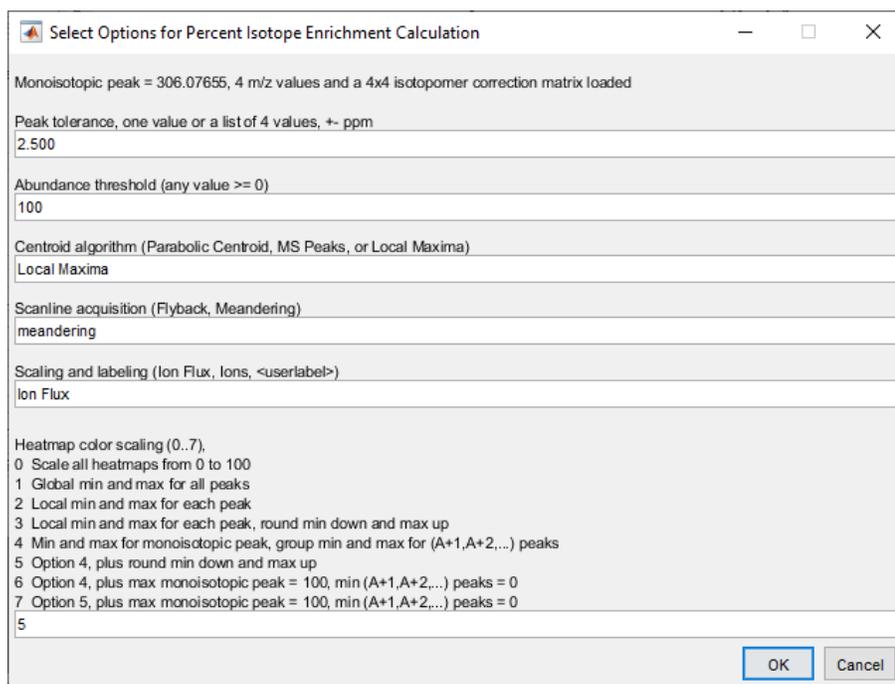
Computing the percent enrichments of GSH within each voxel in many images is a straightforward task, but a daunting one to perform manually. For this, we have created and now introduce the *PIE* tool in MSiReader. The Application Programming Interface (API) for the MSiReader software was used to add a toolbar icon and associated function to perform the percent enrichment calculation on the currently loaded data set for any list of  $m/z$  isotopologues and an abundance correction matrix. The correction matrix is supplied by the user. It may contain either probabilities or relative abundances for each isotope and must be full rank. The construction of the matrix following a method similar to that of Fernandez et al.<sup>38</sup> is described below.

The PIE tool also allows for the input of user defined parameters such as an abundance threshold, mass tolerance, and centroid algorithm (**Figure 6.2**). It performs the described calculations within the tool, and outputs heatmaps describing the percent enrichment of each isotopologue of interest. Users can select one of eight schemes for scaling the color of the heatmaps in the selection menu that opens with the tool. It also outputs an Excel worksheet containing all input parameters and both corrected abundances and final percent enrichments for further quantitative analysis by the user. The calculation can be applied to the entire image or a selected

region of interest. Examples of input isotopologues correction matrix and the Excel output of the PIE tool are included in the Supplemental Information along with a representative imzML data set from enriched mouse liver tissue.

#### 6.2.6 Droplet Analysis of SIL-GSH solutions for PIE Tool Validation

Fifteen solutions were prepared using unlabeled reduced GSH standard (Fisher Scientific) and SIL-GSH (Glycine-labeled [ $^{13}\text{C}_2$ ,  $^{15}\text{N}$ ], Cambridge Isotope Laboratories, Tewksbury, MA) in LC-MS grade water. Each solution was prepared to a 1 mM concentration made up of a ratio of the two standards to measure known percent enrichments. For example, the solution comprising of 0.005 mM SIL-GSH and 0.995 mM GSH should theoretically represent a 0.5% percent isotope enrichment of SIL-GSH. Droplets containing 7  $\mu\text{L}$  of each solution were directly pipetted onto a glass microscope slide and analyzed via IR-MALDESI in a similar manner as described above except for a narrowed mass range of 286-326  $m/z$ . Each droplet was ablated 100 times (for 100 mass spectra) per concentration. The dataset was externally mass calibrated to 2-



**Figure 6.2.** User-defined workflow settings for the PIE tool in MSiReader.

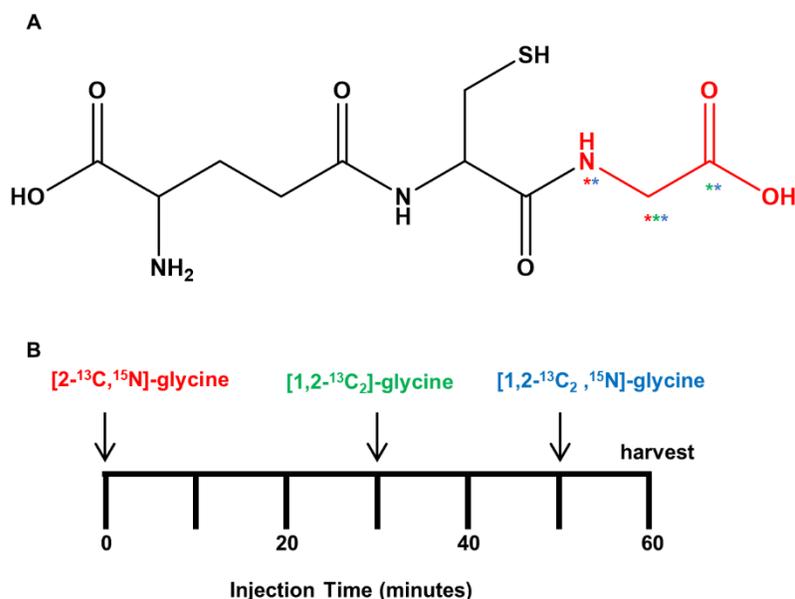
Undecylbenzenesulfonic Acid ( $m/z$  311.1686, putatively identified from chemical formula  $C_{17}H_{28}O_3S$  identified using carbon and sulfur counting) using the MSiReader external calibration tool. Percent isotope enrichments over each of the 100 scans were calculated using the PIE tool and averaged. We then fit a linear regression to determine the fit of these calculated percent enrichments to the theoretical spiked enrichment values as an indication of the accuracy of the PIE tool.

## 6.3 Results and Discussion

### 6.3.1 Multiple Infusion Start Times of Glycine Isotopologues

The workflow presented here for spatially mapping the isotopic enrichment of glycine precursors into mouse liver GSH is summarized above in **Figure 6.1**. Glycine is an immediate precursor of GSH, so the labeled isotopologues directly incorporated into the glycine residue of the tripeptide (**Figure 6.3A**). Mice were first injected with three distinct SIL-glycine isotopologues. Three SIL-glycine isotopologues, [2- $^{13}C$ , $^{15}N$ ]-glycine, [1,2- $^{13}C_2$ ]-glycine, and [1,2- $^{13}C_2$ , $^{15}N$ ]-glycine, were strategically selected because their incorporation into GSH would produce monoisotopic peaks resolvable from one another and unlabeled GSH using a high resolving power mass analyzer to be measured within a single tissue section. The [M-H<sup>+</sup>] species of unlabeled GSH has a calculated  $m/z$  of 306.0765 Da. The  $^{13}C$  and  $^{15}N$  isotopes have mass shifts of 1.0034 Da and 0.997 Da, respectively, from their lighter and more abundant isotopes  $^{12}C$  and  $^{14}N$ . Therefore, we expected elevated signals at the GSH peaks of 308.0769 (2- $^{13}C$ , $^{15}N$ ), 308.0832 (1,2- $^{13}C_2$ ), and 309.0803 (1,2- $^{13}C_2$ , $^{15}N$ ) due to metabolic incorporation of the SIL-glycine labels. We will refer to these labels as SIL1, SIL2, and SIL3 (relative to both increasing  $m/z$  and to injection order) hereafter for brevity. As each of the SIL-GSH molecules are chemically the same, we assumed ionization efficiency was the same for all enriched products. The timing of label

injection is presented in **Figure 6.3B**. The timing intervals were chosen based on an earlier pulsed MIST study of muscle protein synthesis<sup>26</sup> and the rate of glutathione labeling found in porcine liver in a continuous infusion isotope study.<sup>39</sup>

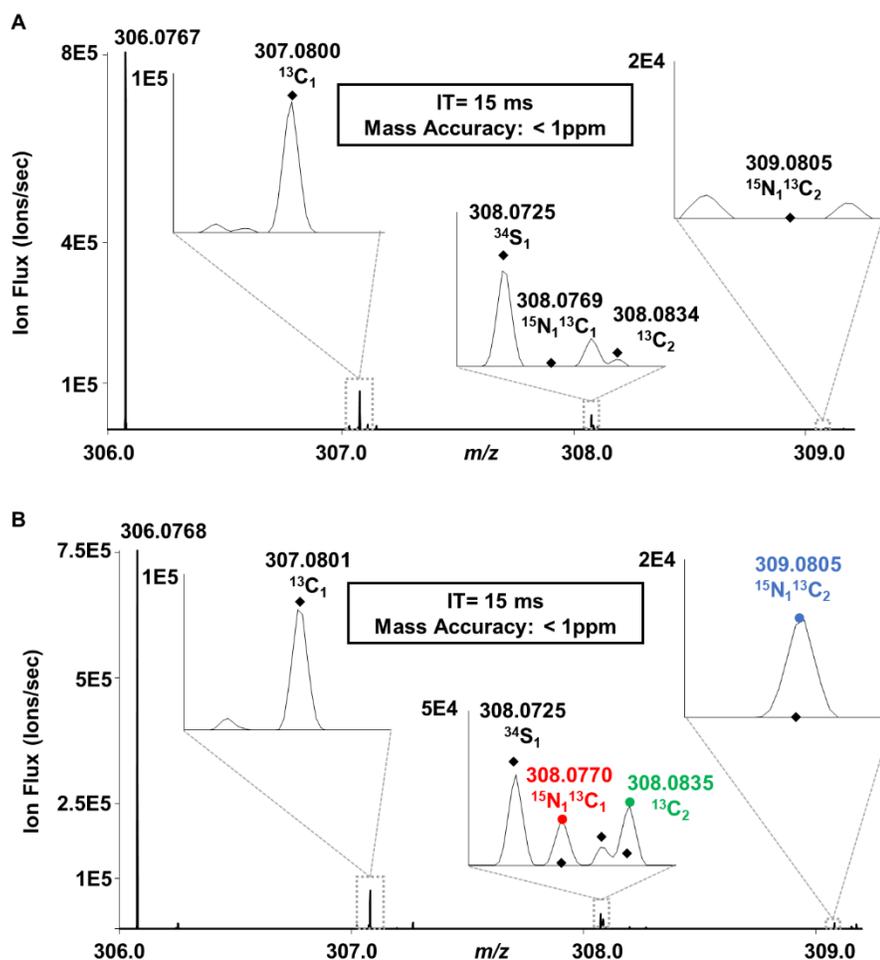


**Figure 6.3.** Summary of glycine incorporation into reduced glutathione. Color is used to distinguish the three glycine labels infused in this protocol and the resulting incorporation in GSH to produce SIL1 (red), SIL2 (green), and SIL3 (blue). **A**) Structure of reduced glutathione. SIL are incorporated in the glycine residue (red). Stars represent location of isotopically labeled atoms in each respective SIL-GSH. **B**) Summary of multiple infusion start time protocol used to inject glycine labels into each mouse over the course of one hour.

### 6.3.2 Confirmed Detection of GSH and SIL-GSH in Mouse Liver Tissue Using Mass Measurement and Spectral Accuracy

Tandem mass spectrometry (MS/MS) is usually employed to produce precursor defining fragments for confident identifications of compounds. However, in the mass spectrometry imaging field, due to the significant amount of time required to generate images, tandem MS is not typically performed unless there is not sufficient information available in the MS1 data to confer identification. To confidently confirm the detection of GSH and isotopologues without MS/MS,

high mass and spectral accuracy, and biological context were used. As shown in **Figure 6.4**, an abundant monoisotopic peak was measured within one ppm of the theoretical  $m/z$  of deprotonated GSH (306.0765) in both the unlabeled and labeled mouse liver tissue. The putative  $^{13}\text{C}$  and  $^{34}\text{S}$  GSH isotopic peaks were also detected and fully resolved from other peaks in the isotopic distribution. Based on the similarity between the theoretical distribution of native GSH (black diamonds, **Figure 6.4**) and the measured peaks, we concluded this was a confident identification



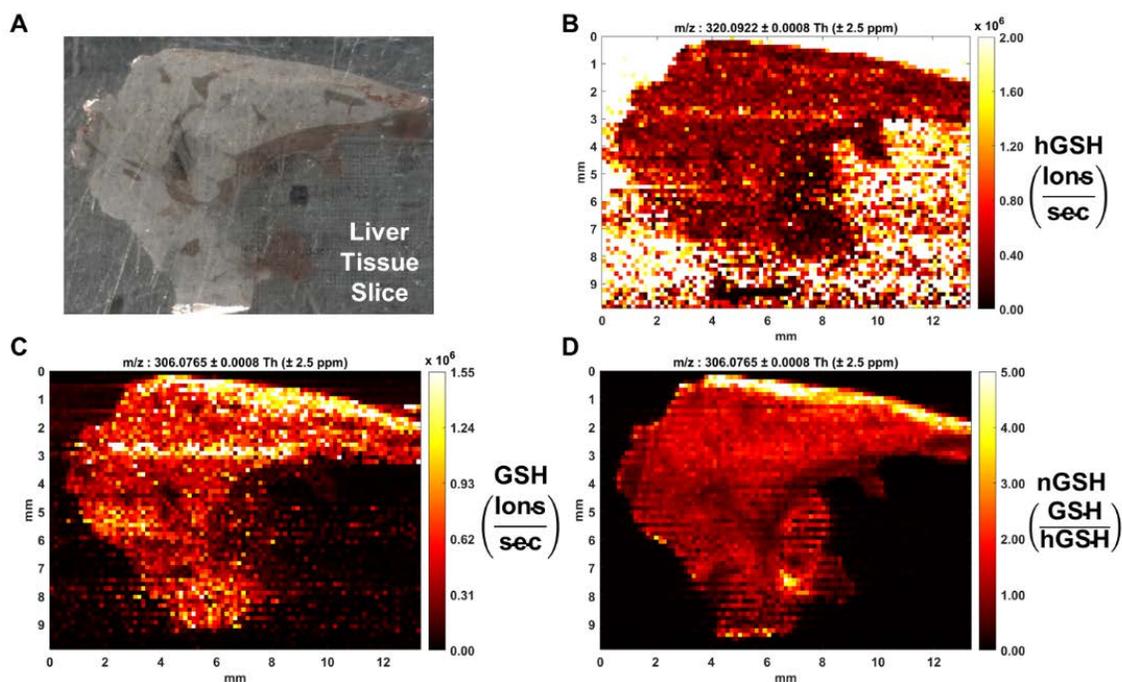
**Figure 6.4.** Measured mass spectrum of GSH from liver tissue sections. The theoretical isotopic distribution of unlabeled GSH is shown by the black diamonds. The measured  $^{13}\text{C}$  and  $^{34}\text{S}$  peak match the theoretical distribution in both the **A**) unlabeled and **B**) labeled liver section. Peaks corresponding to SIL-enrichment in labeled liver are indicated with red, green, and blue circles. Signal at these peaks was much higher than the theoretical distribution in **B** only, confirming enrichment via SIL-glycine incorporation.

of native GSH. Additionally, we utilized the spectral accuracy tool in MSiReader to quantify the differences in theoretical and estimated number of carbons and sulfur atoms using carbon/sulfur counting.<sup>40</sup> The majority of carbon and sulfur peaks measured across the tissue were estimated within one atom of the theoretical number of atoms (**Figure C.1**), supporting confident identification across the tissue sections. We compared the spectra measured in an unlabeled mouse liver (**Figure 6.4A**) with spectra in a labeled mouse liver (**Figure 6.4B**) to confirm the absence or low signal of SIL peaks in the isotopic distribution of naturally abundant GSH where we expected enrichment based on calculated mass shifts from the SIL (**Figure 6.4B**). The elevated signal of these peaks in the labeled mouse was initial evidence of successful incorporation of the SIL-glycine isotopologues into the GSH pool.

### *6.3.3 Mapping of Labels Across Tissue Sections and Normalization*

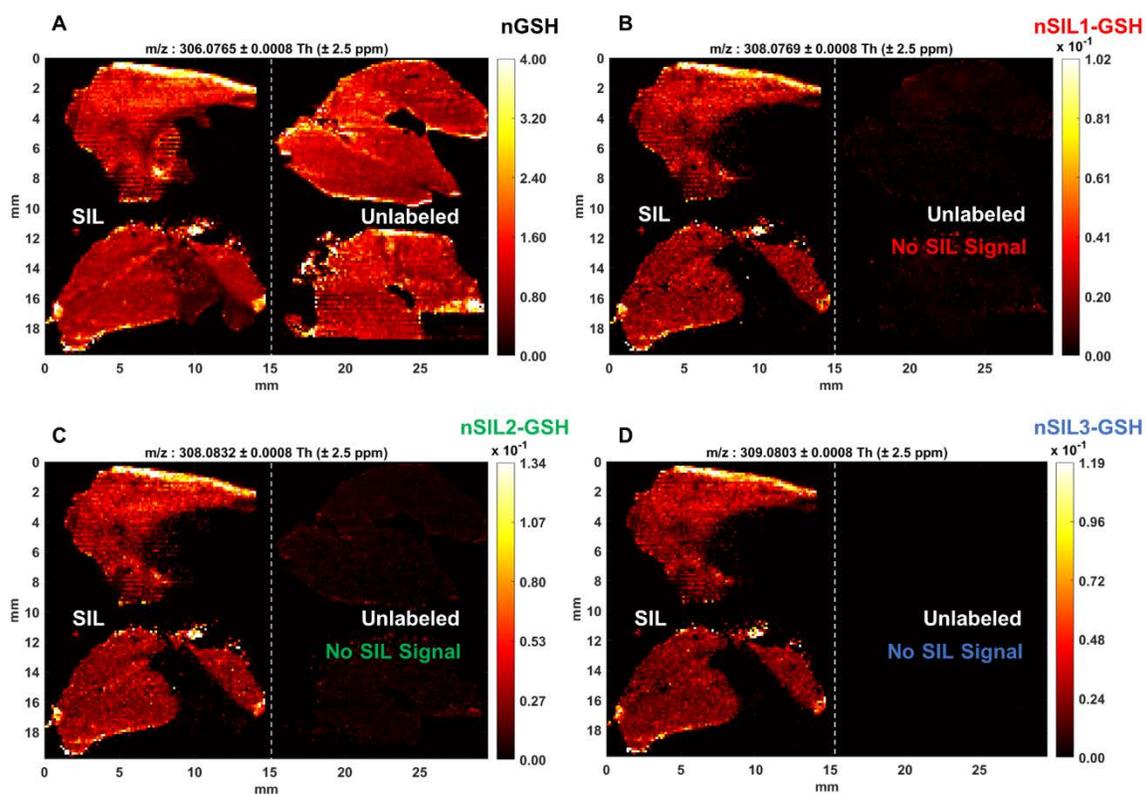
Following observation of the presence of SIL-GSH peaks within the raw mass spectra collected from the tissue, we used MSiReader to quantify the relative abundance of these peaks across the liver tissue slice. **Figure 6.5A** shows an optical image of a liver tissue section, within which several features are visible. Variability between voxels due to variations in analytical factors (e.g., ionization efficiency, electrospray stability, or variability in tissue density) can often hinder accurate visibility of changes in analyte signal across anatomical features in a tissue section.<sup>41</sup> This is significant challenge faced in the MSI community and was discussed further by Nazari and coworkers,<sup>11</sup> who described a method for normalization of GSH in hen ovarian tissue using IR-MALDESI-MSI. Towards the similar goal of GSH normalization, we chose the compound homogluthathione (hGSH), a peptide found in plants, as a normalization standard, which is chemically very similar to GSH (with the addition of one additional CH<sub>2</sub> group). A similar ionization efficiency is ideal for a normalization compound to account for any changes in the

ablation and ESI plume dynamics from voxel-to-voxel. The identified monoisotopic peak of GSH was normalized to the detected hGSH peak in each voxel. As expected for a normalization compound, very high signal at the  $m/z$  peak 320.0922 was observed off tissue (**Figure 6.5B**), while compound ionized from beneath the ablated tissue was within approximately an order of magnitude of the GSH peak. Upon normalization, we observed a successful reduction in voxel-to-voxel variation across the image, leading to a significant improvement in feature visualization (**Figure 6.5C/D**) within the steady-state tissue images. This not only improved visual comparison of data collection over several instrument days but also aided ROI selection within MSiReader. Post-processing normalization strategies for images, discussed extensively by Tu and Muddiman in 2019<sup>41</sup> and available in MSiReader, were also tested and compared with images normalized to the reference peak (**Figure C.2**).



**Figure 6.5.** Normalization of GSH peaks to hGSH-coated slide. **A)** Image of mounted liver tissue section. **B)** Heatmap of measured hGSH signal on and off tissue ( $m/z$  320.0922). **C)** Heatmap of measured GSH signal on and off tissue ( $m/z$  306.0765). **D)** Heatmap of normalized GSH signal. Physical features across the liver tissue slice are easily visible.

Following normalization, we evaluated the isotopologues' signals across multiple replicates of liver sections, from both the labeled mouse and the unlabeled mouse. **Figure 6.6A** shows normalized native abundances of GSH. Based on these images, we found that the natural abundances GSH were comparable (i.e., within an order of magnitude after normalization) between tissue slices from the same animal and the biological replicates, indicating GSH metabolism in the mice was not severely disrupted by the glycine injections. Similar signals were observed for the three SIL isotopologues across and between the labeled liver sections, while little to no signal was observed in the unlabeled tissue for these peaks (**Figure 6.6B-D**), matching the



**Figure 6.6.** Normalized heatmaps of labeled (left) and unlabeled (right) liver tissue measuring the monoisotopic peaks of **A)** unlabeled GSH ( $m/z$  306.0765) **B)** SIL1 ( $m/z$  308.0769) **C)** SIL2 ( $m/z$  308.0832) and **D)** SIL3 ( $m/z$  309.0803). While unlabeled GSH levels are similar (within one order of magnitude) between replicate sections, signal at SIL peaks is much higher in labeled tissue, as expected. This confirms the enrichment of these peaks due to labeling rather than endogenous interferences.

initial observations found directly in the mass spectra. The unlabeled tissue continued to serve as a negative control in confirming the presence of these peaks in the labeled sample were solely due to SIL infusion, and not due to a severe difference in naturally abundant GSH that enhance signals at these mass shifts. The opposite is also supported by this data, that the lack of signal at these peaks is not due to an abnormal lack of naturally abundant species. The scarcity of signal also confirmed the lack of endogenous interferences with these monoisotopic peaks in mouse liver. What little signal was observed was attributed to the natural occurrence of those unlabeled isotopologues in areas where GSH is highly abundant. This observation emphasized the importance of correcting for the contribution of the natural abundance of these isotopologues of GSH to avoid overestimating the enrichment of a specific label.

#### *6.3.4 Development and Validation of MSiReader PIE Tool*

Following inspection of the steady-state images in **Figure 6.6**, we next investigated the kinetic mapping of SIL incorporation. To calculate percent isotope enrichment (PIE) or fractional enrichment of each label within the GSH pool, it was necessary to correct for both the natural abundance of each isotopologue and the contribution of any overlapping signals from other labels (e.g., the A+1 peak of the [2-<sup>13</sup>C,<sup>15</sup>N] labeled GSH peak with the monoisotopic peak of the [1,2-<sup>13</sup>C<sub>2</sub>, <sup>15</sup>N] labeled GSH peak.) To account for all overlaps between the four peaks of interest, we computed a correction matrix.<sup>23,38</sup>

First, the probability of occurrence of each relevant isotopic peak for the labeled and unlabeled species were determined (**Table C.1**). We used the IsoSpec<sup>42</sup> package in R to calculate probabilities of isotopic peaks using IUPAC estimations of isotope natural abundances. Assuming the same probability of isotopes occurring at each  $m/z$  between the unlabeled and labeled molecules would result in an inaccurate estimation of probabilities. Therefore, only the probability of

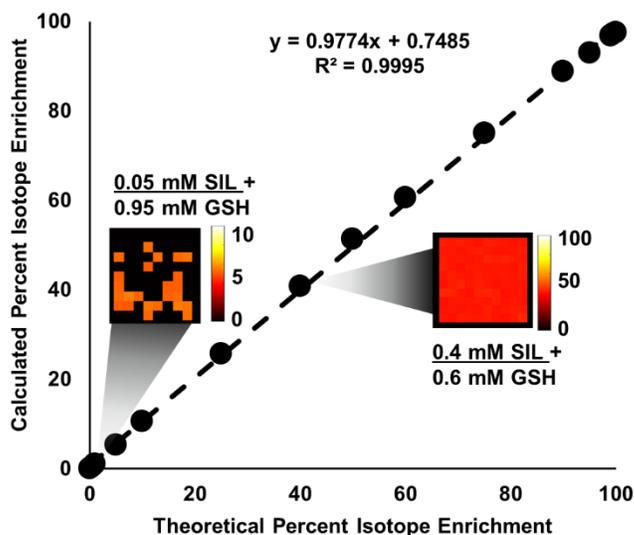
unlabeled atoms in each molecule were considered. For example, the probability of one  $^{13}\text{C}$  atom occurring in a GSH molecule (containing ten carbons total, and assuming all ten are “unlabeled”) is higher than the probability of one occurring in a SIL-GSH molecule where only nine of the carbons are naturally occurring (i.e., SIL1) due to the synthetic addition of a single  $^{13}\text{C}$ . The probability of an isotopic peak was normalized to the probability of the monoisotopic peak of the distribution to compute the expected relative abundance (RA) of that peak. The upper triangular correction matrix was filled with the computed RA ratio for each expected peak. **Equation 6.1** below relates the measured relative abundance to the corrected abundance for each isotopic peak:

$$[MA_0 \quad \dots \quad MA_n] = [CA_0 \quad \dots \quad CA_n] \times \begin{bmatrix} RA_0^0 & \dots & RA_n^0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & RA_n^n \end{bmatrix} \quad (6.1)$$

where “CA” denotes the corrected abundance for species “0” (unlabeled GSH) to “n” (in this case, n=3 and represents SIL-3 measured at  $m/z$  309.0803.) “MA” represents the measured abundances at each peak from species “0” to “n”. In order to correct each peak abundance of interest, the vector containing measured values can be multiplied by the inverse of the correction matrix<sup>38</sup> or the system of linear equations described by **Equation 6.1** can be solved by other means (e.g., QR matrix decomposition<sup>43</sup>). This results in a vector containing corrected abundances (CA) that represent only the remaining signal attributed to the enrichment of that peak with a SIL precursor. From the CAs, the PIE is computed as the ratio of each CA belonging to an unlabeled/labeled peak to the sum of all the corrected peaks as shown in **Equation 6.2**:

$$\text{Percent Isotopic Enrichment (PIE)} = \frac{CA_i}{\sum_{i=0}^n CA_i} \quad (6.2)$$

To evaluate the validity of the PIE tool prior to tissue MSI analysis, we analyzed a series of droplets with known ratios of unlabeled GSH ( $m/z$  306.0765) and SIL-GSH ( $m/z$  309.0803) making up a combined 1 mM GSH solution. A total of 15 solutions were analyzed to fit a curve to the theoretical enrichment of the respective standards in the solution. Concentrations of each standard were stepped to evaluate a range of enrichments. We fit a linear regression to compare the spiked concentrations of the standard in each solution to the enrichments computed using the PIE tool. This curve is shown in **Figure 6.7**. Heatmaps generated using the PIE tool are shown for two different solution concentrations on the curve. Even at low enrichment values of SIL (5% enrichment displayed by the top heatmap in **Figure 6.7** the enrichment is visible within the



**Figure 6.7.** Validation of PIE tool via IR-MALDESI droplet analysis of 15 solutions with known percent isotope enrichment. Heatmaps produced by the PIE tool are shown for two solutions (0.05 mM SIL-GSH + 0.95 mM GSH and 0.4 mM SIL-GSH + 0.6 mM GSH). Fitted curve of average SIL-GSH enrichments calculated using PIE tool vs known enrichments of the solutions. The curve fit the data points very well ( $R > 0.99$ ) and has a slope of approaching unity ( $m > 0.97$ ) indicating the PIE tool’s accuracy across a wide range of concentration ratios simulating “enrichment ratios”. Intercept  $> 0$  indicates a small level of error at low enrichments likely due to isotopic impurity of the SIL-GSH standard as well as potential differences between the natural abundances of isotopes (*e.g.*,  $^{13}\text{C}$ ,  $^{34}\text{S}$ ) reported by IUPAC vs specific natural abundances of isotopes in the standards [39].

heatmap images. We found that the curve had a slope of approaching unity ( $m > 0.97$ ) and a correlation coefficient ( $R$ )  $> 0.99$ , demonstrating the accuracy of the percent isotope enrichment calculations.

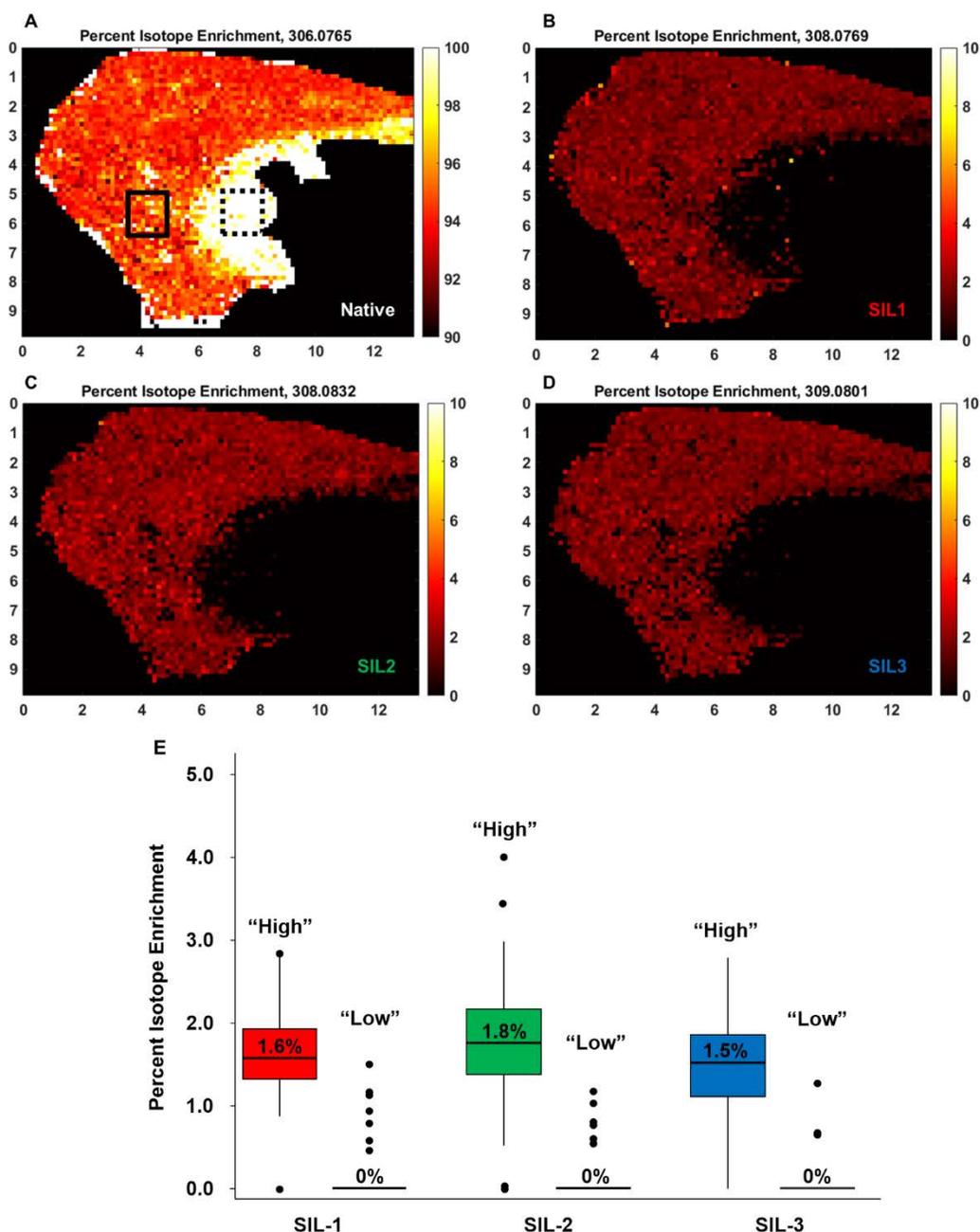
### 6.3.5 Mapping Percent Enrichment of GSH Isotopologues in Mouse Liver Tissue

We used the new MSiReader PIE tool to visualize heterogeneity in percent enrichment across the analyzed liver tissues. The overall median PIE results from each tissue are shown in **Table 6.1**. Median enrichments between replicates sections differ by less than 0.2%.

**Table 6.1.** Median percent isotope enrichment values of analyzed tissue

	Labeled Section 1	Labeled Section 2	Unlabeled Section 1	Unlabeled Section 2
SIL-1	1.7 %	1.9%	0.2%	0.2%
SIL-2	2.0 %	2.1%	0.1%	0.1%
SIL-3	1.7 %	1.8%	0.1%	0.2%

Resulting heatmaps from the analysis of “Labeled Section 1” are shown in **Figure 6.8**. **Figure 6.8A** shows the PIE distribution for unlabeled GSH in the total GSH pool. Similarly, **Figures 6.8B-D** show the PIE distribution of SIL1-3, respectively, across the liver tissue. Based on these heat maps we detected similar levels of incorporation of all three glycine labels in the GSH metabolic pool across most of the liver tissue section. Interestingly, differences in incorporation were revealed across different features of the liver section. The visualization of these differences allows us to select specific ROIs for further quantification and comparison of enrichment. For example, bright areas in **Figure 6.8A** (dotted box) indicate most of the GSH in this region was unlabeled while the orange-red region (solid box) indicates the voxels with lower amounts of unlabeled GSH and increased levels of SIL-GSH. **Figure 6.8E** displays a boxplot of the enrichment measurements



**Figure 6.8.** Percent isotope enrichment of labeled liver section. Heatmaps of **A)** unlabeled GSH ( $m/z$  306.0765) **B)** SIL1 ( $m/z$  308.0769) **C)** SIL2 ( $m/z$  308.0832) and **D)** SIL3 ( $m/z$  309.0803). An area of high incorporation of the SILs is highlighted with a solid black square and an area of low incorporation is highlighted with a dotted black square. The quantities of incorporation in these ROIs are quantified in **E**.

made in these specific ROIs. We found a median enrichment of 1.6%, 1.8%, and 1.5% of each respective SIL in the “high” ROI and approximately 94% of the GSH pool remained unlabeled. In

the “low” ROI, we found almost no labeling aside from a few outlier voxels. The similarity in PIE values for the median enrichment values for each SIL may reflect the flat portion of the kinetic curve during the transition from label uptake to washout where similar enrichments would be expected. Alternatively, the length of the intervals between bolus doses may be too short compared to the overall glutathione biosynthesis resulting in similar levels of incorporation. Work is underway to change the timing of the infusion protocols to confirm the kinetic profiles of glutathione biosynthesis. Heatmaps from the second labeled tissue and the unlabeled replicates are shown in **Figure C.3**. These data show that the method can be successfully utilized for mapping the variation in the rate of glycine uptake into glutathione, i.e., glutathione biosynthetic activity across liver tissues. Interestingly, areas of the tissues where high steady-state GSH signal were observed did not necessarily correspond to areas of high GSH flux. This emphasizes the importance of studying GSH metabolic activity rather than relying on overall GSH levels to elucidate tissue function.

## **6.4 Conclusions**

A method for kinetically labeling GSH in mouse liver tissues using a multiple infusion start time protocol is presented here. By injecting three SIL-glycine precursors at different time points, we showed the utility of IR-MALDESI-MSI for spatially mapping heterogeneity of the metabolic incorporation of those labels into the GSH product pool across mouse liver tissue sections. We also introduced the PIE tool in MSiReader which can be adapted for any kinetic labeling experiments. The method presented here will be adapted for future studies of different tissues for analysis of GSH metabolic heterogeneity in healthy and diseased states.

## **6.5 Acknowledgements**

All IR-MALDESI-MSI experiments were completed in the Molecular Education, Technology and Research Innovation Center (METRIC) at North Carolina State University. The authors gratefully acknowledge the financial support received from the National Institutes of Health (R21GM134219).

## 6.6 References

- (1) Bansal, A.; Simon, M. C. Glutathione Metabolism in Cancer Progression and Treatment Resistance. *J Cell Biol* **2018**, *217* (7), 2291–2298. <https://doi.org/10.1083/jcb.201804161>.
- (2) Gamcsik, M. P.; Kasibhatla, M. S.; Teeter, S. D.; Colvin, O. M. Glutathione Levels in Human Tumors. *Biomarkers* **2012**, *17* (8), 671–691. <https://doi.org/10.3109/1354750X.2012.715672>.
- (3) Forman, H. J.; Zhang, H.; Rinna, A. Glutathione: Overview of Its Protective Roles, Measurement, and Biosynthesis. *Mol Aspects Med* **2009**, *30* (1–2), 1–12. <https://doi.org/10.1016/j.mam.2008.08.006>.
- (4) Townsend, D. M.; Tew, K. D.; Tapiero, H. The Importance of Glutathione in Human Disease. *Biomed Pharmacother* **2003**, *57* (3–4), 145–155.
- (5) Ballatori, N.; Krance, S. M.; Notenboom, S.; Shi, S.; Tieu, K.; Hammond, C. L. Glutathione Dysregulation and the Etiology and Progression of Human Diseases. *Biol Chem* **2009**, *390* (3), 191–214. <https://doi.org/10.1515/BC.2009.033>.
- (6) Stockwell, B. R.; Friedmann Angeli, J. P.; Bayir, H.; Bush, A. I.; Conrad, M.; Dixon, S. J.; Fulda, S.; Gascón, S.; Hatzios, S. K.; Kagan, V. E.; Noel, K.; Jiang, X.; Linkermann, A.; Murphy, M. E.; Overholtzer, M.; Oyagi, A.; Pagnussat, G. C.; Park, J.; Ran, Q.; Rosenfeld, C. S.; Salnikow, K.; Tang, D.; Torti, F. M.; Torti, S. V.; Toyokuni, S.; Woerpel, K. A.; Zhang, D. D. Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease. *Cell* **2017**, *171* (2), 273–285. <https://doi.org/10.1016/j.cell.2017.09.021>.

- (7) Wu, G.; Fang, Y.-Z.; Yang, S.; Lupton, J. R.; Turner, N. D. Glutathione Metabolism and Its Implications for Health. *The Journal of Nutrition* **2004**, *134* (3), 489–492. <https://doi.org/10.1093/jn/134.3.489>.
- (8) Beckman, K. B.; Ames, B. N. The Free Radical Theory of Aging Matures. *Physiological Reviews* **1998**, *78* (2), 547–581. <https://doi.org/10.1152/physrev.1998.78.2.547>.
- (9) Maher, P. The Effects of Stress and Aging on Glutathione Metabolism. *Ageing Research Reviews* **2005**, *4* (2), 288–314. <https://doi.org/10.1016/j.arr.2005.02.005>.
- (10) Johnson, W. M.; Wilson-Delfosse, A. L.; Mieyal, J. J. Dysregulation of Glutathione Homeostasis in Neurodegenerative Diseases. *Nutrients* **2012**, *4* (10), 1399–1440. <https://doi.org/10.3390/nu4101399>.
- (11) Nazari, M.; Bokhart, M. T.; Loziuk, P. L.; Muddiman, D. C. Quantitative Mass Spectrometry Imaging of Glutathione in Healthy and Cancerous Hen Ovarian Tissue Sections by Infrared Matrix-Assisted Laser Desorption Electrospray Ionization (IR-MALDESI). *Analyst* **2018**, *143* (3), 654–661. <https://doi.org/10.1039/C7AN01828B>.
- (12) Traverso, N.; Ricciarelli, R.; Nitti, M.; Marengo, B.; Furfaro, A. L.; Pronzato, M. A.; Marinari, U. M.; Domenicotti, C. Role of Glutathione in Cancer Progression and Chemoresistance. *Oxidative Medicine and Cellular Longevity* **2013**, *2013*, 1–10. <https://doi.org/10.1155/2013/972913>.
- (13) Thelwall, P. E.; Simpson, N. E.; Rabbani, Z. N.; Clark, M. D.; Pourdeyhimi, R.; Macdonald, J. M.; Blackband, S. J.; Gamcsik, M. P. In Vivo MR Studies of Glycine and Glutathione Metabolism in a Rat Mammary Tumor. *NMR in Biomedicine* **2012**, *25* (2), 271–278. <https://doi.org/10.1002/nbm.1745>.

- (14) Kennedy, L.; Sandhu, J. K.; Harper, M.-E.; Cuperlovic-Culf, M. Role of Glutathione in Cancer: From Mechanisms to Therapies. *Biomolecules* **2020**, *10* (10). <https://doi.org/10.3390/biom10101429>.
- (15) Rae Buchberger, A.; DeLaney, K.; Johnson, J.; Li, L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal Chem* **2018**, *90* (1), 240–265. <https://doi.org/10.1021/acs.analchem.7b04733>.
- (16) Jang, C.; Chen, L.; Rabinowitz, J. D. Metabolomics and Isotope Tracing. *Cell* **2018**, *173* (4), 822–837. <https://doi.org/10.1016/j.cell.2018.03.055>.
- (17) Bokhart, M. T.; Muddiman, D. C. Infrared Matrix-Assisted Laser Desorption Electrospray Ionization Mass Spectrometry Imaging Analysis of Biospecimens. *Analyst* **2016**, *141* (18), 5236–5245. <https://doi.org/10.1039/c6an01189f>.
- (18) Caleb Bagley, M.; Garrard, K. P.; Muddiman, D. C. The Development and Application of Matrix Assisted Laser Desorption Electrospray Ionization: The Teenage Years. *Mass Spectrom Rev* **2021**. <https://doi.org/10.1002/mas.21696>.
- (19) Robichaud, G.; Barry, J. A.; Garrard, K. P.; Muddiman, D. C. Infrared Matrix-Assisted Laser Desorption Electrospray Ionization (IR-MALDESI) Imaging Source Coupled to a FT-ICR Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (1), 92–100. <https://doi.org/10.1007/s13361-012-0505-9>.
- (20) Robichaud, G.; Barry, J. A.; Muddiman, D. C. IR-MALDESI Mass Spectrometry Imaging of Biological Tissue Sections Using Ice as a Matrix. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (3), 319–328. <https://doi.org/10.1007/s13361-013-0787-6>.
- (21) Sampson, J. S.; Murray, K. K.; Muddiman, D. C. Intact and Top-Down Characterization of Biomolecules and Direct Analysis Using Infrared Matrix-Assisted Laser Desorption

- Electrospray Ionization Coupled to FT-ICR Mass Spectrometry. *J Am Soc Mass Spectrom* **2009**, *20* (4), 667–673. <https://doi.org/10.1016/j.jasms.2008.12.003>.
- (22) Nazari, M.; Muddiman, D. C. Polarity Switching Mass Spectrometry Imaging of Healthy and Cancerous Hen Ovarian Tissue Sections by Infrared Matrix-Assisted Laser Desorption Electrospray Ionization (IR-MALDESI). *Analyst* **2016**, *141* (2), 595–605. <https://doi.org/10.1039/C5AN01513H>.
- (23) Buescher, J. M.; Antoniewicz, M. R.; Boros, L. G.; Burgess, S. C.; Brunengraber, H.; Clish, C. B.; DeBerardinis, R. J.; Feron, O.; Frezza, C.; Ghesquiere, B.; Gottlieb, E.; Hiller, K.; Jones, R. G.; Kamphorst, J. J.; Kibbey, R. G.; Kimmelman, A. C.; Locasale, J. W.; Lunt, S. Y.; Maddocks, O. D.; Malloy, C.; Metallo, C. M.; Meuillet, E. J.; Munger, J.; Nöh, K.; Rabinowitz, J. D.; Ralser, M.; Sauer, U.; Stephanopoulos, G.; St-Pierre, J.; Tennant, D. A.; Wittmann, C.; Vander Heiden, M. G.; Vazquez, A.; Vousden, K.; Young, J. D.; Zamboni, N.; Fendt, S.-M. A Roadmap for Interpreting <sup>13</sup>C Metabolite Labeling Patterns from Cells. *Current Opinion in Biotechnology* **2015**, *34*, 189–201. <https://doi.org/10.1016/j.copbio.2015.02.003>.
- (24) Klein, S.; Heinzle, E. Isotope Labeling Experiments in Metabolomics and Fluxomics. *WIREs Systems Biology and Medicine* **2012**, *4* (3), 261–272. <https://doi.org/10.1002/wsbm.1167>.
- (25) Wilkinson, D. J. Historical and Contemporary Stable Isotope Tracer Approaches to Studying Mammalian Protein Metabolism. *Mass Spectrometry Reviews* **2018**, *37* (1), 57–80. <https://doi.org/10.1002/mas.21507>.

- (26) Zhang, X.-J.; Chinkes, D. L.; Wolfe, R. R. Measurement of Muscle Protein Fractional Synthesis and Breakdown Rates from a Pulse Tracer Injection. *Am J Physiol Endocrinol Metab* **2002**, *283* (4), E753-764. <https://doi.org/10.1152/ajpendo.00053.2002>.
- (27) Lane, A. N.; Yan, J.; Fan, T. W.-M. <sup>13</sup>C Tracer Studies of Metabolism in Mouse Tumor Xenografts. *Bio Protoc* **2015**, *5* (22), [bio-protocol.org/e1650](https://doi.org/10.1093/bio-protoc/bt1650).
- (28) Yuan, M.; Kremer, D. M.; Huang, H.; Breitkopf, S. B.; Ben-Sahra, I.; Manning, B. D.; Lyssiotis, C. A.; Asara, J. M. Ex Vivo and in Vivo Stable Isotope Labelling of Central Carbon Metabolism and Related Pathways with Analysis by LC–MS/MS. *Nat Protoc* **2019**, *14* (2), 313–330. <https://doi.org/10.1038/s41596-018-0102-x>.
- (29) Agostini, F.; Libera, L. D.; Rittweger, J.; Mazzucco, S.; Jurdana, M.; Mekjavic, I. B.; Pišot, R.; Gorza, L.; Narici, M.; Biolo, G. Effects of Inactivity on Human Muscle Glutathione Synthesis by a Double-Tracer and Single-Biopsy Approach. *J Physiol* **2010**, *588* (Pt 24), 5089–5104. <https://doi.org/10.1113/jphysiol.2010.198283>.
- (30) Cl, P.; Am, M.; Aw, P.; Pv, F.; G, P.; Fc, B. A Multiple Infusion Start Time (MIST) Protocol for Stable Isotope Studies of Fetal Blood. *Placenta* **2001**, *22* (2–3), 171–176. <https://doi.org/10.1053/plac.2000.0599>.
- (31) Dudley, M. A.; Burrin, D. G.; Wykes, L. J.; Toffolo, G.; Cobelli, C.; Nichols, B. L.; Rosenberger, J.; Jahoor, F.; Reeds, P. J. Protein Kinetics Determined in Vivo with a Multiple-Tracer, Single-Sample Protocol: Application to Lactase Synthesis. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **1998**, *274* (3), G591–G598. <https://doi.org/10.1152/ajpgi.1998.274.3.G591>.
- (32) Bokhart, M. T.; Nazari, M.; Garrard, K. P.; Muddiman, D. C. MSiReader v1.0: Evolving Open-Source Mass Spectrometry Imaging Software for Targeted and Untargeted Analyses.

- J. Am. Soc. Mass Spectrom.* **2018**, *29* (1), 8–16. <https://doi.org/10.1007/s13361-017-1809-6>.
- (33) MSiReader, <https://msireader.wordpress.ncsu.edu/>
- (34) Bokhart, M. T.; Rosen, E.; Thompson, C.; Sykes, C.; Kashuba, A. D. M.; Muddiman, D. C. Quantitative Mass Spectrometry Imaging of Emtricitabine in Cervical Tissue Model Using Infrared Matrix-Assisted Laser Desorption Electrospray Ionization. *Anal Bioanal Chem* **2015**, *407* (8), 2073–2084. <https://doi.org/10.1007/s00216-014-8220-y>.
- (35) Garrard, K. P.; Ekelöf, M.; Khodjaniazova, S.; Bagley, M. C.; Muddiman, D. C. A Versatile Platform for Mass Spectrometry Imaging of Arbitrary Spatial Patterns. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (12), 2547–2552. <https://doi.org/10.1021/jasms.0c00128>.
- (36) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; 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, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nature Biotechnology* **2012**, *30* (10), 918–920. <https://doi.org/10.1038/nbt.2377>.
- (37) Race, A. M.; Styles, I. B.; Bunch, J. Inclusive Sharing of Mass Spectrometry Imaging Data Requires a Converter for All. *Journal of Proteomics* **2012**, *75* (16), 5111–5112. <https://doi.org/10.1016/j.jprot.2012.05.035>.

- (38) Fernandez, C. A.; Rosiers, C. D.; Previs, S. F.; David, F.; Brunengraber, H. Correction of  $^{13}\text{C}$  Mass Isotopomer Distributions for Natural Stable Isotope Abundance. *Journal of Mass Spectrometry* **1996**, *31* (3), 255–262. [https://doi.org/10.1002/\(SICI\)1096-9888\(199603\)31:3<255::AID-JMS290>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1096-9888(199603)31:3<255::AID-JMS290>3.0.CO;2-3).
- (39) Jahoor, F.; Wykes, L. J.; Reeds, P. J.; Henry, J. F.; del Rosario, M. P.; Frazer, M. E. Protein-Deficient Pigs Cannot Maintain Reduced Glutathione Homeostasis When Subjected to the Stress of Inflammation. *The Journal of Nutrition* **1995**, *125* (6), 1462–1472. <https://doi.org/10.1093/jn/125.6.1462>.
- (40) Khodjaniyazova, S.; Nazari, M.; Garrard, K. P.; Matos, M. P. V.; Jackson, G. P.; Muddiman, D. C. Characterization of the Spectral Accuracy of an Orbitrap Mass Analyzer Using Isotope Ratio Mass Spectrometry. *Anal. Chem.* **2018**, *90* (3), 1897–1906. <https://doi.org/10.1021/acs.analchem.7b03983>.
- (41) Tu, A.; Muddiman, D. C. Systematic Evaluation of Repeatability of IR-MALDESI-MS and Normalization Strategies for Correcting the Analytical Variation and Improving Image Quality. *Anal Bioanal Chem* **2019**, *411* (22), 5729–5743. <https://doi.org/10.1007/s00216-019-01953-5>.
- (42) Łacki, M. K.; Startek, M.; Valkenburg, D.; Gambin, A. IsoSpec: Hyperfast Fine Structure Calculator. *Anal. Chem.* **2017**, *89* (6), 3272–3277. <https://doi.org/10.1021/acs.analchem.6b01459>.
- (43) Weisstein, E. W. QR Decomposition <https://mathworld.wolfram.com/QRDecomposition.html> (accessed 2021 -07 -19).

## CHAPTER 7: Mapping Glutathione Metabolic Activity in Mouse Mammary Tumors Using Functional Mass Spectrometry Imaging

Submitted: Mellinger, A.L., Kibbe, R.R., Rabbani, Z.N., Meritet, D., Muddiman, D.C., Gamcsik, M.P., *Free Radical Biology & Medicine*, 2022.

### 7.1 Introduction

Glutathione is one of the primary cellular antioxidants and plays a central role in neutralizing oxidative stress in tumor tissue imparted by reprogrammed metabolism, the harsh hypoxic and acidic environments and the additional stress produced by therapy.<sup>1-4</sup> For these reasons, a number of clinical studies sought to link glutathione levels in patient tumors to tumor clinical stage, histological grade or therapy response but have yielded mixed results.<sup>5</sup> These results are likely due to the difficulty in relating the static level of a single biomarker to the functional capability of tumor tissue to replenish and maintain antioxidant capacity in response to stress. This functional capability can be assessed by measuring metabolic rates as these are the net result of the interactions of an entire network of molecular components in the tissue under regulatory control and environmental influences.<sup>6-8</sup> Since tumor tissue is notoriously heterogeneous in molecular content and in microenvironmental conditions, the metabolic activity will vary substantially across the tissue. Therefore, functional imaging is needed to map these variations across the tumor. In prior work, we used magnetic resonance methods to noninvasively detect the heterogeneity of isotope incorporation into tumor glutathione, but the low sensitivity limited the amount of kinetic information obtainable from these experiments.<sup>9,10</sup>

We developed a new method of functional mass spectrometry imaging (fMSI)<sup>11</sup> that used timed administration of isotopologues of glycine that yielded both glutathione metabolic activity maps at each spatial location in ex vivo liver tissue sections.<sup>12</sup> This single-sample isotope tracer

method to provide detailed kinetic data is based on the work of Dudley et al.<sup>13</sup>, and later dubbed multiple-infusion start time (MIST)<sup>14</sup> and is particularly advantageous to studying heterogeneous tumor tissue. This method can provide multiple timepoint kinetic data from a single tissue sample. In this report, mice bearing orthotopically implanted 4T1 mammary tumors were infused with three isotopologues of glycine, [2-<sup>13</sup>C,<sup>15</sup>N]-glycine, [1,2-<sup>13</sup>C<sub>2</sub>]-glycine, [1,2-<sup>13</sup>C<sub>2</sub>,<sup>15</sup>N]-glycine, at timed intervals followed by a single tissue harvest. Mass spectrometry analysis of the tissue shows isotope labeling patterns that reflect the time dependent uptake of each glycine isotopologue and its conversion into glutathione allowing kinetic analysis of metabolic activity. The results show that glycine uptake is heterogeneous, and glutathione metabolic activity is highest at the tumor periphery.

## 7.2 Materials and Methods

### 7.2.1 Animal Studies

Female Balb/c mice (6-8 weeks old) were obtained from Charles River (Wilmington, MA). Approximately 10<sup>5</sup> 4T1 tumor cells (ATCC) were injected into the mammary pad of the mice. After 12-21 days, the tumors range from 0.5 to 1 cm<sup>3</sup> in size. Mice were anesthetized (isoflurane/oxygen) and the tail vein was cannulated. The three stable isotope-labeled (SIL) isotopologues of glycine, [2-<sup>13</sup>C,<sup>15</sup>N]-glycine (abbreviated: **<sup>1</sup>gly**), [1,2-<sup>13</sup>C<sub>2</sub>]-glycine (**<sup>2</sup>gly**), and [1,2-<sup>13</sup>C<sub>2</sub>,<sup>15</sup>N]-glycine (**<sup>3</sup>gly**) (**Figure 7.1A**), and [<sup>13</sup>C<sub>2</sub>,<sup>15</sup>N]- urea (**\*urea**) were obtained from Cambridge Isotope Laboratories (Tewksbury, MA) and were dissolved in phosphate-buffered saline. The isotopologues were administered using the MIST protocol to the anesthetized mice by bolus injection (0.05 mL) at doses of 300 μmoles/kg via the tail vein cannula at 0-, 40- and 60-minute timepoints. At the 119-minute timepoint, a 50 μmol bolus of isotope-labeled urea (**\*urea**) was administered as a perfusion marker and mice were euthanized at 120 min (**Figure 7.1C**).

Control experiments infused unlabeled glycine (<sup>0</sup>gly) over the same timing sequence. Tissues were harvested and immediately frozen in isopentane cooled in liquid nitrogen. Tissues were stored at -80°C until sectioning for MSI studies. The use of animals and all associated procedures were reviewed and approved by the NC State University Institutional Animal Care and Use Committee. The 4T1 cell line was authenticated by short tandem repeat profiling by the ATCC.

### *7.2.2 Sample Preparation*

Microscope slides were evenly sprayed with a solution of 1 mg/mL homoglutathione (Bachem) in 50% methanol using a pneumatic sprayer as described previously.<sup>12</sup> The tumor tissues were sliced to a 20 µm thickness with a cryostat prior to being thaw mounted on the prepared slide for mass spectrometry imaging. Adjacent 10 µm thick sections were prepared for histological examination.

### *7.2.3 Mass Spectrometry Imaging, Data Analysis, and Preparation*

All tissue analyses were performed using a home-built infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI) source coupled with an Exploris 240 Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Bremen Germany) for MSI by the methods previously described.<sup>12</sup> Recorded mass spectra and isotopic distributions were directly viewed and analyzed in XCalibur. Specifics of data analysis and the software tool used to generate metabolite isotopic enrichments appear in our recent publication.<sup>12</sup> Glutathione was detected in negative ion mode whereas glycine and urea were collected in positive ion mode. In positive ion mode, the  $m/z$  range 45-225 was recorded. Internal calibration was used to achieve sub-ppm mass accuracy. We used an ESI solvent of 50% acetonitrile and 0.2% formic acid. The amount of isotope incorporated into metabolites was performed using the Percent Isotope Enrichment (PIE) tool in the MSiReader software.<sup>12</sup> The MSI data is collected from tissue 20 µm thick tissue section with resulting sampled

volume elements (voxels) of  $150 \times 150 \times 20 \mu\text{m}$  size. Functional heatmaps displaying voxel-by-voxel slopes were generated from MSiReader PIE tool output in RStudio using the nlme, viridis, and ggplot2 packages.

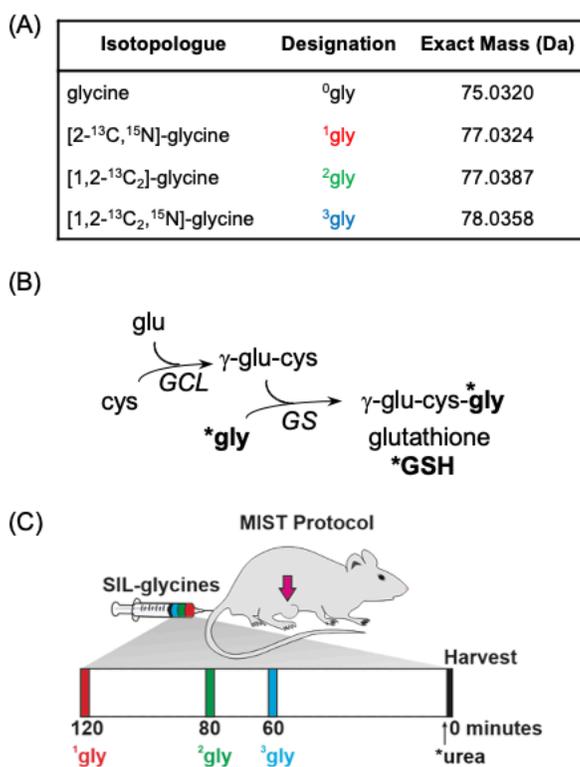
#### 7.2.4 Histology and Immunohistochemistry

Ten-micron tissue sections were obtained adjacent to those used for mass spectrometry analysis. One section was stained with hematoxylin-eosin stain. This tissue section was analyzed to detect regions of necrosis and neutrophil infiltration by a board-certified veterinary anatomic pathologist (DM) using light microscopy.

### 7.3 Results

#### 7.3.1 Distribution and Labeling of Glutathione

All fMSI experiments used heavy SIL glycines, designated as **\*gly**, with control experiments infusing unlabeled, naturally abundant glycine (**<sup>0</sup>gly**). The labeling patterns of the three isotope-enriched isotopologues of glycine (**<sup>1</sup>gly**, **<sup>2</sup>gly**, **<sup>3</sup>gly**), are shown in **Figure 7.1A** and yield distinct species with increasing exact masses that can be resolved by high resolution mass spectrometry.<sup>12</sup> These **\*gly** species can be incorporated into glutathione in a two-step biosynthetic process catalyzed by the enzymes glutamate cysteine ligase (*GCL*) and glutathione synthetase (*GS*) to yield isotopologues of glutathione (**\*GSH**) that can also be resolved by high resolution mass spectrometry (**Figure 7.1B**). The MIST protocol was used to administer bolus doses of **<sup>1</sup>gly**, **<sup>2</sup>gly**, **<sup>3</sup>gly** at 120, 80, and 60 minutes, respectively, prior to tissue harvest as shown in **Figure 7.1C**. SIL-urea (**\*urea**) was administered one minute before harvest as a perfusion marker.<sup>15</sup> Incorporation of the three glycine isotopologues for three different time intervals resulted in the enrichment of three isotopologues of glutathione **<sup>1</sup>GSH**, **<sup>2</sup>GSH**, **<sup>3</sup>GSH**, after 120, 80 or 60 minutes, respectively. Therefore, each tissue sample will contain unlabeled glutathione, **<sup>0</sup>GSH**, and **<sup>1</sup>GSH**, **<sup>2</sup>GSH**, **<sup>3</sup>GSH**.

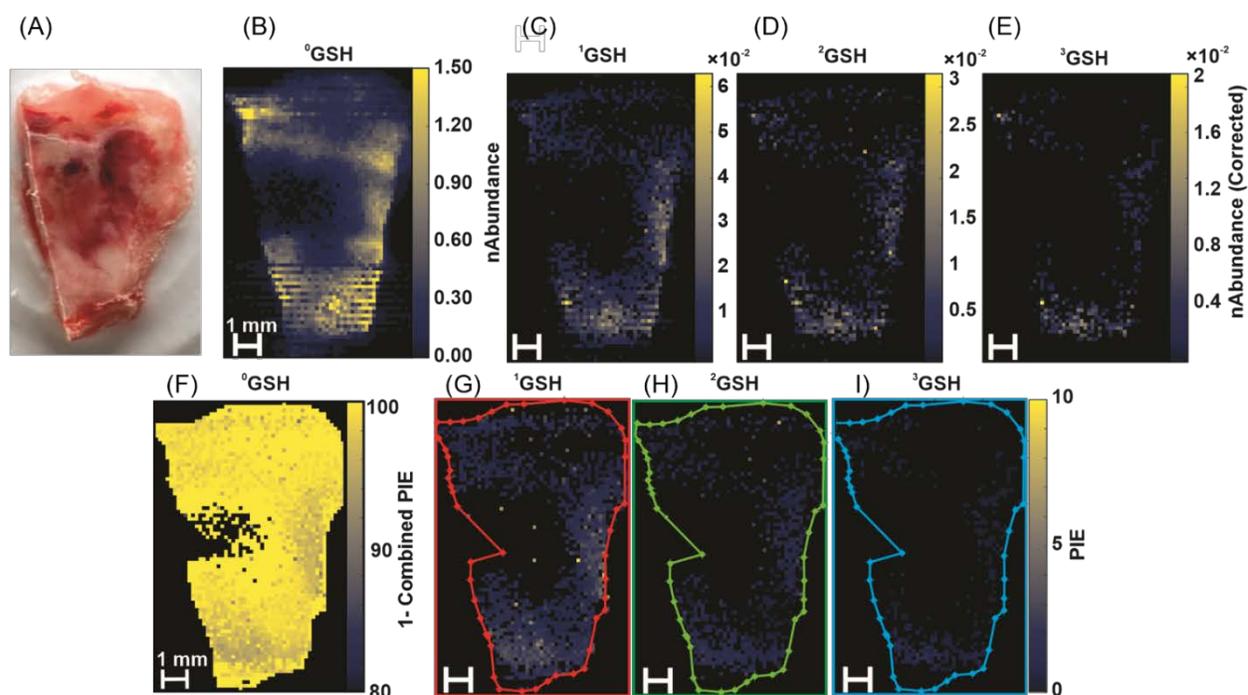


**Figure 1.** A) Glycine isotopologues used in this study with listed exact masses. B) Glutathione is biosynthesized in two steps from three amino acids catalyzed by glutamate cysteine ligase (*GCL*) and glutathione synthetase (*GS*). The three glycine isotopologues (<sup>\*</sup>gly) yield the corresponding glutathione isotopologues (<sup>\*</sup>GSH). C) Timing of the MIST administration of glycine isotopologues (<sup>\*</sup>gly) and isotope labeled urea (<sup>\*</sup>urea).

The percent isotope enrichment (PIE) of each isotopologue in the sample, obtained from the mass spectrometry imaging data, represents the ratio of isotope-labeled to total glutathione from the three time points.

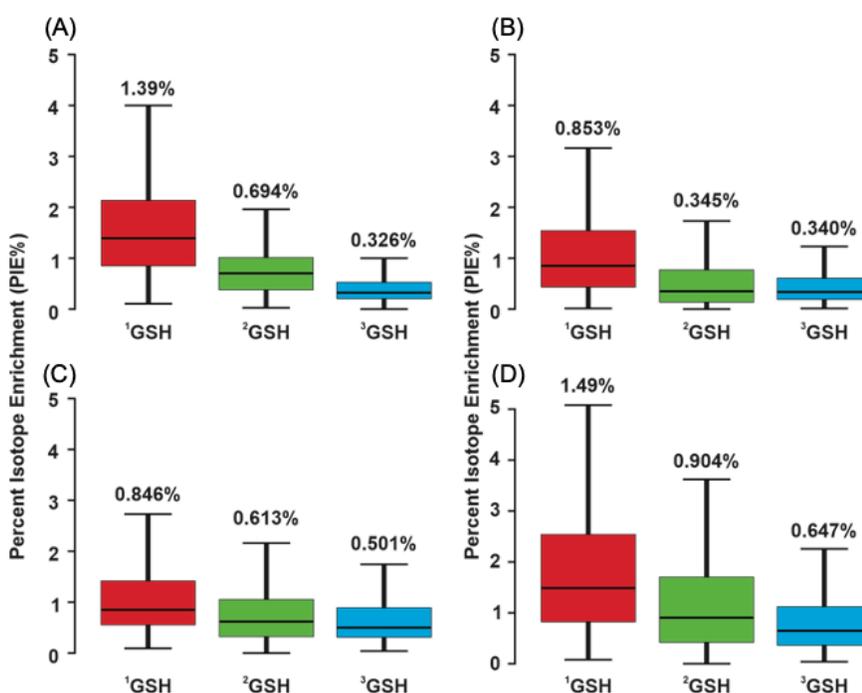
A harvested 4T1 tumor tissue sample that was used to obtain thin sections for mass spectrometry and histological analyses is shown in **Figure 7.2A**. The overall cross-sectional dimensions of the tissue shown in **Figure 7.2A** was approximately 11 x 6 mm. Thin 20  $\mu$ m sections of this tumor were analyzed by MSI. The steady-state levels of unlabeled glutathione (<sup>0</sup>GSH), detected at  $m/z$  306.0765, and normalized to the homoglutathione standard, is shown in **Figure 7.2B**. This image shows high glutathione levels at the periphery of the tumor with a large central

region with low/no detectable glutathione. Similarly, heatmaps of the steady-state abundances of the three GSH isotopologues detected at  $m/z$ 's 308.0769 ( $^1\text{GSH}$ ), 308.0832 ( $^2\text{GSH}$ ), 308.0803 ( $^3\text{GSH}$ ) are shown in **Figures 7.2C-E**, respectively. As expected, the  $^1\text{GSH}$ , synthesized from  $^1\text{gly}$  over a 120-min exposure is highest in concentration, followed by  $^2\text{GSH}$  (80-min exposure) and  $^3\text{GSH}$  (60-min exposure) to the other glycine isotopologues. The PIE heatmaps of all glutathione species are presented in **Figures 7.2F-I**. Due to the concentration of the administered  $^*\text{gly}$  and the 120 min timeframe, most (~95%) of the glutathione in the tissue is unlabeled resulting in **Figure 7.2F** that shows a high PIE of unlabeled glutathione. The PIE maps for isotopologues  $^1\text{GSH}$ ,  $^2\text{GSH}$  and  $^3\text{GSH}$  are shown in **Figures 2G-I**. Similar data was obtained from 4T1 tumors excised from



**Figure 7.2.** A) Excised 4T1 tumor from MIST labeled mouse mounted on specimen disc for cryosectioning. MSI heatmaps depicting steady state abundances of B) glutathione ( $^0\text{GSH}$ ), C)  $^1\text{GSH}$  at  $m/z$  308.0769 D)  $^2\text{GSH}$  at  $m/z$  308.0832 and E)  $^3\text{GSH}$  at  $m/z$  309.0803. SIL abundances are corrected for isotopologue signal overlap. All heatmaps are normalized relative to the homoglutathione standard. PIE heatmaps of F) unlabeled glutathione,  $^0\text{GSH}$ , G)  $^1\text{GSH}$ , H)  $^2\text{GSH}$  and I)  $^3\text{GSH}$  detected. Region of interest analyzed with PIE tool is outlined. Scale bar = 1mm.

three mice treated with the identical MIST protocol and all show high levels of glutathione along the outer edge of the tumors and a central core with little to no glutathione detected (**Figure D.1**). The PIE data for each of these three tumors are shown in **Figure D.2**. In all tumors, the distribution of the three glutathione isotopologues is also heterogenous. The relative levels of glutathione isotopologues detected in the four tumors administered <sup>1</sup>gly, <sup>2</sup>gly, <sup>3</sup>gly are summarized by the box and whisker plots in **Figure 7.3** showing the PIE data collected from all tissue voxels in which isotopologues <sup>1</sup>GSH, <sup>2</sup>GSH, <sup>3</sup>GSH were detected. All tumors show relative PIE levels <sup>1</sup>GSH > <sup>2</sup>GSH > <sup>3</sup>GSH as expected for tissue exposed for 120, 80 and 60 min to the labeled glycine substrates. These data illustrate the ability of this fMSI method to generate three timepoint kinetic data from a single tissue sample. The GSH heatmap for a negative control tumor, treated with three doses of unlabeled glycine, also shows high levels of glutathione at the edges of the tumors, and a

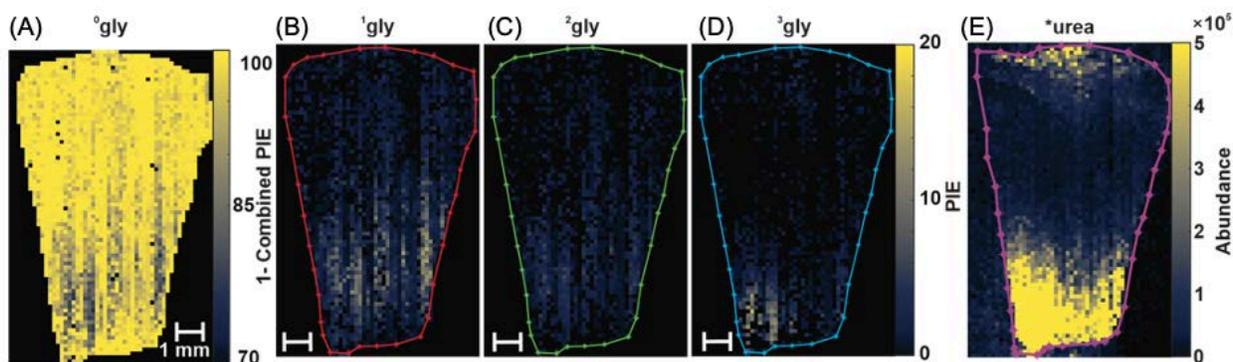


**Figure 7.3.** A) Box and whisker plots summarizing the PIE data from all detectable labeled isotopologues in the tumor shown in **Figure 7.2**. The number over each plot is the median PIE for the isotopologue data. (B)-(D) Box and whisker plots summarizing the \*GSH PIE data for the three 4T1 tumors shown in **Figures D.1/D.2**.

central region with low glutathione (**Figure D.3A**). Abundances of  $^1\text{GSH}$ ,  $^2\text{GSH}$ ,  $^3\text{GSH}$  were not detected above signals expected from the natural abundance of these isotopologues (**Figures D.3B**). The PIE data for this tumor shows high levels of unlabeled  $^0\text{GSH}$  (**Figure D.3E**), as expected, and no  $^1\text{GSH}$ ,  $^2\text{GSH}$ ,  $^3\text{GSH}$  (**Figures D.3F-H**).

### 7.3.2 Distribution and Uptake of Labeled Glycines

The relative rate of incorporation of isotope labels from glycine into glutathione is dependent upon delivery of the glycine isotopologues to the tissue which, in turn, is a function of how well the tissue is perfused. Perfusion of tumor tissue is known to be heterogeneous so mapping the uptake of the glycine species across the tumor provides an indication of the variation of the delivery of these substrates. Unlike the glutathione data shown in **Figure 7.3** and **Figures D.1/D.2**, mass spectrometry detection of glycine is most sensitive in positive ion mode. Therefore, a 20  $\mu\text{m}$  tumor section, collected within 100  $\mu\text{m}$  to the one used to collect the data in **Figure 7.2**, was used to collect the positive ion mode glycine data in **Figure 7.4**. **Figure 7.4A** shows the PIE distribution for unlabeled  $^0\text{gly}$  in the tissue. The MSI PIE maps of the glycine isotopologues are shown in **Figures 7.4B-D**. Similar to the  $^*\text{GSH}$  data, the glycine isotopologue uptake shown in the tumor section in **Figures 7.4B-D** is heterogeneous with the highest levels in the lower region



**Figure 7.4.** PIE heatmaps of: **A)** unlabeled glycine  $^0\text{gly}$ , **B)**  $^1\text{gly}$ , **C)**  $^2\text{gly}$  **D)**  $^3\text{gly}$  detected in a thin section of the 4T1 tumor adjacent to the section shown in Figure 2. **E)** MSI data showing the abundance heatmap of perfusion marker SIL-urea ( $^*\text{urea}$ ) detected in the same thin-section as in (**A-D**). Scale bar = 1mm.

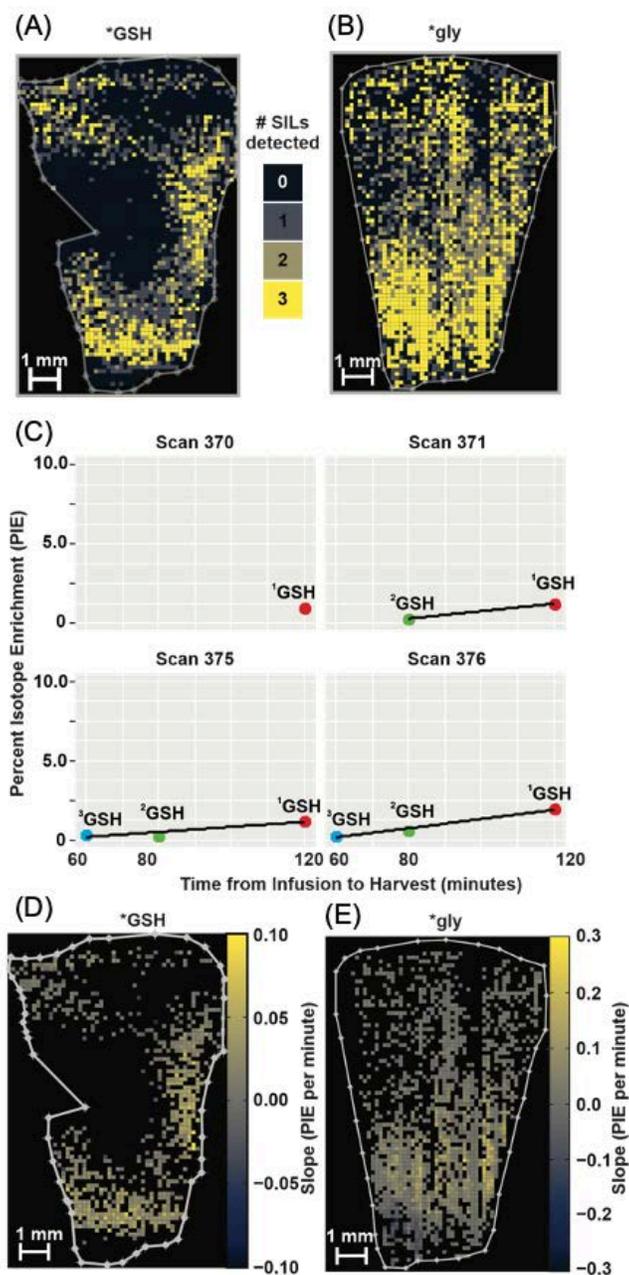
of the tissue. To confirm that **\*gly** uptake represents a measure of tissue perfusion, the positive ion MSI map of **\*urea** uptake is shown in **Figure 7.4E**. Since **\*urea** was administered one minute prior to tissue harvest, the uptake indicates the regions of highest perfusion in the tissue which correlates well with the regions indicating high glycine uptake. The data in **Figure 7.4** shows more extensive uptake of **<sup>1</sup>gly** followed by **<sup>2</sup>gly** and then **<sup>3</sup>gly**. The PIE maps of the glycine isotopologues and **\*urea** for the three other tumors studied are shown in **Figure D.4**.

### *7.3.3 Time Dependence of Label Uptake and Incorporation*

The preceding data indicates that glycine isotopologue uptake and conversion into glutathione is heterogeneous resulting in differences in the total number of isotopologues detected in each tissue voxel. **Figure 7.5A** shows the variation in the number of glutathione isotopologues found in each voxel. From the data mapped in this figure, the bottom half of the tumor shows a high concentration of voxels in which all three isotopologues were detected. The upper half, show a high level of voxels in which fewer than three isotopologues were detected. There is a large central region in which no glutathione isotopologues are detected. **Figure 7.5B** shows the map of the number of **\*gly** isotopologues that were detected in the adjacent tissue section that was extracted from the data shown in **Figure 7.4**. The glycine isotopologues are detected across more of the tissue than for glutathione when you compare **Figures 7.5A to 7.5B** showing substrate delivery was not limiting for the lack of glutathione production in some regions of the tumor. **Figure 7.5A** shows many voxels in which only one isotopologue was detected. Primarily **<sup>1</sup>GSH** was detected in these voxels. Voxels containing only one glycine isotopologue are less frequently observed (**Figure 7.5B**).

**Figure 7.5C** shows a random sampling of **\*GSH** PIE data from individual voxels (comprising the heat maps shown in **Figure 7.2** and boxplots in **Figure 7.3**) as a function of the

time interval over which each isotopologue was synthesized (i.e., difference between infusion and time of tumor harvest.) All three  $^*\text{GSH}$  isotopologues were detected in Scans 375 and 376 but



**Figure 7.5.** A) Map of the number of  $^*\text{GSH}$  isotopologues found at each tissue voxel. B) Map of the number of  $^*\text{gly}$  isotopologues detected in each tissue voxel. C)  $^*\text{GSH}$  PIE levels over time from sample voxels in the MSI data D) Heat map showing the slopes of the lines fitted to  $^*\text{GSH}$  PIE data from all voxels containing  $\geq 2$  data points. E) Heat map showing the slopes of the lines fitted to  $^*\text{gly}$  PIE data from all voxels containing  $\geq 2$  data points. Region of interest analyzed with PIE tool is outlined.

fewer in other voxels (e.g., Scans 370 and 371 in **Figure 7.5C**.) Even with only the one species detected (Scan 370), the PIE level of this isotopologue indicates active metabolism at this location. In many voxels, no enriched glutathione was detected. Regions in which <3 isotopologues were detected, are likely due to lower/intermittent tissue perfusion. For voxels in which 2 or more \*GSH species were detected, a line was fit to the data where the slope of this line provides an estimate of the rate of production of glutathione, i.e., PIE/min, from the labeled glycines in that voxel.

Heatmaps of the slope data, reflecting the PIE per minute for glutathione uptake of label from glycine in every tissue voxel are shown in **Figure 7.5D** for the tumor tissue data. In most tissue voxels in this tumor, positive slopes of varying magnitude were observed for the \*GSH data. The slopes were fit to tissue voxels containing two or more isotopologues (excluded voxels remain black.) The data in **Figure 7.5D** show greatest label uptake on the central right and lower left portion of the tissue. Interestingly, the upper left corner of the tumor showed both high steady-state glutathione levels (**Figure 7.2B**) and perfusion of all three labels (**Figure 7.5A**), but slopes of lower magnitude than other areas of the tumor. The slope data for glutathione PIE versus time for the other tumors data are shown in **Figure D.5**. Similar to **Figure 7.5D**, the data for label uptake into glutathione in these other tumors show positive slopes indicating labeling of the glutathione pool was increasing during the 120 min experimental interval in all tumors.

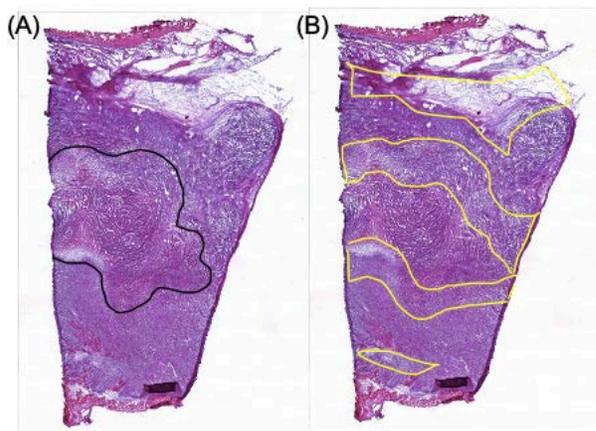
Slope data, reflecting the rate of glycine uptake into the tissue in all voxels are shown in **Figure 7.5E**. In this case, some regions of the tumor exhibit a negative slope for the glycine data, and these are mainly localized to the lower part of the tumor. Glycine slope data from the other tumors are shown in **Figure D.6**. Interestingly, the data in **Figure D.6A** primarily displays positive glycine slopes, whereas the data for tumor tissue in **Figure D.6B** are dominated by negative slopes illustrating the intertumoral heterogeneity in tissue perfusion. Since these are bolus injections, the

negative slope indicates the glycine substrates are in the washout phase in certain locations in the tissue. Theoretically, bolus wash-in and wash-out kinetics are not linear and more complex line fits are needed to account for the actual behavior of the dosed isotopic substrates.

The data in **Figure 7.5A** show many voxels in which zero, one or two \*GSH isotopologues are detected which is likely due to no or low perfusion (zero isotopologues detected) or intermittent perfusion (one or two isotopologues detected.) In support of this conclusion, we extracted data from our earlier study on liver tissue.<sup>12</sup> Since healthy liver is a well-perfused tissue, a MIST protocol using the same glycine isotopologues should detect all three glutathione isotopologues across the tissue. The data in **Figure D.7** confirms that all three \*GSH isotopologues are found across most of the tissue.

#### *7.3.4 Necrosis and Neutrophil Infiltration Identified by Tumor Histology*

Additional tumor tissue thin sections obtained from the tissues were stained with hematoxylin-eosin and the data for the section obtained from the tumor shown in **Figure 7.2** is shown in **Figure 7.6**. A region of tissue outlined in black in **Figure 7.6A** was identified as a region



**Figure 7.6.** **A)** Hematoxylin-eosin-stained tissue thin-section obtained from the tumor shown in **Figure 7.2A**. The black line indicates a region of coagulative necrosis. **B)** The same section with yellow lines indicating regions of the tissue with increased neutrophil infiltration.

of coagulative necrosis. This region corresponds to levels of little to no detectable unlabeled glutathione and the absence of glutathione isotopologues indicating no metabolic activity (**Figure 7.5D**). However, compared to the lack of glutathione isotopologues in this region, there are a number glycine isotopologues present, although this region has a high density of voxels with less than three glycine isotopologues (**Figure 7.5B**). In addition to identifying tissue necrosis, several regions of the tissue had a high level of neutrophil infiltration indicative of localized inflammation and are delineated by yellow borders in **Figure 7.6**. Two areas of increased neutrophil infiltration border on the central necrotic region. All regions displaying neutrophil infiltrations all show a low rate of isotope incorporation into glutathione (**Figure 7.5A**).

#### **7.4 Discussion and Conclusions**

Measures of metabolic activity or flux reflect the net outcome of the interactions of genes, proteins and metabolites under regulatory control operating within the local tissue microenvironments. Glutathione is a key player in the development, progression, and therapy response in cancer<sup>4, 6-19</sup> and as such, a measure of its functional activity will likely be a better indicator of cancer aggressiveness and predictor of treatment outcome than steady-state levels. In addition, the innate metabolic and microenvironmental heterogeneity of tumor tissue requires this functional activity be assessed across the tissue. This is accomplished through functional imaging methods which often use isotope tracking technologies. Perhaps the best-known functional imaging method used in clinical oncology is positron emission tomography (PET) to detect glucose delivery, uptake, and phosphorylation. To date there is no PET probe to detect glutathione metabolic activity. We developed a magnetic resonance method to detect glutathione metabolism noninvasively in rodent tumor tissues<sup>9,10</sup> but the low sensitivity of the method limited the amount of kinetic and heterogeneity data that could be obtained. An additional advantage of some

noninvasive imaging methods is the capability to provide multiple time point kinetics. However noninvasive methods such as PET and magnetic resonance can be limited in the metabolic pathways that can be probed and often yield images with poor spatial resolution.

Classic isotope tracing experiments can be used to track many more metabolic pathways, but these methods are often invasive, and a commonly used approach is to administer a single tracer and harvest tissue samples at multiple time intervals after infusion. This works well in some settings but is not readily amenable to MSI analysis as it is not possible to obtain data from identical tissue spatial locations in each serial sampling. This is especially true in tumor tissue which is comprised of multiple cell types and tissue environments with few, if any, anatomical features in which to orient similar sampling sites. The MIST protocol replaces a single tracer infusion and multiple tissue harvests approach with multiple tracer administration and a single tissue harvest to provide kinetic data. This approach has been used previously to follow protein biosynthesis<sup>13,14,20</sup> and glutathione metabolic rates<sup>21</sup> in tissue extracts. The combination of MIST with MSI that is the basis of this fMSI method demonstrated here is not as elegant as many of the noninvasive functional imaging methods available but offers some distinct advantages. First, the MSI method can unambiguously identify and map multiple molecular species and pathways beyond the capabilities of most other imaging methods. Second, immunohistochemistry analysis of adjacent thin sections can be used to correlate classic tissue biomarkers to functional metabolic activities. Third, tissue perfusion measures can be acquired at each spatial location that is necessary to assessing variations in metabolic activity. Fourth, many MSI methods provide much higher spatial resolution than methods such as PET or functional magnetic resonance. Finally, a single-sample MIST approach precludes the need for multiple tissue samples, or animal subjects, that is frequently needed in acquiring dynamic data.

Similar to our magnetic resonance results, the fMSI data shows higher glutathione labeling located near the outer periphery of the tumor. However, our magnetic resonance method, although noninvasive, had poor sensitivity and spatial resolution. Due to this poor sensitivity, label incorporation could only be detected at a single timepoint. We were later able to use magnetic resonance to noninvasively monitor the time dependence of glutathione synthesis in rat and human liver but were not able to map variations in metabolic activity differences across the tissue.<sup>22</sup> The value of obtaining multiple timepoints as demonstrated by this fMSI method is that the trend in the metabolic kinetics may be determined at each spatial location. For example, the data indicates that incorporation of isotope-labeled glycine is still increasing in most tissue regions in the 4T1 tumor over the 120 min time course of the experiment shown in **Figure 7.5** but is in a washout phase in other regions of the tumor. A single timepoint reading would miss these differences. Glycine isotopologue wash-in or wash-out will also affect glutathione kinetics further highlighting the need for developing models of label transit through the glutathione metabolic network. The metabolic activity data in this study was obtained through linear fits of the PIE data as a function of time and is a convenient way to visualize the data. However, the uptake and metabolism of the labeled substrates would be more accurately represented by a more complex function. Particularly in this study, where bolus doses are administered, label wash-in and wash-out kinetics need to reflect both substrate uptake and washout and glutathione synthesis and consumption. The MIST procedure has previously been used with both a bolus injection<sup>20</sup> and continuous infusion<sup>13, 14,21</sup> protocols and both these approaches can provide unique data on dynamic biological processes. We are currently working on different infusion protocols and timing schemes and developing more physiologically accurate metabolic models to fit the data and incorporating this into the fMSI

analysis software. However, these results also emphasize that perfusion must be considered when assessing metabolic heterogeneity.

There are regions in all four tumors where glutathione is absent but glycine isotopologues are present indicating the glutathione metabolic machinery is not active in these regions. Perfusion differences detected by variations in glycine isotopologue uptake were supported by unique MSI data that mapped isotope-labeled urea uptake. Urea is a metabolite found at high concentration in most tissues and rapidly diffuses from the blood and is quickly taken up by tissue.<sup>23</sup> The use of isotope-labeled urea in perfusion imaging was demonstrated in hyperpolarized magnetic resonance studies<sup>15</sup> and was adopted in this study for use with MSI. This method readily detected well-perfused regions of tumor tissue. In contrast, poorer perfusion is detectable in some tumor regions by low uptake of the three glycine isotopologues. The data also shows that timed MIST protocol may be capable of detecting intermittent perfusion as only one or two glutathione isotopologues were detected in some regions. Re-examination of our liver data supports the conclusion that intermittent perfusion may be responsible for detection of fewer than three isotopologues in the tissue. Intermittent perfusion, leading to cyclical variations in tissue oxygen levels, is often observed in regions of tumor tissue<sup>24</sup> and may be a driver of tumor progression<sup>25</sup> so detection and analysis of such tumor fractions is critical assessing the metabolic consequences of limited blood flow. The single point data shown in **Figure 7.5C** (i.e., Scan 370) shows intermittent perfusion does not stop label incorporation into glutathione. How these data are fit into a metabolic model, or visually displayed, is still under consideration. In addition, fMSI experiments to better characterize regions of intermittent perfusion are planned. Although the slopes of the linear fits to the enrichment data are an approximation of the metabolic rates, they do illustrate the

heterogeneity in both glycine uptake and metabolism into glutathione and demonstrate the value of the fMSI technique using MIST infusion protocols.

The MIST infusion protocol shown in **Figure 7.1C** worked well in mapping metabolic and perfusion variations in the 4T1 tumor. However, our initial attempts to study glutathione metabolism in the 4T1 tumor used a MIST timing protocol identical to that used in our liver studies<sup>12</sup> resulted in no detectable uptake of isotope labels into glutathione. In the liver study, the glycine isotopologues were administered at time points of 60, 30 and 10 min and MSI detected all three glutathione isotopologues across the liver tissue. For the 4T1 tumor, glutathione metabolic activity is slower, so a longer MIST protocol totaling 120 min with appropriate time separation between doses was necessary to allow time for label incorporation and to generate the data shown. Therefore, the overall length of MIST infusion protocol and intervals between doses needs to be tailored to the metabolic rates to yield data reflecting the dynamics of the system.

This study utilized an IR-MALDESI source paired with orbitrap-based mass analysis to generate imaging data with a spatial resolution of 150  $\mu\text{m}$ . This is a far higher resolution than available from some functional imaging methods such as PET. There are a number of other MSI ionization and detection methods that can map isotope uptake with  $<1 \mu\text{m}$  spatial resolution.<sup>11</sup> Those methods, paired with MIST infusions, offer the possibility of obtaining single-cell or subcellular functional data. Advanced molecular profiling methods have already collected reams of data at the single cell to whole-tissue level and the addition of fMSI methods to the analytical arsenal over these size scales can provide unprecedented data on the spatial relationships between molecular components and biological function; the objective of the field of spatial biology.<sup>26</sup> We will continue to probe differences in glutathione metabolic activity in tumor tissue to establish functional measures in tumors from different tissue types and levels of aggressiveness. This

communication focuses on the ability of fMSI to measure glutathione metabolism, but this method is amenable to any process in which multiple isotopologues of a metabolic substrate are available.

### **7.5 Acknowledgements**

All IR-MALDESI-MSI experiments were completed in the Molecular Education, Technology and Research Innovation Center (METRIC) at North Carolina State University. The authors gratefully acknowledge the financial support received from the National Institutes of Health (R21GM134219).

## 7.6 References

- (1) Fiaschi, T., Chiarugi, P. Oxidative stress, tumor microenvironment, and metabolic reprogramming: a diabolic liaison. *International Journal of Cell Biology*, **2012**, 2012, 762825.
- (2) Gandhi, N., Das, G.M., Metabolic Reprogramming in Breast Cancer and Its Therapeutic Implications. *Cells*, **2019**, 8 (2), 89.
- (3) Schiliro, C., Firestein, B.L., Mechanisms of Metabolic Reprogramming in Cancer Cells Supporting Enhanced Growth and Proliferation, *Cells*, **2021**, 10 (5), 1056.
- (4) Kennedy, L., Sandhu, J.K., Harper, M.E., Cuperlovic-Culf, M. Role of Glutathione in Cancer: From Mechanisms to Therapies. *Biomolecules*. **2020**, 10 (10), 1429.
- (5) Gamcsik, M.P., Kasibhatla, M.S., Teeter, S.D., Colvin, O.M. Glutathione levels in human tumors. *Biomarkers*. **2012**, 17(8), 671-91.
- (6) Niklas, J., Heinzle, E. Metabolic flux analysis in systems biology of mammalian cells. *Adv Biochem Eng Biotechnol*. **2012**, 127, 109-32.
- (7) Lagziel, S., Lee, W.D. & Shlomi, T. Studying metabolic flux adaptations in cancer through integrated experimental-computational approaches. *BMC Biol*. **2019**, 17, 51.
- (8) Badur, M.G., Metallo, C.M. Reverse engineering the cancer metabolic network using flux analysis to understand drivers of human disease. *Metab Eng*. **2018**, 45, 95-108.
- (9) Thelwall, P. E., Simpson, N. E., Rabbani, Z. N., Clark, M. D., Pourdeyhimi, R., Macdonald, J. M., Blackband, S. J., Gamcsik, M. P. In vivo MR studies of glycine and glutathione metabolism in a rat mammary tumor. *NMR in biomedicine*, **2012**, 25(2), 271–278.

- (10) Thelwall, P. E., Yemin, A. Y., Gillian, T. L., Simpson, N. E., Kasibhatla, M. S., Rabbani, Z. N., Macdonald, J. M., Blackband, S. J., & Gamcsik, M. P. Noninvasive in vivo detection of glutathione metabolism in tumors. *Cancer research*, **2005**, *65*(22), 10149–10153.
- (11) Mellinger, A.L., Muddiman, D.C., Gamcsik, M.P., *J. Proteome Res.*, **2022**, in press.
- (12) Mellinger, A.L, Garrard, K.P., Khodjanizyazova, S., Rabbani, Z.N., Gamcsik, M.P., Muddiman, D.C. Multiple Infusion Start Time Mass Spectrometry Imaging of Dynamic SIL-Glutathione Biosynthesis Using Infrared Matrix-Assisted Laser Desorption Electrospray Ionization, *J. Proteome Res.*, **2022**, *21*(3), 747-757.
- (13) Dudley, M. A., Burrin, D. G., Wykes, L. J., Toffolo, G., Cobelli, C., Nichols, B. L., Rosenberger, J., Jahoor, F., & Reeds, P. J. Protein kinetics determined in vivo with a multiple-tracer, single-sample protocol: application to lactase synthesis. *Am. J. Physiol.*, **1998**, *274*(3), G591–G598.
- (14) Paolini, C. L., Marconi, A. M., Pike, A. W., Fennessey, P. V., Pardi, G., & Battaglia, F. C. A multiple infusion start time (MIST) protocol for stable isotope studies of fetal blood. *Placenta*, **2001**, *22*(2-3), 171–176.
- (15) von Morze, C., Larson, P. E., Hu, S., Keshari, K., Wilson, D. M., Ardenkjaer-Larsen, J. H., Goga, A., Bok, R., Kurhanewicz, J., & Vigneron, D. B. Imaging of blood flow using hyperpolarized [(13)C]urea in preclinical cancer models. *J. Magn. Reson. Imaging*, **2011**, *33*(3), 692–697.
- (16) Bansal, A., Simon, M. C. Glutathione metabolism in cancer progression and treatment resistance. *The Journal of cell biology*, **2018**, *217*(7), 2291–2298.

- (17) Peiris-Pagès, M., Martinez-Outschoorn, U. E., Sotgia, F., & Lisanti, M. P. Metastasis and Oxidative Stress: Are Antioxidants a Metabolic Driver of Progression? *Cell metabolism*, **2015**, 22(6), 956–958.
- (18) Grek, C. L., & Tew, K. D. Redox metabolism and malignancy. *Current opinion in pharmacology*, **2010**, 10(4), 362–368.
- (19) Hayes, J. D., Dinkova-Kostova, A. T., & Tew, K. D. Oxidative Stress in Cancer. *Cancer Cell*, **2020**, 38(2), 167–197.
- (20) Zhang, X. J., Chinkes, D. L., & Wolfe, R. R. Measurement of muscle protein fractional synthesis and breakdown rates from a pulse tracer injection. *American journal of physiology. Endocrinology and Metabolism*, **2002**, 283(4), E753–E764.
- (21) Agostini, F., Dalla Libera, L., Rittweger, J., Mazzucco, S., Jurdana, M., Mekjavic, I. B., Pisot, R., Gorza, L., Narici, M., & Biolo, G. Effects of inactivity on human muscle glutathione synthesis by a double-tracer and single-biopsy approach. *The Journal of Physiology*, **2010**, 588(Pt 24), 5089–5104.
- (22) Skamarauskas, J. T., Oakley, F., Smith, F. E., Bawn, C., Dunn, M., Vidler, D. S., Clemence, M., Blain, P. G., Taylor, R., Gamcsik, M. P., & Thelwall, P. E. Noninvasive in vivo magnetic resonance measures of glutathione synthesis in human and rat liver as an oxidative stress biomarker. *Hepatology*, **2014**, 59(6), 2321–2330.
- (23) Iwamoto, K., Watanabe, J., Imai, K., Shibata, M., Hirate, J., Ozeki, S., Disposition of urea following intravenous administration to rats, *Chem Pharm Bull (Tokyo)*, **1982**, 30(4), 1422-9.
- (24) Dewhirst M. W. Relationships between cycling hypoxia, HIF-1, angiogenesis and oxidative stress. *Radiation Research*, **2009**, 172(6), 653–665.

- (25) Saxena, K., & Jolly, M. K. Acute vs. Chronic vs. Cyclic Hypoxia: Their Differential Dynamics, Molecular Mechanisms, and Effects on Tumor Progression. *Biomolecules*, **2019**, 9(8), 339.
- (26) Palla, G., Fischer, D. S., Regev, A., & Theis, F. J. Spatial components of molecular tissue biology. *Nature Biotechnology*, **2022**, 40(3), 308–318.

## **CHAPTER 8: Investigating Glycerate Production from Intestinal Fructose Metabolism**

### **Elevated by Dietary Fat in Mouse Pancreas using Mass Spectrometry Imaging**

This work was previously published in: Yanru, W., Wong, C.W., Chiles, E.N., Mellinger, A.L., Bae, H., Jung, S., Peterson, T., Wang, J., Negrete, M., Xiang, K., Huang, Q., Wang, L., Jang, C., Muddiman, D.C., Su, X., Williamson, I., Shen, X., “Glycerate Production from Intestinal Fructose Metabolism Elevated by Dietary Fat Induces Glucose Intolerance Through  $\beta$ -cell Damage,” *Cell Metabolism*, **2022**, *34*, 7, 1042-1053, e6. Copyright © Elsevier Inc.

### **8.1 Introduction**

Type 2 diabetes mellitus (T2DM) has doubled in prevalence over the past two decades, becoming a major threat to global health.<sup>1,2</sup> Although T2DM is associated with genetics, the development of glucose intolerance is also a consequence of lifestyle and dietary patterns.<sup>1,3,4</sup> Ample evidence associates the consumption of fructose-containing sweetened beverages<sup>5,6</sup> and dietary fat,<sup>7,8</sup> prominent components of the western diet, with the development of T2DM. Pathogenic effects of fructose are shown to be more prominent among mice fed a high-fat diet, although the mechanisms remain unknown.<sup>9</sup>

Recent reports demonstrated that the intestine largely shields the liver and other internal organs from the effects of fructose consumption, such as enhanced lipogenesis and ectopic fat deposition, by converting a substantial amount of ingested fructose to various metabolites.<sup>10,11</sup> Only when ingested fructose exceeds the maximum fructose metabolism capacity of the small intestine does unconverted fructose “spillover” to the liver (via the portal vein) and the colon.<sup>12</sup> These findings indicate that the metabolism of fructose is not localized to the liver and that a systematic evaluation of its fate is warranted. However, little is known about the contribution of

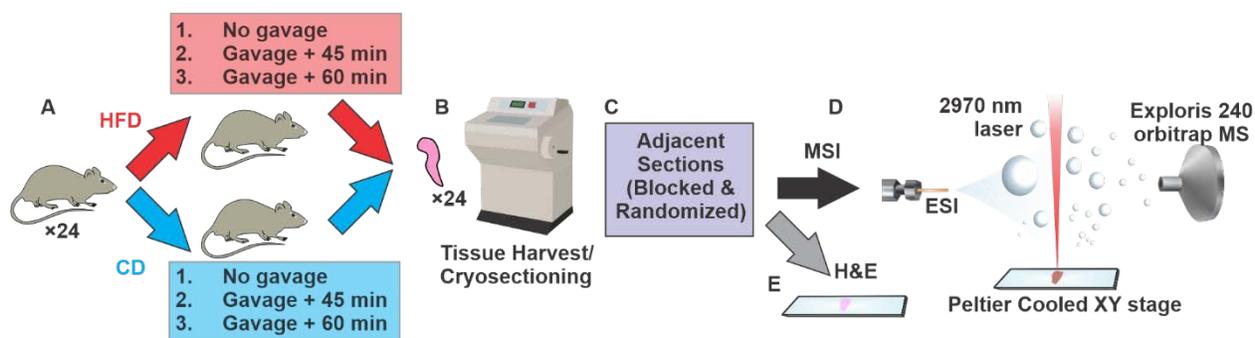
fructose-derived metabolites from the small intestine during the development of T2DM, particularly in a liver-independent fashion.

Wong and coworkers observed that a high-fat diet (HFD) altered fructose metabolism in the small intestine of mice, resulting in an enrichment of organic acid pools from the fructolysis pathway but not the citric acid cycle. Fructose derived glycerate specifically was sustained at high levels in systemic circulation, and chronic glycerate treatment caused damage to pancreatic islet cells, especially  $\beta$  cells, *in vivo*. These data suggested that fat consumption enhances intestinal fructose metabolism and circulation of its derived glycerate, which can chronically damage pancreatic islet cells and lead to subsequent glucose intolerance. The model presented provides a mechanism linking T2DM pathogenesis to the consumption of a western diet rich in fructose and fat.

The goal of this abbreviated report is to expand on the use of mass spectrometry imaging to support the findings presented in the listed work by Wong et. al. Using mass spectrometry imaging (MSI) via infrared matrix assisted laser desorption electrospray ionization (IR-MALDESI) and stable isotope labeling (SIL) we measured differences in the deposition of fructose derived glycerate within pancreas tissues of mice fed a high fat and a control diet.

## 8.2 Materials and Methods

**Figure 8.1** summarizes the workflow followed for the mass spectrometry imaging analysis of the pancreas tissues. Twenty-four pancreas tissues were imaged, with four replicates per the six conditions studied: mice fed a HFD and gavaged with  $U^{13}C$ -fructose and harvested after 45/60 minutes, mice fed a control diet (CD) and gavaged with  $U^{13}C$ -fructose and harvested after 45/60 minutes, or mice fed either diet but without a fructose gavage. Further details can be found within the stated reference for this work.



**Figure 8.1.** **A)** 24 mice were split into 6 conditions, receiving either a high fat diet or a control diet along with SIL-fructose administered via gavage 45 or 60 minutes prior to sacrifice, or no gavage. **B)** Pancreas tissues were harvested and sectioned to even thickness prior to thaw mounting on microscope slides. **C)** Two sections were saved per pancreas for **D)** MSI analysis and **E)** H&E staining for islet counts. Sections were blocked for analysis days and randomized prior to analysis.

### 8.2.1 Mass spectrometry imaging by IR-MALDESI-MSI

Pancreas tissues were first equilibrated to  $-15^{\circ}\text{C}$  then sectioned to  $20\ \mu\text{m}$  thickness using a Leica CM1950 cryostat (Buffalo Grove, IL, USA). Cut sections were then thaw-mounted on clean microscope slides (1 mm height, plain, Fisher Scientific, Pittsburgh, PA) and stored at  $-80^{\circ}\text{C}$  until IR-MALDESI-MSI analysis. Tissues were first blocked into four groups, each containing one replicate of each condition. Tissues were both cut and imaged in randomized order within these blocks to minimize sampling bias. Additionally, jejunum tissue from a fructose-gavaged mouse was analyzed in a similar fashion to validate the usage of M+3 glycerate as a tracer of fructose-derived glycerate.

Mass spectrometry imaging (MSI) was performed using the infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI) coupled to an Exploris 240 Orbitrap mass spectrometer (Thermo Scientific, Bremen, Germany) to achieve high resolution and accurate mass measurements. To facilitate desorption of targeted analytes, a thin layer of ice was applied to the surface each tissue section by first purging the source enclosure with nitrogen gas to 10% humidity,

stabilizing the sample stage at -8 °C, and exposing the stage to ambient humidity until the ice matrix formed<sup>13</sup>. The enclosure was then resealed and purged again until humidity returned to 10%, where it was maintained throughout analysis.

To ablate targeted tissue regions a 2970-nm wavelength laser was used with a single burst of ten pulses to produce 1 mJ of energy at a rate of 10 kHz. X and Y stage movements of 100  $\mu\text{m}$  were used to achieve oversampling.<sup>14</sup> Ablated analytes were post-ionized in the orthogonal electrospray plume, established by applying a voltage of approximately 3 kV to the electrospray solvent (1 mM acetic acid in 50:50 water/acetonitrile.) Mass spectrometry analysis of ionized molecules was performed in negative mode with internal calibrant used to achieve high mass accuracy (<2.5 ppm) within the 85-225  $m/z$  range. Automatic gain control (AGC) was disabled. A mass resolution power of 240,000<sub>FWHM</sub> at 200  $m/z$  was used with a fixed injection time (15 ms) to synchronize timing of the ablation plume with ion collection in the C-trap of the mass spectrometer.

### 8.2.2 Data Collection and Analysis

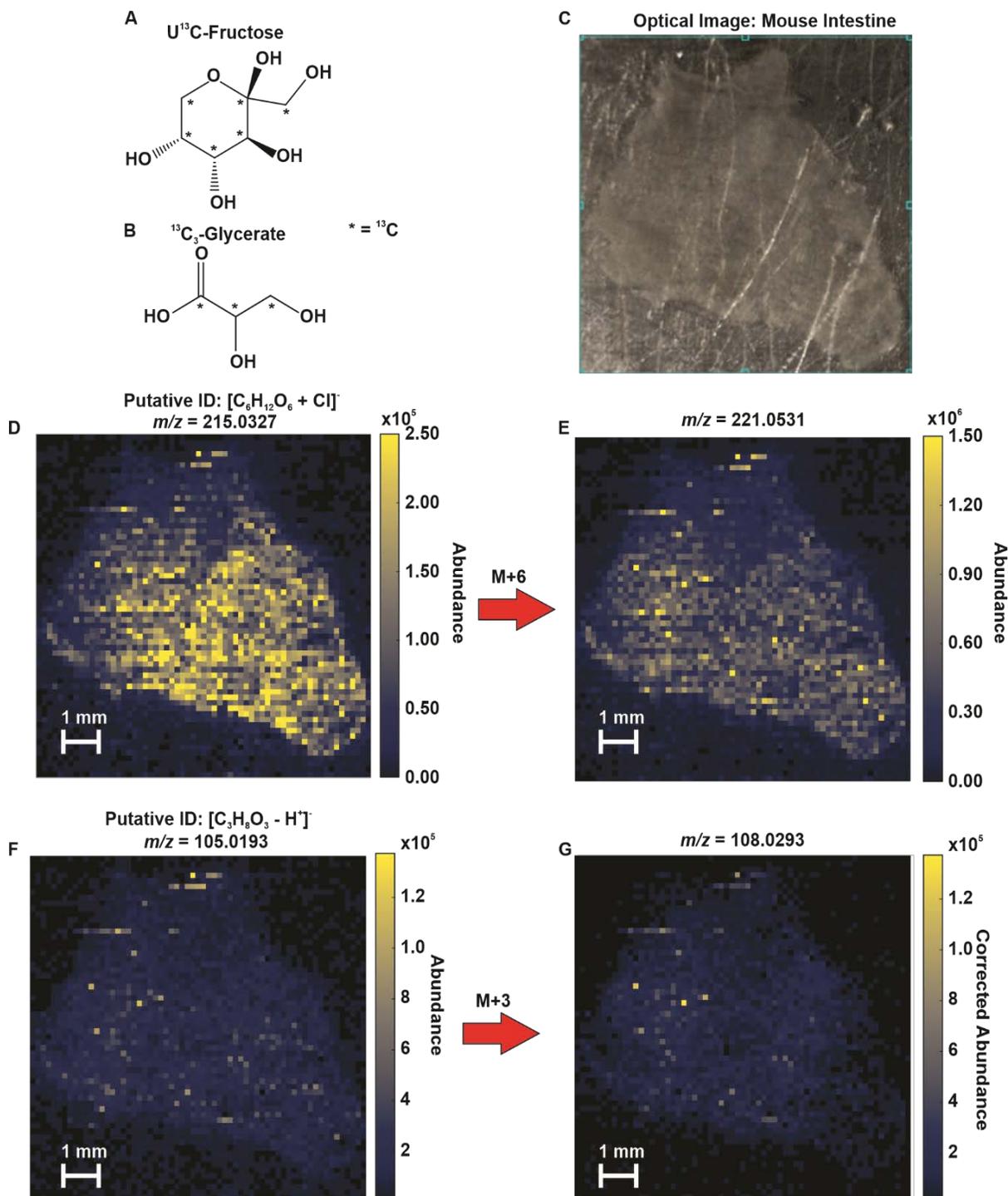
Collected raw data files were converted from the XCalibur.RAW format to mzML files via the ProteoWizard tool, MSConvert and then to imzML format using imzMLConverter. Images were generated from the imzML files in MSiReader v1.03c.<sup>15</sup> For MSI data, abundances of stable isotope-labeled peaks (M+3 glycerate, 108.0293  $m/z$ ) were corrected for potential signal overlap from the naturally abundant species of each molecule using the percent isotope enrichment (PIE) tool in MSiReader.<sup>16</sup> Isotopic distributions were computed using the R-based IsoSpec package.<sup>17</sup> Tissue regions of interest were selected to include pancreas tissue only and exclude peripheral fat tissue found in adjacent tissue staining when counting glycerate peaks.

## 8.3 Results and Discussion

### 8.3.1 Validating SIL-Fructose and SIL-Glycerate Detected in Mouse Intestine

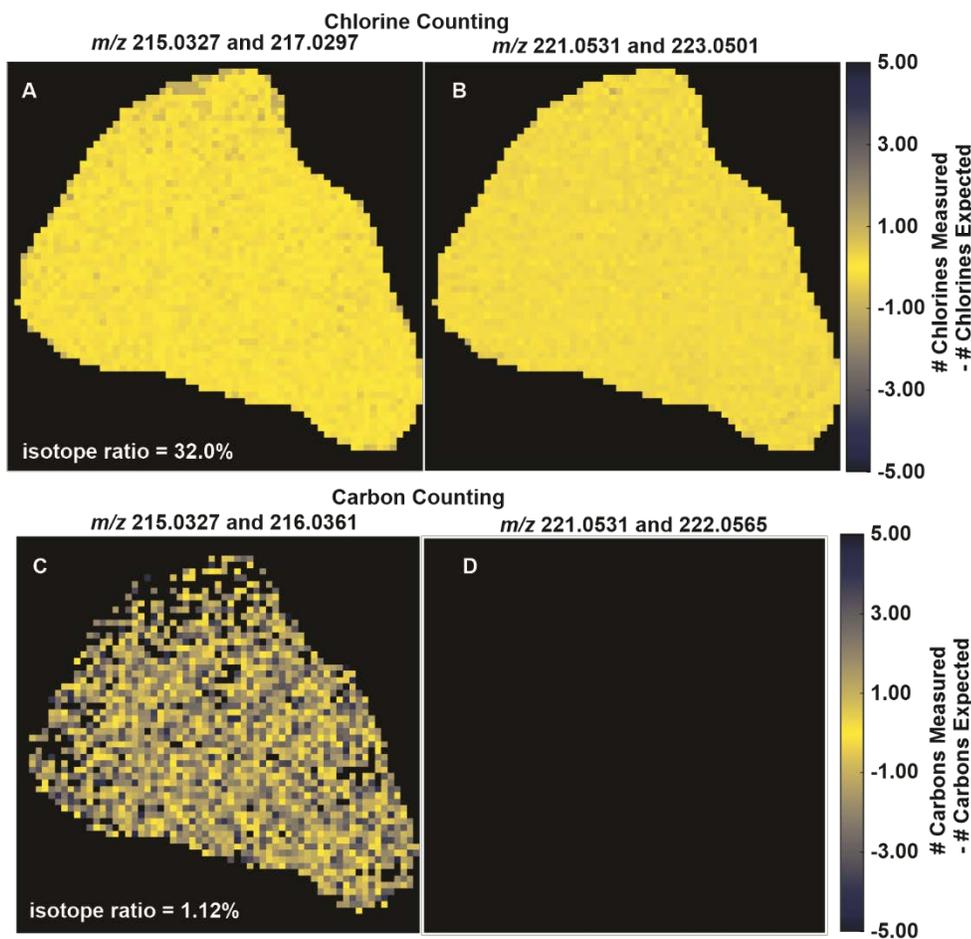
Prior to investigation of pancreatic tissues, a mouse was injected with the U<sup>13</sup>C-fructose standard (structure shown in **Figure 8.2A**) to validate the use of M+3 glycerate (**Figure 8.2B**) as a tracer of the fructose-derived glycerate. The intestinal tissue (specifically the jejunum tissue, **Figure 8.2C**) was imaged using IR-MALDESI-MSI as previously described. We successfully detected the chlorine adducted ion of C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>, which was putatively identified as glucose (an isomer of fructose) in the mouse tissue based on the high mass accuracy of the MS1 peak *m/z*. This data is shown in **Figure 8.2D** and allowed us to initially assess the data collected via IR-MALDESI-MSI. We then searched for the M+6 peak of fructose and found that it was highly abundant across this tissue (almost an order of magnitude more abundant than the unlabeled putative glucose peak in **D**), indicating successful arrival of SIL-fructose to the tissue. We note that this data was not corrected for abundance overlap. However, the probability of the naturally occurring M+6 isotopologue of glucose is very low (approximately 3E-13) and would therefore not affect this data.

We were then able to detect a high abundance of both the naturally abundant glycerate (**Figure 8.2E**) and the SIL-glycerate (M+3, **Figure 8.2F**) in this tissue. This allowed us to move forward with using this molecule as a tracer of fructose-derived glycerate in tissues from the treated mice. To remain extremely conservative, the abundance of the fructose-derived glycerate was corrected for any potential overlap in signal due to the natural abundance of the M+3 isotopologue of glycerate and other potential isomers.<sup>16</sup> While this is highly unlikely based on the probability of a <sup>13</sup>C occurring within a 3-carbon molecule, the purpose of this was to remain conservative in our assumptions.



**Figure 8.2.** **A)** Chemical structure of SIL-fructose with six <sup>13</sup>C isotopes labeled. **B)** Expected SIL structure of SIL-fructose-derived glycerate. **C)** Optical image of intestine section mounted on microscope slide. Heatmaps showing relative abundances in intestine of **D)** unlabeled glucose (expected *m/z* of unlabeled fructose) **E)** SIL-fructose, **F)** naturally abundant (unlabeled) glycerate and **G)** glycerate. SIL-glycerate abundance was corrected for signal overlap of naturally abundant glycerate.

To validate our putative identifications of the six-carbon sugar, we utilized the spectral accuracy tool in MSiReader (**Figure 8.3.**) It should be noted that the  $[M-H]^+$  species were also detected in this tissue, but with lower abundances. Additionally, molecules containing chlorine create very distinct isotopic distributions due to the high abundance of the  $^{37}\text{Cl}$  isotope, resulting in a highly abundant  $M+2$  peak. By using the chlorine adducts of these molecules, we were able to validate our putative identifications using spectral accuracy. **Figure 8.3A/B** shows the error in chlorine estimations across the intestine tissue for both the glucose molecule and the  $M+6$  fructose molecule, respectively. This data indicates that estimations are within  $\pm 0.5$  chlorine atoms and increases our confidence in this putative identification. We also performed carbon counting to



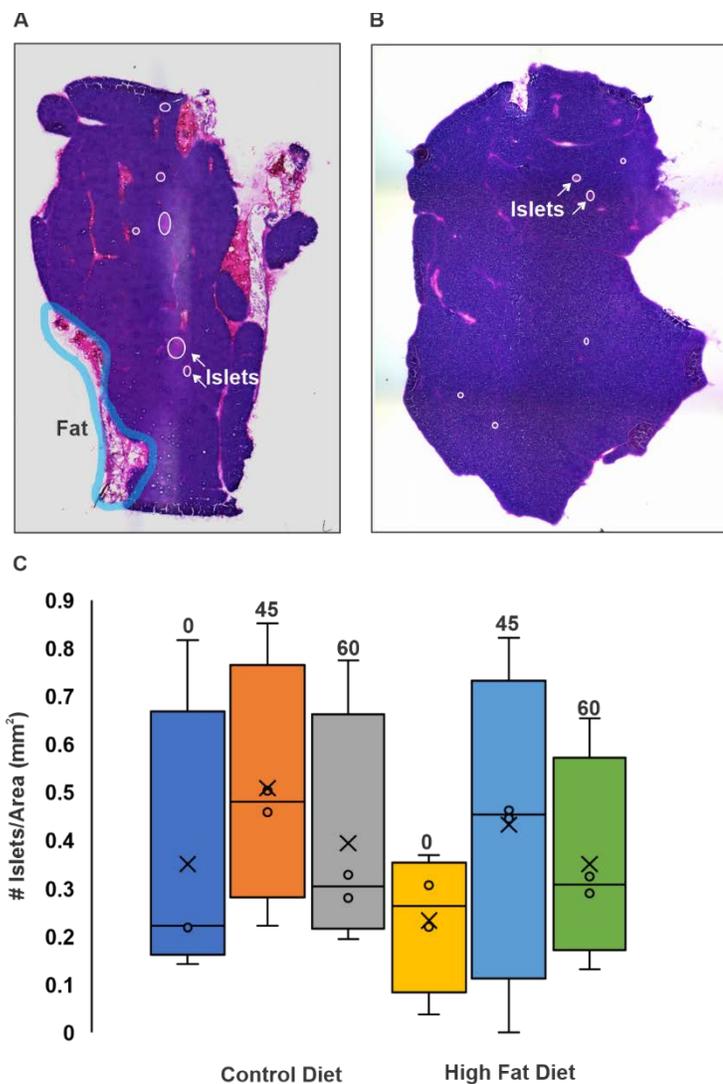
**Figure 8.3.** Spectral accuracy in mouse intestine tissue. Chlorine counting for **A)** naturally abundant glucose and **B)** SIL-fructose. Carbon counting for **C)** naturally abundant glucose and **D)** SIL-fructose.

further increase our confidence in this identification, although the probability of  $^{13}\text{C}$  isotopes is much lower than that of the  $^{37}\text{Cl}$  and therefore should result in a lower M+1 peak. **Figure 8.3C** shows the error in the carbon estimation for the unlabeled peak. Since the SIL-fructose peak was universally labeled, we would not expect an M+1 peak due to a naturally occurring  $^{13}\text{C}$ . This was confirmed by the data in **Figure 8.3D**.

With only three carbons, we did not find spectral accuracy to be a useful validation tool in the glycerate data. However, this peak was not detected in negative control tissues of preliminary experiments (data not shown). Additionally, by checking the METLIN database and METASPACE peak identifications, we were able to rule out other MS1 identifications for this peak furthering our confidence in the identifications.

### 8.3.2 Islet Estimation in Pancreas Tissues

The interest of this project and other data included, was to link glycerate deposition and damage to islets. Prior to MSI analysis, adjacent tissue sections were stained and investigated for both islet cells and fat content. Due to the increased deposition of fructose in fatty tissues and the lack of fat across every tissue analyzed, these areas were excluded from SIL-glycerate deposition counts. **Figure 8.4A** shows an example of fat tissue that was excluded from one tissue section. This prevented skewing results towards higher deposition counts for tissues with fat deposits within the sections and allowed specifically endocrine tissue to be investigated. We also performed a count of islet cells across tissue sections to account for any bias towards numbers of islet cells in the areas of the tissue sections analyzed. **Figures 8.4A and 8.4B** shows examples of islets in stained tissues (circled in white). **Figure 8.4C** shows boxplots of the islet count distributions within each condition analyzed. This data is normalized to pancreas area and shows a similar distribution was collected from each of the six groups.

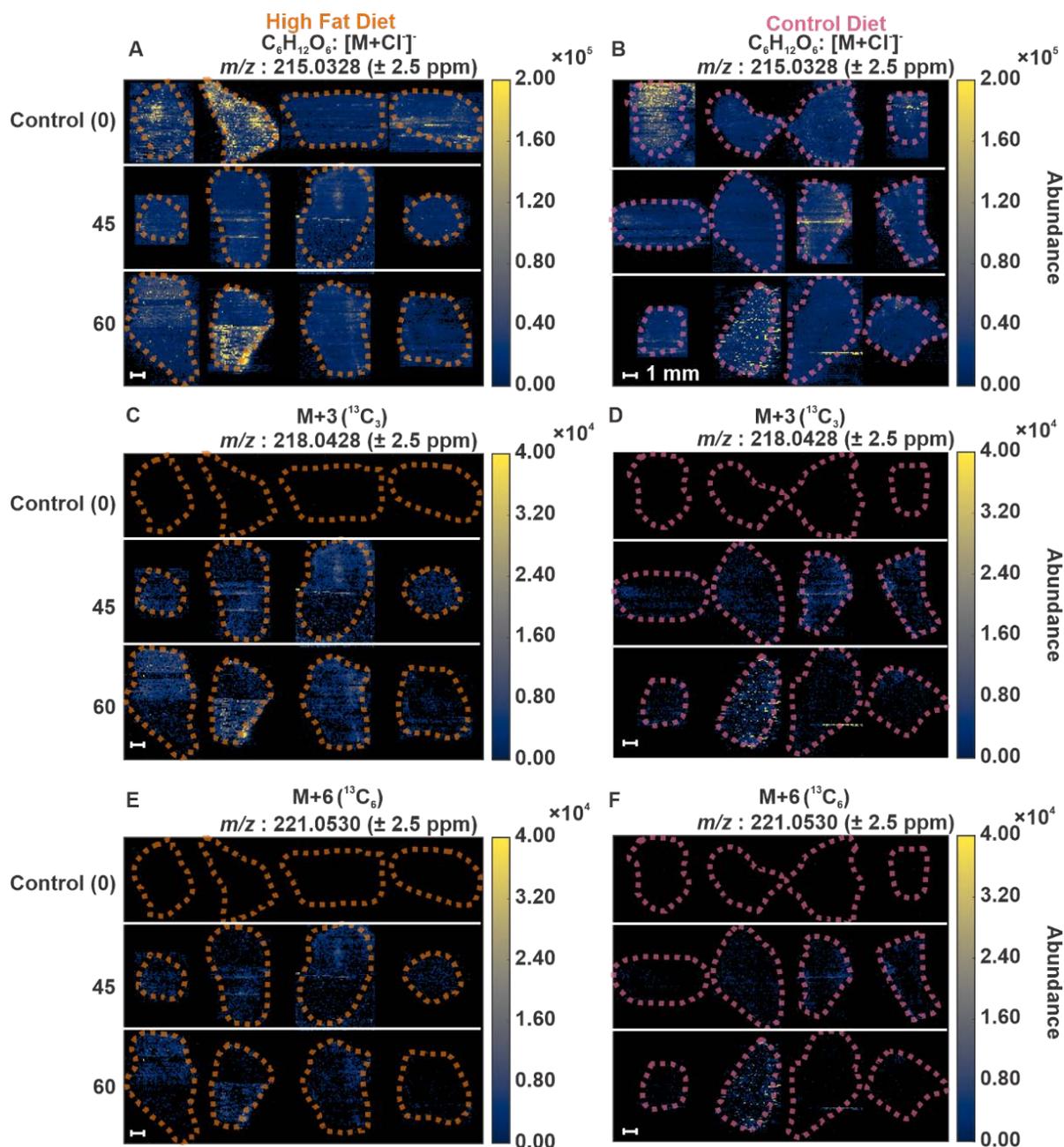


**Figure 8.4.** A-B) Optical images of stained pancreas tissue sections. Identified islets are circled in white and fatty tissue is circled in blue. C) Boxplot showing islet counts normalized to the measured area of each pancreas section analyzed.

### 8.3.3 SIL-Fructose Deposition in Gavage Mouse Pancreas Sections

The mouse pancreas sections (adjacent to stained tissues) were then analyzed by IR-MALDESI-MSI. SIL-fructose and SIL-glycerate were targeted in these studies, performed in negative mode. Similar to the positive control intestine tissue, we were able to detect the unlabeled chlorine adduct of  $C_6H_{12}O_6$ , putatively identified as glucose. We detected this peak in all 24

replicates and across all six conditions (**Figure 8.5A/B**). The high abundance of this molecule in tissues that were not labeled with the SIL-fructose gavage allowed us to assess the quality of pancreas data collected via IR-MALDESI. The M+3 species (i.e., three carbons were labeled with  $^{13}\text{C}$ ,  $m/z$  218.0428) was detected across pancreas sections from gavage labeled mice. As shown by



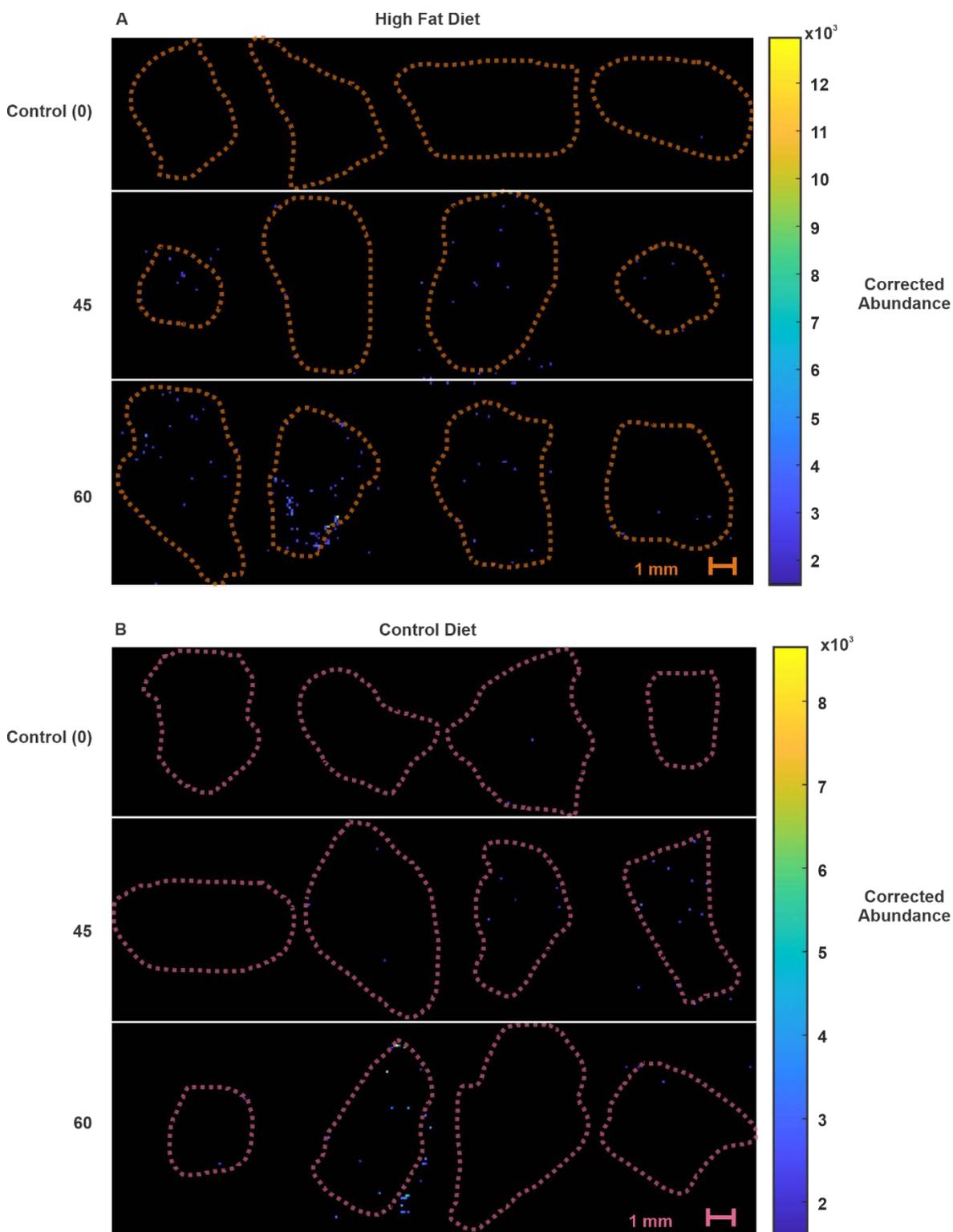
**Figure 8.5.** Heatmaps displaying abundances at  $m/z$ 's of **A-B**) naturally abundant glucose/fructose, **C-D**) metabolized SIL-fructose (M+3) and **E-F**) the infused SIL-fructose (M+6) in all analyzed pancreas tissues.

Jang et al. this is likely the result of three-carbon units of SIL-fructose being metabolized in the small intestine into glucose before entering both portal vein and systemic circulation.<sup>12</sup> No visible difference between the spatial distributions of deposition in mice fed a HFD (**Figure 8.5C**) and mice fed a CD (**Figure 8.5D**) was observed. Importantly, no signal from this peak was detected in the pancreas of the mice that were not given the SIL gavage. Abundances of the SIL-fructose (M+6) appeared slightly lower in abundance across tissues compared to the putatively identified M+3 glucose, likely due to the metabolism of this exogenous compound. However, it was still detected across all gavage-labeled groups within the two diets and remained absent in the negative control tissue sections (**Figure 8.5E and F**).

#### *8.3.4 Increased Deposition of SIL-Fructose Glycerate in HFD Mouse Pancreas*

Differences in the deposition of SIL-glycerate tracer in the mouse pancreas tissues was investigated between groups. We first performed an abundance correction to remain conservative in our detection counts using the PIE tool. Therefore, any peaks due to the high abundance of the unlabeled and natural isotopologues of glycerate (and other potential isomers) would be excluded. This data is shown in **Figure 8.6**, with regions of interest outlined, pancreas from HFD fed mice (**A**) in orange and CD (**B**) in pink. As expected, we did not observe deposition in the unlabeled tissues across both groups. These “dark” heatmaps are shown at in the top rows of **Figure 8.6A and 8.6B**.

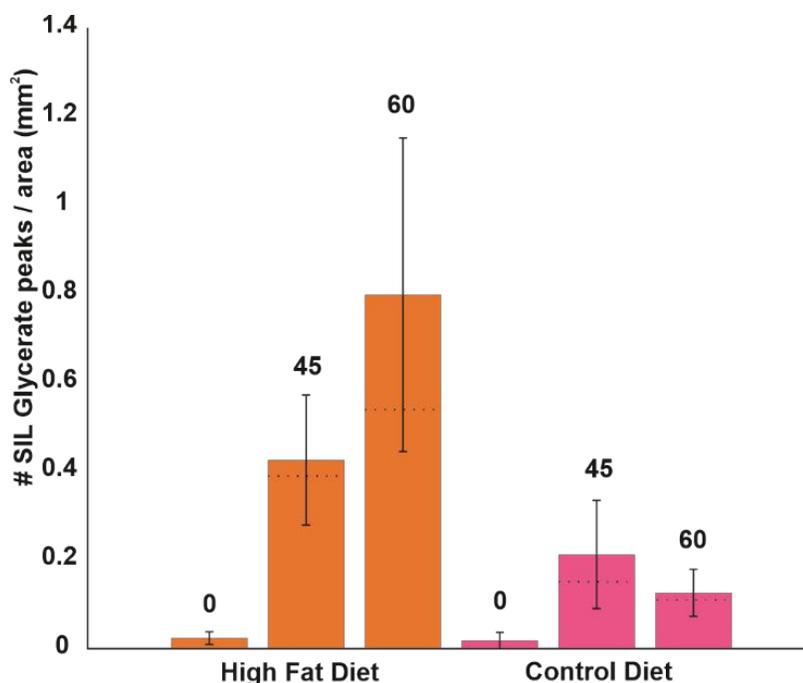
Unlike the intestine tissue, detected SIL-glycerate peaks in the pancreas tissue are expected from the result of “spillover” from the intestine, and therefore are not deposited across the entirety of the tissue. Therefore, we were more interested in detection frequency across these tissues and chose to use the higher contrasting “viridis” scale for these heatmaps. From these heatmaps, we



**Figure 8.6.** Abundance heatmaps of SIL-fructose derived glycerate ( $M+3$ ,  $m/z$  108.0293) detected in analyzed pancreas tissues. Abundances were corrected for overlapping signal from naturally abundant glycerate. **A)** Data from mice fed a high fat diet (tissue edges outlined in orange.) **B)** Data from mice fed a control diet (tissue edges outlined in pink.)

were not able to visibly observe any difference between tissues within the diets sacrificed at the different time points (45 vs 60 minutes.) Since our interest was specifically in the deposition of this metabolite in pancreatic tissue, any fat tissue (identified via the H&E stain, the optical image, and increased abundances of the fructose/glucose species) was excluded from these counts. For both diets, the most obvious example of this is in the 60-minute sample replicate that is second from the left (bottom row, **Figure 8.6A and 8.6B.**) Aside from these samples where glycerate deposited more frequently in the fatty tissue, there was no visible pattern of deposition across the tissues.

While it appeared that the high fat diet groups had a higher number of peaks in all tissues, this was a subjective conclusion and therefore prompted more substantial quantitative evidence. Hence, we counted the number of detections made in each tissue and normalized these counts to the area of tissue analyzed. We then took the average of these normalized counts within each condition block for comparison. This data is shown in **Figure 8.7.** The HFD (left) data is represented by the orange bars while the CD is represented in pink bars. From this data, we confirmed that the deposition of the SIL-fructose derived M+3 glycerate peaks was much higher in the HFD fed pancreas tissues. However, the HFD- 60-minute group and the CD-45-minute group were both influenced by outliers. For this reason, a dotted black line is included with each bar to also show the median of each group. While overall relationships remained the same when comparing the medians between diets (HFD mice had higher deposition than CD) the differences between timings decreases within the groups when comparing medians.



**Figure 8.7.** Bar graph showing the average numbers of SIL-glycerate peaks detected by group normalized to pancreas tissue areas. A dotted black bar represents the median of each group. Solid black bar = standard error. High fat diet data is shown in orange (left) and control diet data is shown in pink (right).

## 8.4 Conclusions

Using mass spectrometry imaging of fructose-derived glycerate in mouse pancreas tissues, we were able to support the findings presented by Wong et al. We traced the deposition of M+3 glycerate in pancreas from mice fed either a HFD or a CD and found that mice fed a HFD had an increased deposition of the labeled metabolite in their endocrine tissue.

## 8.5 Acknowledgements

All IR-MALDESI measurements were completed in the Molecular Education, Technology and Research Innovation Center (METRIC) at North Carolina State University. The authors gratefully acknowledge the financial support received from the National Institutes of Health (R01GM087964.)

## 8.6 References

- (1) Bhupathiraju, S. N.; Hu, F. B. Epidemiology of Obesity and Diabetes and Their Cardiovascular Complications. *Circulation Research* **2016**, *118* (11), 1723–1735. <https://doi.org/10.1161/CIRCRESAHA.115.306825>.
- (2) Unnikrishnan, R.; Pradeepa, R.; Joshi, S. R.; Mohan, V. Type 2 Diabetes: Demystifying the Global Epidemic. *Diabetes* **2017**, *66* (6), 1432–1442. <https://doi.org/10.2337/db16-0766>.
- (3) Wang, H.; Wang, J.; Liu, M.-M.; Wang, D.; Liu, Y.-Q.; Zhao, Y.; Huang, M.-M.; Liu, Y.; Sun, J.; Dong, G.-H. Epidemiology of General Obesity, Abdominal Obesity and Related Risk Factors in Urban Adults from 33 Communities of Northeast China: The CHPSNE Study. *BMC Public Health* **2012**, *12* (1), 967. <https://doi.org/10.1186/1471-2458-12-967>.
- (4) Nguyen, D. M.; El-Serag, H. B. The Epidemiology of Obesity. *Gastroenterology Clinics of North America* **2010**, *39* (1), 1–7. <https://doi.org/10.1016/j.gtc.2009.12.014>.
- (5) Hannou, S. A.; Haslam, D. E.; McKeown, N. M.; Herman, M. A. Fructose Metabolism and Metabolic Disease. *J Clin Invest* **2018**, *128* (2), 545–555. <https://doi.org/10.1172/JCI96702>.
- (6) Cordain, L.; Eaton, S. B.; Sebastian, A.; Mann, N.; Lindeberg, S.; Watkins, B. A.; O’Keefe, J. H.; Brand-Miller, J. Origins and Evolution of the Western Diet: Health Implications for the 21st Century. *The American Journal of Clinical Nutrition* **2005**, *81* (2), 341–354. <https://doi.org/10.1093/ajcn.81.2.341>.
- (7) Forouhi, N. G.; Krauss, R. M.; Taubes, G.; Willett, W. Dietary Fat and Cardiometabolic Health: Evidence, Controversies, and Consensus for Guidance. *BMJ* **2018**, *361*, k2139. <https://doi.org/10.1136/bmj.k2139>.
- (8) Soltis, A. R.; Kennedy, N. J.; Xin, X.; Zhou, F.; Ficarro, S. B.; Yap, Y. S.; Matthews, B. J.; Lauffenburger, D. A.; White, F. M.; Marto, J. A.; Davis, R. J.; Fraenkel, E. Hepatic

- Dysfunction Caused by Consumption of a High-Fat Diet. *Cell Reports* **2017**, *21* (11), 3317–3328. <https://doi.org/10.1016/j.celrep.2017.11.059>.
- (9) Softic, S.; Gupta, M. K.; Wang, G.-X.; Fujisaka, S.; O'Neill, B. T.; Rao, T. N.; Willoughby, J.; Harbison, C.; Fitzgerald, K.; Ilkayeva, O.; Newgard, C. B.; Cohen, D. E.; Kahn, C. R. Divergent Effects of Glucose and Fructose on Hepatic Lipogenesis and Insulin Signaling. *J Clin Invest* **2017**, *127* (11), 4059–4074. <https://doi.org/10.1172/JCI94585>.
- (10) Jang, C.; Wada, S.; Yang, S.; Gosis, B.; Zeng, X.; Zhang, Z.; Shen, Y.; Lee, G.; Arany, Z.; Rabinowitz, J. D. The Small Intestine Shields the Liver from Fructose-Induced Steatosis. *Nat Metab* **2020**, *2* (7), 586–593. <https://doi.org/10.1038/s42255-020-0222-9>.
- (11) Kim, M.; Astapova, I. I.; Flier, S. N.; Hannou, S. A.; Doridot, L.; Sargsyan, A.; Kou, H. H.; Fowler, A. J.; Liang, G.; Herman, M. A. Intestinal, but Not Hepatic, ChREBP Is Required for Fructose Tolerance. *JCI Insight* **2** (24), e96703. <https://doi.org/10.1172/jci.insight.96703>.
- (12) Jang, C.; Chen, L.; Rabinowitz, J. D. Metabolomics and Isotope Tracing. *Cell* **2018**, *173* (4), 822–837. <https://doi.org/10.1016/j.cell.2018.03.055>.
- (13) Robichaud, G.; Barry, J. A.; Muddiman, D. C. IR-MALDESI Mass Spectrometry Imaging of Biological Tissue Sections Using Ice as a Matrix. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (3), 319–328. <https://doi.org/10.1007/s13361-013-0787-6>.
- (14) Nazari, M.; Muddiman, D. C. Cellular-Level Mass Spectrometry Imaging Using Infrared Matrix-Assisted Laser Desorption Electrospray Ionization (IR-MALDESI) by Oversampling. *Anal Bioanal Chem* **2015**, *407* (8), 2265–2271. <https://doi.org/10.1007/s00216-014-8376-5>.

- (15) Bokhart, M. T.; Nazari, M.; Garrard, K. P.; Muddiman, D. C. MSiReader v1.0: Evolving Open-Source Mass Spectrometry Imaging Software for Targeted and Untargeted Analyses. *J. Am. Soc. Mass Spectrom.* **2018**, 29 (1), 8–16. <https://doi.org/10.1007/s13361-017-1809-6>.
- (16) Mellinger, A. L.; Garrard, K. P.; Khodjaniazova, S.; Rabbani, Z. N.; Gamcsik, M. P.; Muddiman, D. C. Multiple Infusion Start Time Mass Spectrometry Imaging of Dynamic SIL-Glutathione Biosynthesis Using Infrared Matrix-Assisted Laser Desorption Electrospray Ionization. *J. Proteome Res.* **2021**,. <https://doi.org/10.1021/acs.jproteome.1c00636>.
- (17) Łacki, M. K.; Valkenborg, D.; Startek, M. P. IsoSpec2: Ultrafast Fine Structure Calculator. *Anal. Chem.* **2020**, 92 (14), 9472–9475. <https://doi.org/10.1021/acs.analchem.0c00959>.

**APPENDICES**

**Appendix A: Supplemental Information for Chapter 2**

## A.1 Supplemental Table

**Table A.1** Summary of raw file sources for building CSF spectral library. In-house generated raw files from a previous study were also utilized.

Proteome Xchange Identifier	Title	Repository	Instrument	Publishing Author	Principal Investigator	Date	Digestion Enzyme	Sample	Appendix Reference
PXD009623	Characterization of cerebrospinal fluid via data-independent acquisition mass spectrometry	PRIDE	Q Exactive	Barkovits et al. (2018)	Katrin Marcus	2/14/2019	Trypsin	CSF	[1, 2]
PXD008098	Quantitative proteomic profiling of cerebrospinal fluid to identify candidate biomarkers for Alzheimer's disease	PRIDE	Q Exactive	Sathe et al. (2018)	Akhilesh Pandey	1/10/2019	Trypsin; Lys C	Depleted CSF	[3, 4]
PXD010705	Characterization of cerebrospinal fluid via data-independent acquisition mass spectrometry	PRIDE	Q Exactive	Barkovits et al. (2018)	Katrin Marcus	9/17/2018	Trypsin	CSF	[1, 5]
PXD010690	Characterization of cerebrospinal fluid via data-independent acquisition mass spectrometry	PRIDE	Q Exactive	Barkovits et al. (2018)	Katrin Marcus	9/17/2018	Trypsin	Substantia Nigra	[1, 6]
PXD010708	Characterization of cerebrospinal fluid via data-independent acquisition mass spectrometry	PRIDE	Q Exactive	Barkovits et al. (2018)	Katrin Marcus	9/17/2018	Trypsin	Fractionated CSF	[1, 6]
PXD010698	Characterization of cerebrospinal fluid via data-independent acquisition mass spectrometry	PRIDE	Q Exactive	Barkovits et al. (2018)	Katrin Marcus	9/17/2018	Trypsin	CSF	[1, 7]

**Table A.1** (continued.)

PXD006357	Cerebrospinal fluid markers to distinguish bacterial meningitis from cerebral malaria in children	MassIVE	Q Exactive	James M. Njunge (2017)	James M. Njunge	4/24/2017	Trypsin	CSF	[6, 8]
PXD004572	CSF comparison MS and controls	PRIDE	Q Exactive	Kroksveen et al. (2016)	Frode S. Berven	10/13/2016	Trypsin	Depleted CSF	[10, 11]
PXD003075	Human CSF peptidomics of gamma secretase inhibition	PRIDE	Q Exactive	Höltkä et al. (2016)	Johan Gobom	3/14/2016	N/A	Endogeneous CSF peptides	[12, 13]
PXD000608	Proteomics of cerebrospinal fluid microvesicles	PRIDE	Q Exactive	Chiasserini et al. (2014)	Connie R. Jiménez	5/7/2014	Trypsin	CSF microvesicles	[14, 15]
PXD001654	Identification of a biomarker in cerebrospinal fluid for neuronopathic forms of Gaucher disease	PRIDE	Q Exactive	Zigdon et al. (2015)	Yishai Levin	6/24/2016	Trypsin	CSF	[12, 16]
PXD007842	Proteomics analysis of biological fluids of sleeping sickness patients	PRIDE	Q Exactive	Bonnet et al. (2018)	Bertrand Courtioux	11/15/2018	Trypsin; Lys C	Depleted CSF	[17, 18]

## A.2 References

- (1) Barkovits K, Linden A, Galozzi S, Schilde L, Pacharra S, Mollenhauer B, Stoepel N, Steinbach S, May C, Uszkoreit J, Eisenacher M, Marcus K Characterization of Cerebrospinal Fluid via Data-Independent Acquisition Mass Spectrometry. *J Proteome Res.* **2018**, *17*, 3418-3430.
- (2) ProteomeXchange Dataset PXD009623.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD009623>. Accessed 7 Feb 2020
- (3) Sathe G, Na CH, Renuse S, Madugundu AK, Albert M, Moghekar A, Pandey A. Quantitative Proteomic Profiling of Cerebrospinal Fluid to Identify Candidate Biomarkers for Alzheimer's Disease. *PROTEOMICS-Clinical Applications.* **2018**, *13*, 1800105.
- (4) ProteomeXchange Dataset PXD008098.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD008098>. Accessed 7 Feb 2020
- (5) ProteomeXchange Dataset PXD010705.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD010705>. Accessed 7 Feb 2020.
- (6) ProteomeXchange Dataset PXD010690.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD010690>. Accessed 7 Feb 2020
- (7) ProteomeXchange Dataset PXD010698.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD010698>. Accessed 7 Feb 2020

- (8) ProteomeXchange Dataset PXD006357.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD006357>. Accessed 7 Feb 2020
- (9) Njunge JM, Oyaro IN, Kibinge NK, Rono MK, Kariuki SM, Newton CR, Berkley JA, Gitau EN. Cerebrospinal fluid markers to distinguish bacterial meningitis from cerebral malaria in children. *Wellcome Open Res.* **2017**, *2*, 47.
- (10) ProteomeXchange Dataset PXD004572.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD004572>. Accessed 7 Feb 2020
- (11) Kroksveen AC, Guldbrandsen A, Vaudel M, Lereim RR, Barsnes H, Myhr K-M, Torkildsen Ø, Berven FS. In-Depth Cerebrospinal Fluid Quantitative Proteome and Deglycoproteome Analysis: Presenting a Comprehensive Picture of Pathways and Processes Affected by Multiple Sclerosis. *J Proteome Res.* **2017**, *16*, 179-194.
- (12) Hölttä M, Dean RA, Siemers E, Mawuenyega KG, Sigurdson W, May PC, Holtzman DM, Portelius E, Zetterberg H, Bateman RJ, Blennow K, Gobom J. A single dose of the  $\gamma$ -secretase inhibitor semagacestat alters the cerebrospinal fluid peptidome in humans. *Alzheimers Res Ther.* **2016**, *8*, 11.
- (13) ProteomeXchange Dataset PXD003075.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD003075>. Accessed 7 Feb 2020
- (14) ProteomeXchange Dataset PXD000608.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000608>. Accessed 7 Feb 2020.

- (15) Chiasserini D, van Weering JRT, Piersma SR, Pham TV, Malekzadeh A, Teunissen CE, de Wit H, Jiménez CR. Proteomic analysis of cerebrospinal fluid extracellular vesicles: A comprehensive dataset. *Journal of Proteomics*. **2014**, *106*, 191-204.
- (16) ProteomeXchange Dataset PXD001654.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD001654>. Accessed 7 Feb 2020
- (17) ProteomeXchange Dataset PXD007842.  
<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD007842>. Accessed 7 Feb 2020
- (18) Bonnet J, Garcia C, Leger T, Couquet M-P, Vignoles P, Vatunga G, Ndung'u J, Boudot C, Bisser S, Courtioux B. Proteome characterization in various biological fluids of *Trypanosoma brucei* gambiense-infected subjects. *J Proteomics*. **2019**, *196*, 150–161.

## Appendix B: Supplemental Information for Chapter 3

### B.1 R Code for Mixed-Effects Modeling of Longitudinal ALS patient peptide measurements

```

require(ggplot2)
require(nlme)
require(ggcorrplot)
library(ISLR)
library(glmnet)
library(dplyr)
library(tidyr)
library(caret)
library(mgcv)
library(MASS)

#Upload data; Skyline file used after manual filtering for quantitative peptides
## Outlier patient and QC data removed for this analysis
### EX: PeptideData = as.data.frame(read.csv("All_peps_r.csv", header = TRUE))

#Filter any peptides that were missing in 25% of the samples
Peptide_Data_filtered <- PeptideData[, !colSums(PeptideData !=0) < 45]

#Create data frame containing only peak areas for log transformation
abundances2 <- as.data.frame(Peptide_Data_filtered[,12:3208])
abundances2 <-log2(abundances2)
abundances2[abundances2 < 0] <- 0

#Use caret function to return vector of columns to remove to reduce pair-wise correlations
highCorr<- findCorrelation(correlations, cutoff = 0.6)
highCorrmatrix <-as.matrix(highCorr)
filtered <- abundances2[,-c(highCorr)]

#Combine newly logged and filtered data with meta data
logT_PeptideData<- cbind(PeptideData[,1:11], filtered[,1:96])

#Create matrices containing explanatory and response variables for LASSO
X2 <- as.matrix(cbind(logT_PeptideData[,6],logT_PeptideData[,8],logT_PeptideData[,12:107]))
Y2 <- logT_PeptideData$ALSFRS.Total.Score

# Run LASSO analysis

peptide_lasso <- glmnet(X2,Y2,family='gaussian',alpha=1)
peptide_lasso

#99% of the lasso deviation explained with s = 0.04012 with a reasonably narrowed list of peptide coefficients contributing to the model.

```

```

##Created a matrix and separate .csv containing these coefficients to filter "unim
ortant" (coefficient = 0) peptides

peptide_coefficients_0.04012 <- as.matrix(coef(peptide_lasso, x = X2, y = Y2, s = 0.0
4012))
write.csv(peptide_coefficients_0.04012, file = "peptide_coefficients_s_0.04012.csv")

#Filter peptides (put them in the model alone and selected based on the lowest AIC) b
elonging to the same protein and combine with meta data for import and analysis

##EX: PeptideData = as.data.frame(read.csv("All_peps_rmm_r.csv", header = TRUE))

###Create data frame containing only peak areas for log transformation
#### Recombine with the meta data for analysis
abundances2 <- as.data.frame(PeptideData[,12:66])
abundances2 <-log2(abundances2)
abundances2[abundances2 < 0] <- 0
logT_PeptideData<- cbind(PeptideData[,1:11], abundances2[,1:55])

#Generate the mixed model using the lme function in the nlme package
##Used Patient "Subject" identification for random effect
### This is the null model. Does not contain
Model <- lme(ALSFRS.Total.Score ~ 1 ,
             random = ~1|Subject.ID,data=logT_CleanPeptideData, method = "REML")

#Possible analyses
summary(Model)
plot(Model)
ranef(Model)
plot(ranef(Model))
AIC(Model)

#Performed forward selection by adding peptides to minimize AIC as described in text

Final_Model <- lme(ALSFRS.Total.Score ~Disease.Duration..Years.from.symptom.onset.to.
sample.collection. +
  Onset.Site +
  LGADMEDVCGR.1 +
  LGMFNIQHCK.1 +
  LSSLQAGTK +
  SVEGQNFLSEK +
  YLYDVLR +
  EVEELMEDTQHK +
  TVLQNEGTK +
  WVFGGGTK +
  FTFHADICTLSEK +
  EVDSGNDIYGNIPIK +
  ISVTQGASK +
  IISVELPGDAK +
  MALDIEIATYR +
  TTPMLSDSGSFFLYSK +
  EHIVDLEMLTVSSIGTFR +
  GILCSALQK +
  LVESGGGLVKPGGSLR +
  VPTADLEDVLRPLAEDITNILSK +

```

```

ESNPHCSDGQTYGNK +
IGEWIGR +
SVVPTAAAVSLK +
QGQGQSEPGEYEQR +
AGLAASLAGPH +
AELQEGAR +
CNMGYEYSER +
YVMLPVADQDQCIR +
STTAVVTNPK+
HYLLR,
random = ~1|Subject.ID,data=logT_CleanPeptideData, method = "REML")

```

Summary(Final\_Model)

**Table B.1.** Summary of peptides used for global inflammation study

Protein Accession	Gene	Protein Name	Peptide Sequence
P00739	HPR	Haptoglobin related protein	TEGDGVYTLNDK
			LPECEAVCGKPK
			SCAVAEYGVYVK
P00734	F2	Prothrombin	HQDFNSAVQLVENFCR
			TATSEYQTFNPR
			GQPSVLQVNLPIVERPVCK
P02652	APOA2	Apolipoprotein A-II	EPCVESLVSQYFQTVTDYGK
			SPELQAEAK
			SKEQLTPLIK
P02745	C1QA	Complement C1q subcomponent subunit A	PSGNPGKVGYPGSPGLGAR
			SLGFCDTTNK
P08571	CD14	Monocyte differentiation antigen CD14	TTPEPCELDDEDFR
			ELTLEDLK
			VLDLSCNR
P02763	ORM1	Alpha-1-acid glycoprotein 1	WFYIASAFR
			TEDTIFLR
			NWGLSVYADKPETTK
P02751	FN1	Fibronectin	ISCTIANR
			TFYSCTTEGR
			LGVRPSQGGEAPR
O75882	ATRN	Attractin	LTGSSGFVTDGPGNYK
			CTWLIEGQPNR
			SEAACLAAGPGIR
P02765	AHSG	Alpha-2-HS-glycoprotein	APHGPGLIYR
			HGPGLIYR
			HTFMGVVSLGSPSEVSHPR

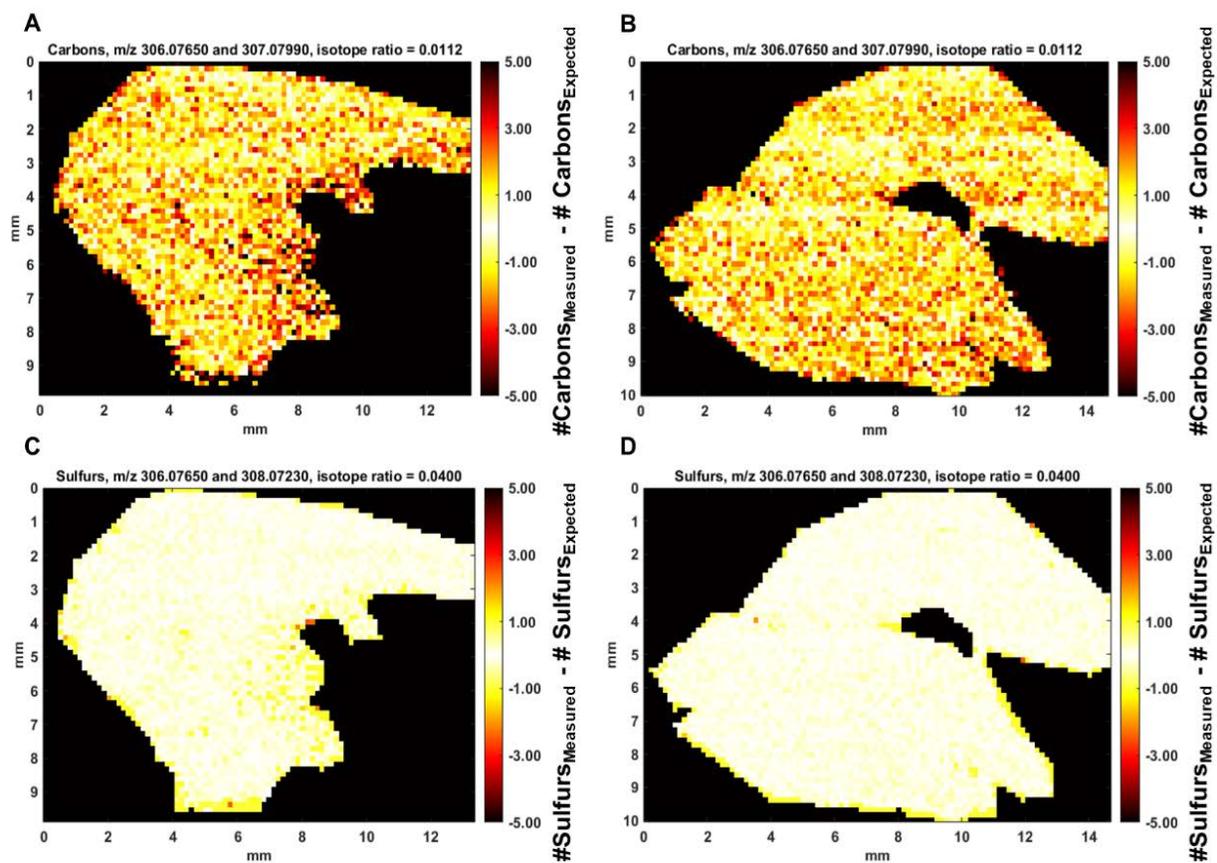
**Table B.1** (continued.)

O75326	SEMA7A	Semaphorin-7A	SEGLLACGTNAR
			AMLVCSDAATNK
			YYLSCPMESSR
P30530	AXL	Tyrosine-protein kinase receptor UFO	TATITVLPQQPR
			APLQGTLLGYR
P01008	SERPINC1	Antithrombin-III	VAEGTQVLELPFK
			GDDITMVLILPKPEK
			IEDGFSLK
P00748	F12	Coagulation factor XII	CFEPQLLR
			NPDNDIRPWCFVLNR
			CLQDRPAPEDLTVVLGQER
P01042	KNG-1	Kininogen-1	TVGSDFYSFK
			KYFIDFVAR
			KLGQSLDCNAEVYVPWEK
Q16610	ECM1	Extracellular matrix protein 1	LLPAQLPAEK
			FSCFQEEAPQPHYQLR
			LTFINDLCGPR
P01009	SERPINA1	Alpha-1-antitrypsin	DTEEEDFHVDQVTTVK
			FNKPFVFLMIEQNTK
			VFLMIEQNTK
P01024	C3	Complement C3	VPVAVQGEDTVQSLTQGDGVAK
			ADIGCTPGSGK
			AELQCPQPAAR

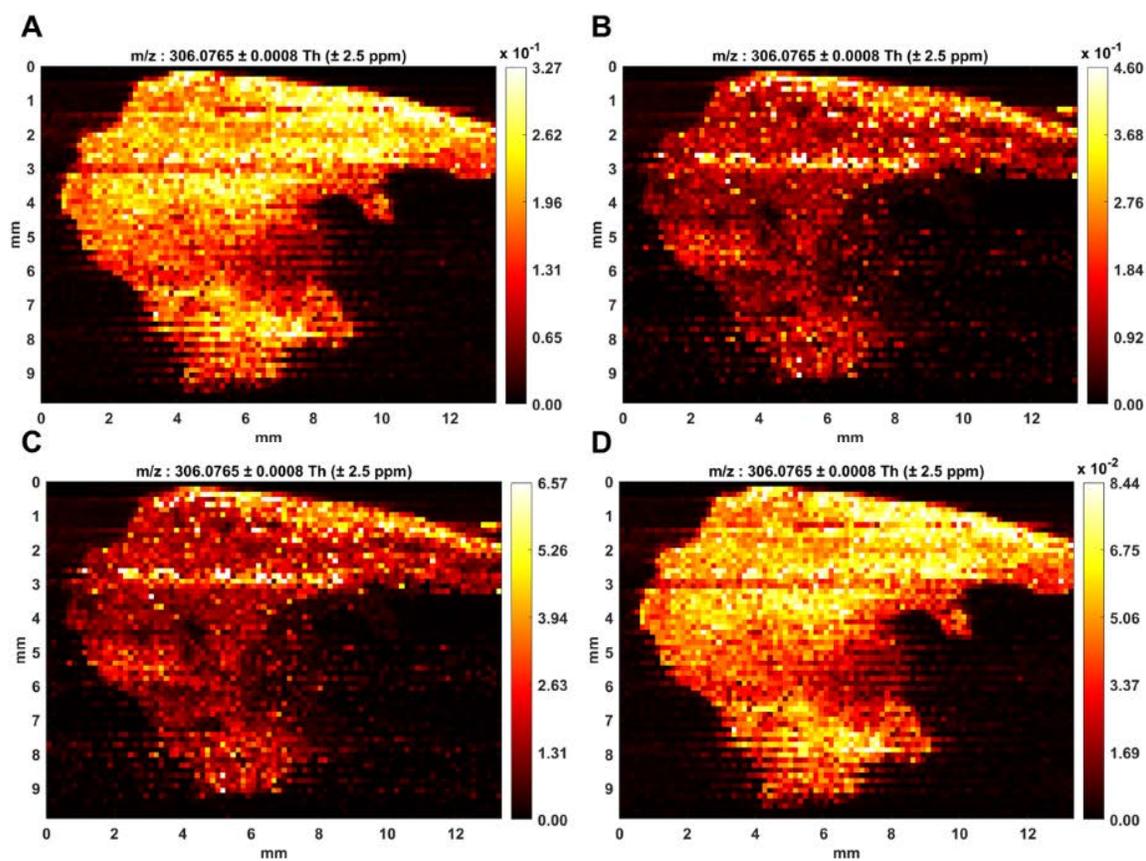
**Table B.2.** Summary of all fixed effects included in linear-mixed effects model to predict ALS-FRS Score

	Peptide	Protein Accession	Gene	Protein Name	Estimate	p-value	Association with Neurological Disease
1	EHIVDLEMLTVSSIGTFR	Q12805	EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	5.49	0.098	
2	AGLAASLAGPH	P08294	SOD3	Extracellular superoxide dismutase	3.87	<b>0.041</b>	
3	ESNPHCGSDGQTYGNK	Q6UWN8	SPINK6	Serine protease inhibitor Kazal-type 6	2.23	0.228	
4	EVDSGNDIYGNIPIK	P16035	TIMP2	Metalloproteinase inhibitor 2	2.01	0.322	
5	ISVTQGASK	Q9Y6R7	FCGBP	FC fragment of IgG binding protein	1.72	<b>0.044</b>	
6	VPTADLEDVPLAEDITNILSK	P02774	GC	Vitamin D-binding protein	1.63	<b>0.006</b>	<b>X</b>
7	CNMGYEYSER	P08603	CFH	Complement factor H	1.62	0.597	
8	LVESGGGLVKPGGSLR	A0A0B4J1V1	IGHV3-21	Immunoglobulin heavy variable 3-21	1.38	0.564	
9	QQQQQSEPGEYEQR	P43121	MCAM	Cell surface glycoprotein MUC18	1.34	0.533	X
10	SVVPTAAAVSLK	P0COL4	C4A	Complement C4-A	0.99	0.631	X
11	AELQEGAR	P02647	APOA1	Apolipoprotein A-I	0.9	0.235	X
12	WVFGGGTK	A0A0A0MT99	IGLJ3	Immunoglobulin lambda joining 3	0.41	0.748	
13	GILCSALQK	P00439	PAH	Phenylalanine-4-hydroxylase	0.21	0.56	
14	LSSLQAGTK	P20062	TCN2	Transcobalamin-2	-0.17	0.86	
15	HYLLR	M9MMK7	DNAH14	Dynein axonemal heavy chain 14	-0.25	0.037	
16	SVEGQNFLSEK	P00451	F8	Coagulation factor VIII	-0.32	0.099	
17	YLYDVLRL	Q9Y6N9	USH1C	Harmonin	-0.34	0.66	
18	LGMFNIQHCK	P01009	SERPINA1	Alpha-1-antitrypsin	-0.94	0.265	X
19	LGADMEDVCGR	P02649	APOE	Apolipoprotein E	-1.05	0.523	X
20	MALDIEIATYR	P08670	VIM	Vimentin	-1.16	0.631	X
21	TTPPMLDSDGSFFLYSK	P01860	IGHG3	Immunoglobulin heavy constant gamma 3	-1.16	0.624	
22	TVLQNETDK	P00450	CP	Ceruloplasmin	-1.18	0.505	X
23	YVMLPVADQDQCIR	P00738	HP	Haptoglobin	-1.65	0.111	X
24	IGEWIGR	Q49A92	C8orf34	Uncharacterized protein C8orf34	-1.8	0.078	
25	EVEELMEDTQHK	Q9UBP4	DKK3	Dickkopf-related protein 3	-2.45	0.412	
26	STTAVVTNPK	P02766	TTR	Transthyretin	-3.97	0.16	X
27	FTFHADICTLSEK	P02768	ALB	Serum albumin	-5.32	<b>0.012</b>	<b>X</b>
28	IISVELPGDAK	P78509	RELN	Reelin	-7.87	<b>0.003</b>	<b>X</b>
			Time from Symptom Onset (Years)		-2.74	<b>0.001</b>	
			Limb Onset		11.23	<b>0.032</b>	

## Appendix C: Supplemental Information for Chapter 6



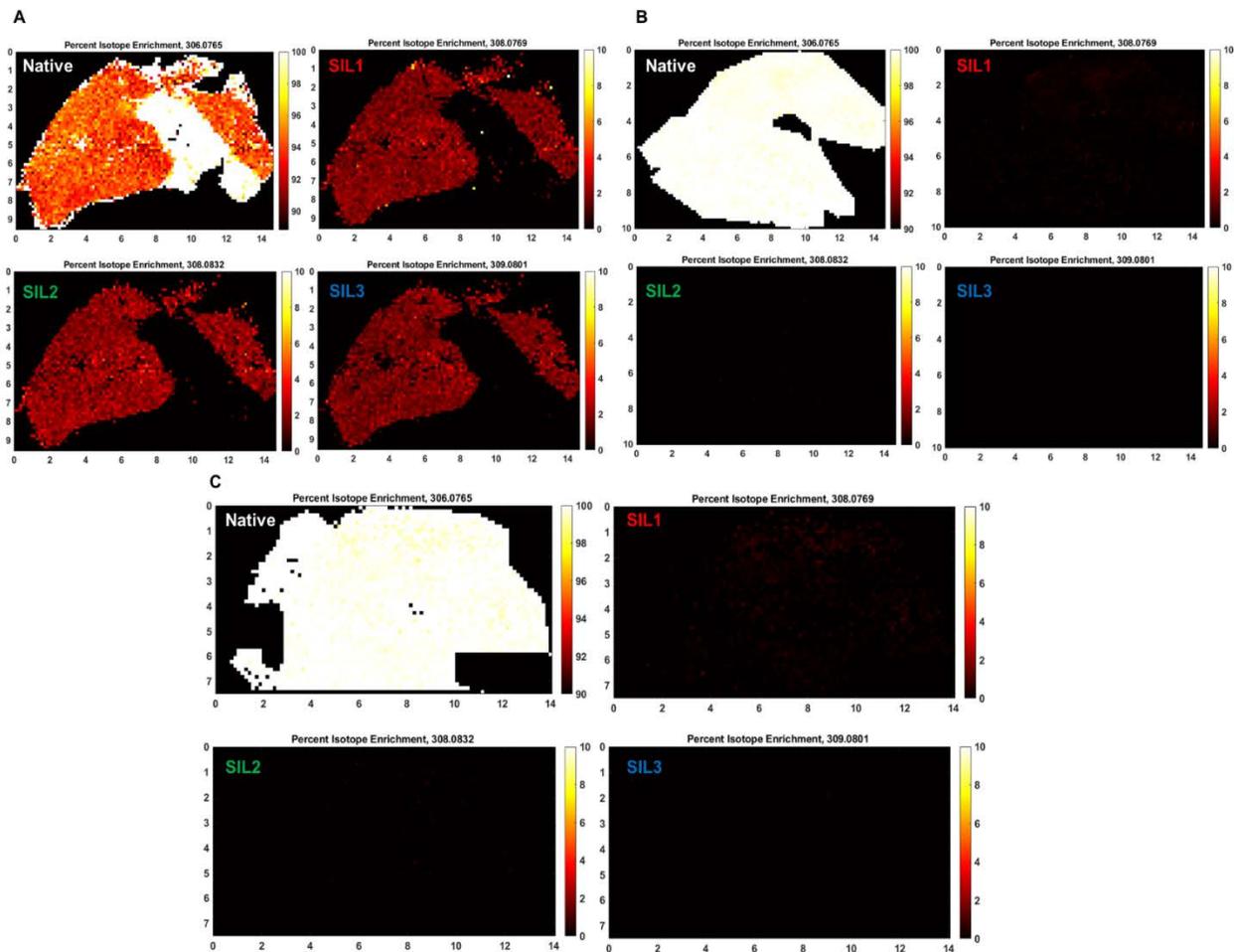
**Figure C.1.** A/B) Carbon and C/D) sulfur counting heatmaps across labeled (A/C) and unlabeled (B/D) liver tissue sections used to confirm identification of native GSH peak via spectral accuracy tool in MSiReader. The images are comparable between labeled/unlabeled, indicating the stable isotope labeling did not disrupt the ability to distinguish these peaks.



**Figure C.2.** A) TIC, B) Maximum, C) Mean, and D) Local TIC normalization of GSH in liver section.

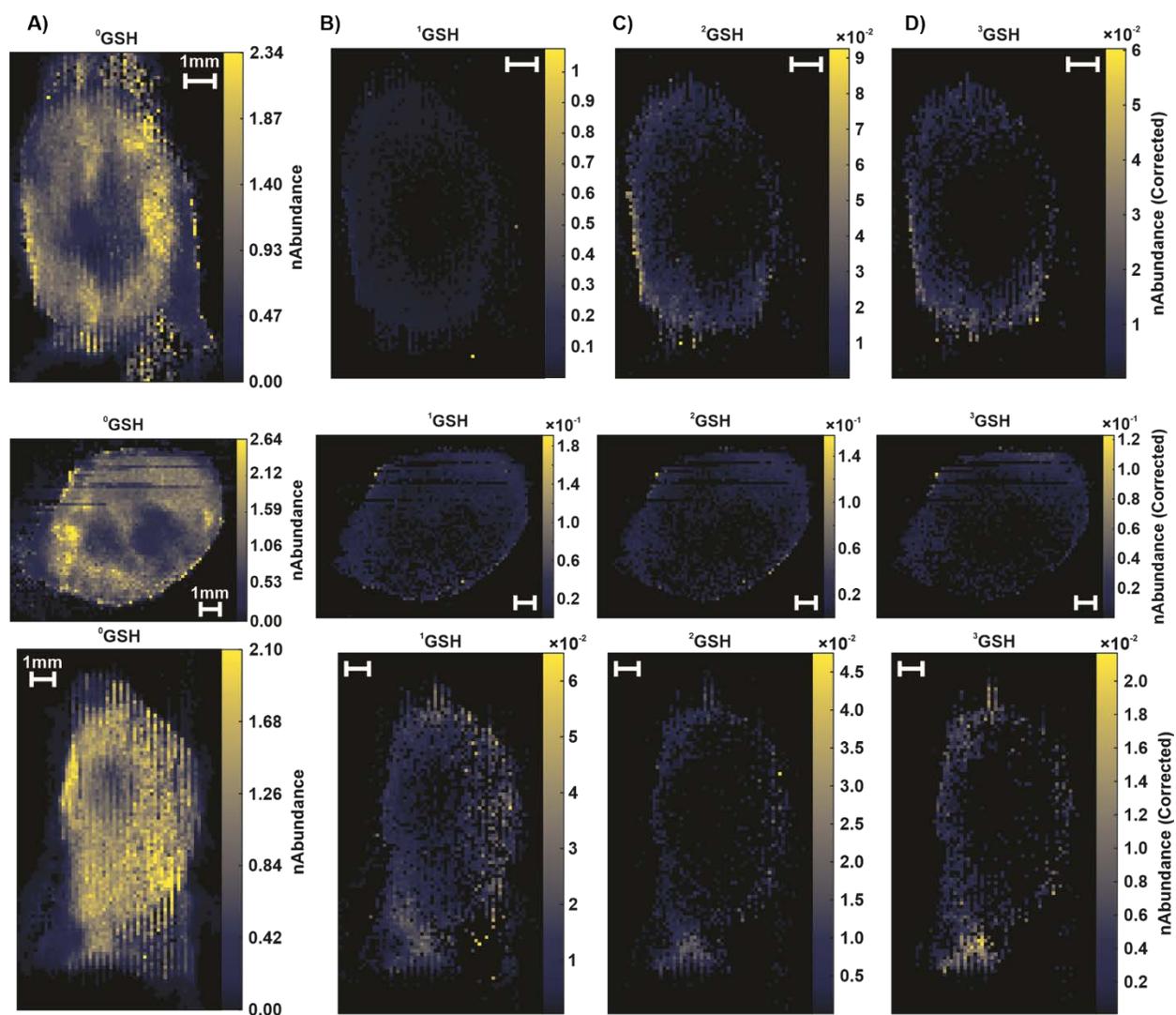
**Table C.1.** GSH Isotopologue Correction Matrix

306.0765	308.0769	308.0832	309.0803
0.82917	0.00099	0.004437773	8.10E-05
0	0.84128	0	0.08257269
0	0	0.8473563	0.009292019
0	0	0	1

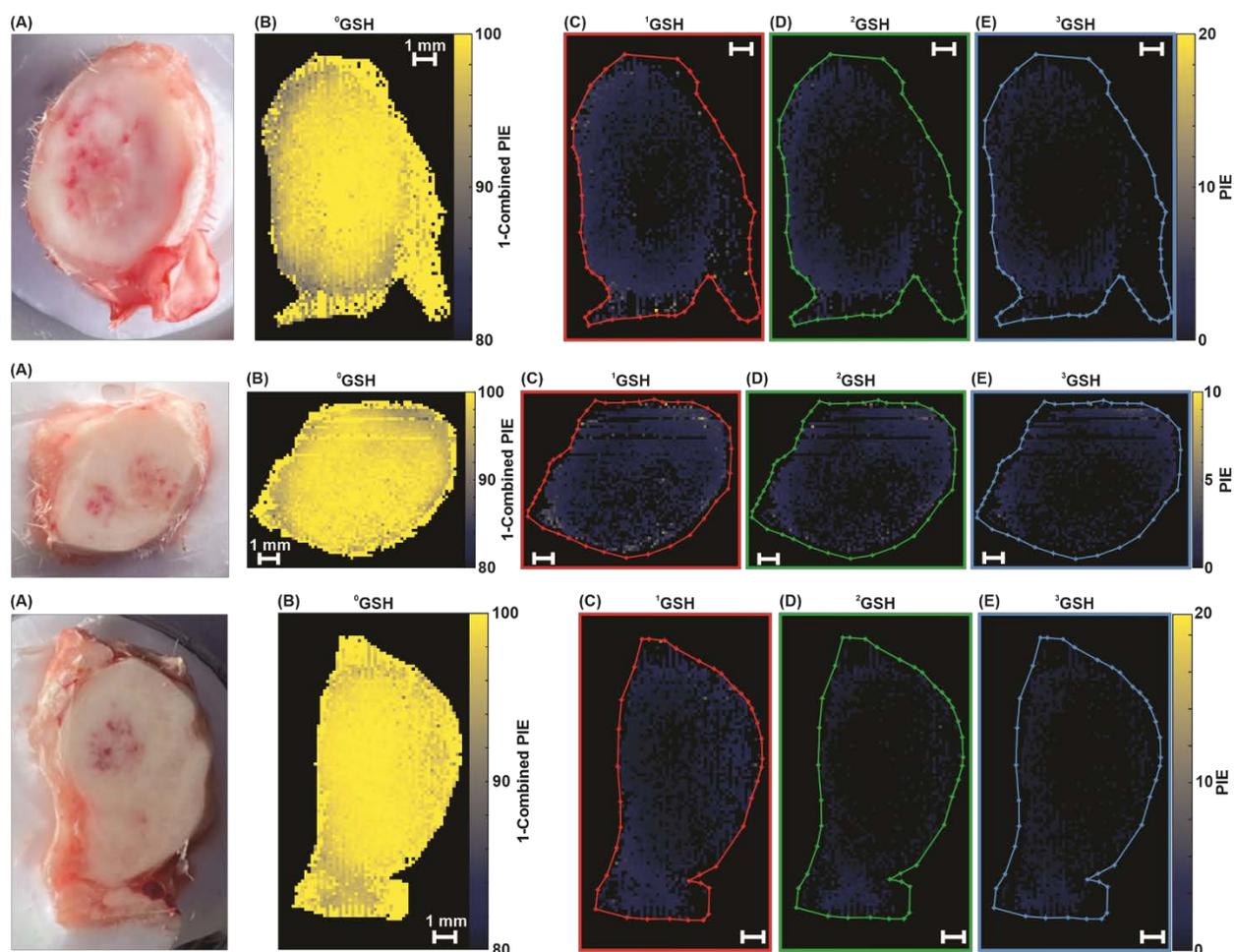


**Figure C.3.** Percent isotope enrichment heatmaps of GSH for **A)** labeled and **B/C)** unlabeled liver sections. No labeling is visible in the unlabeled liver replicate slices, as expected.

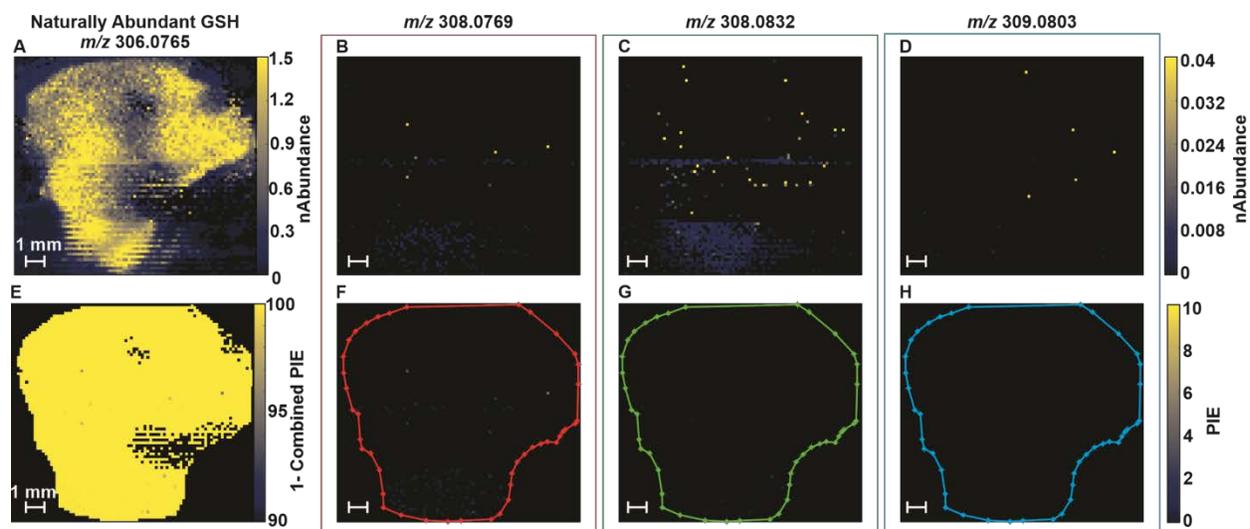
## Appendix D: Supplemental Information for Chapter 7



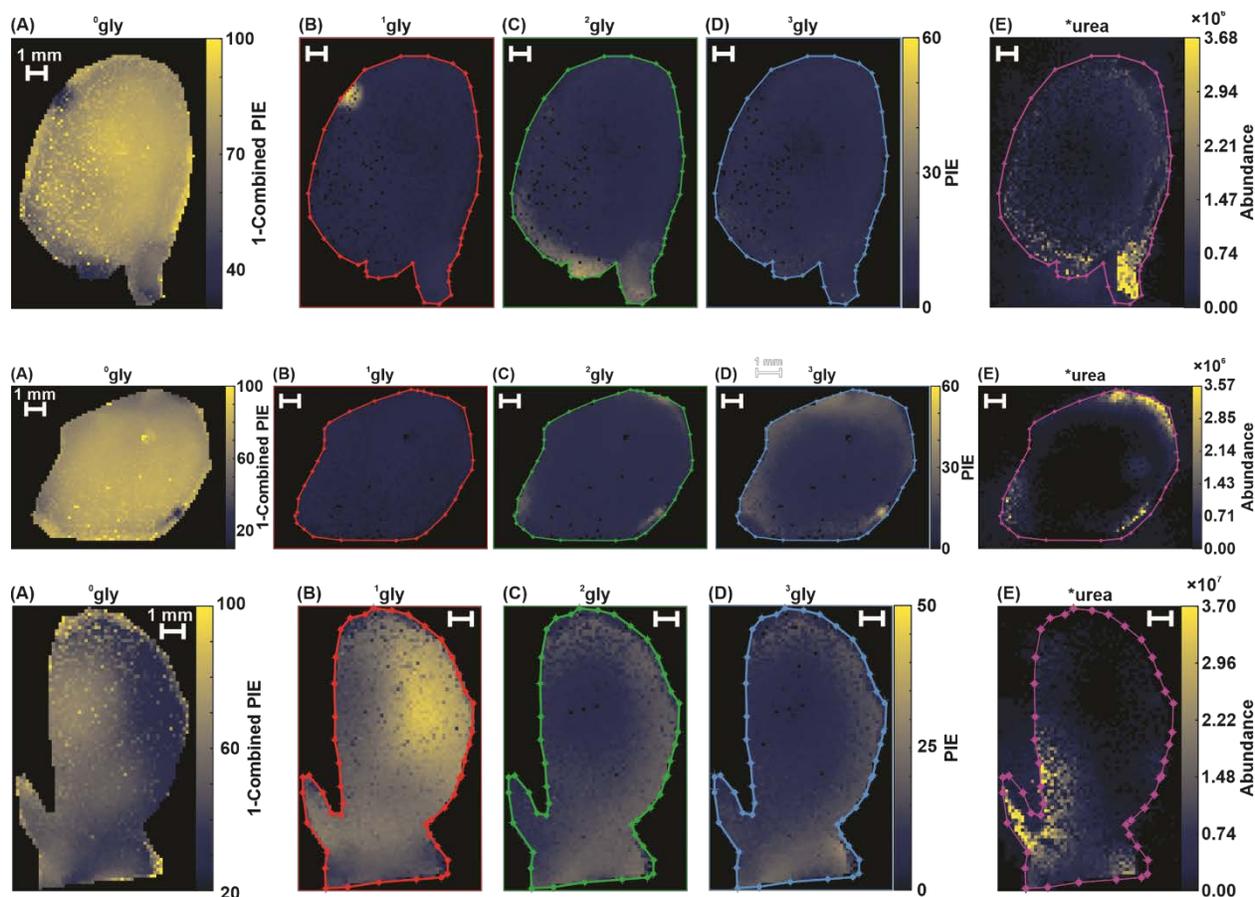
**Figure D.1.** Each row shows MSI data of an additional 4T1 tumor after 120-minute MIST protocol (three replicates total.) Columns represent heatmaps of same  $m/z$  from different tissue sections: Column **A**) MSI heatmaps depicting steady-state abundances of glutathione ( $^0\text{GSH}$ ); **B**)  $^1\text{GSH}$  at  $m/z$  308.0769; **C**)  $^2\text{GSH}$  at  $m/z$  308.0832 and **D**)  $^3\text{GSH}$  at  $m/z$  309.0803. SIL abundances are corrected for isotopologue signal overlap. All heatmaps are normalized to homogluthathione standard. Scale bar = 1mm.



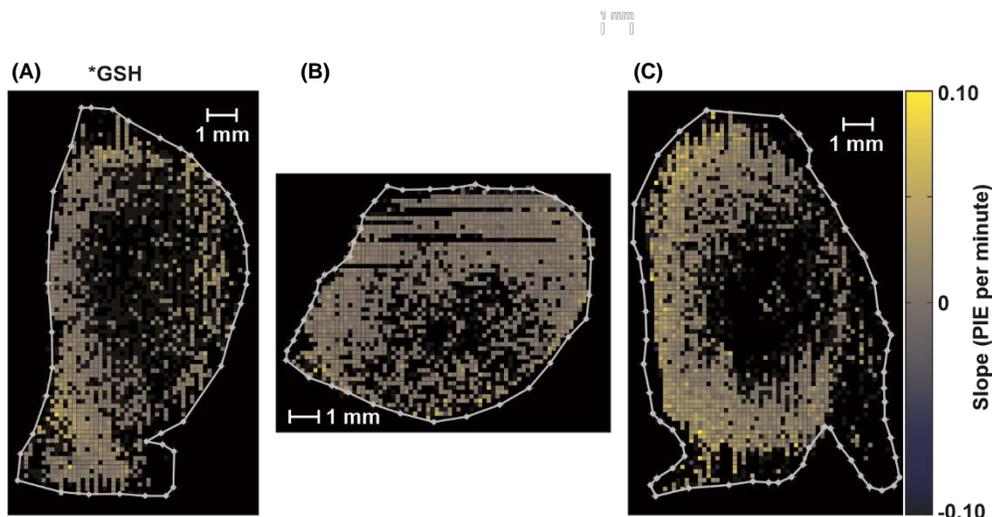
**Figure D.2.** Each row shows data from the additional 4T1 tumor after 120-minute MIST protocol (same sections from **Figure D.1**, in same order) Columns same data from different tissue sections: Column **A**) Optical image of the excised 4T1 tumor. PIE heatmaps of **B**) remaining unlabeled glutathione ( $^0\text{GSH}$ ); **C**)  $^1\text{GSH}$  at  $m/z$  308.0769; **D**)  $^2\text{GSH}$  at  $m/z$  308.0832 and **E**)  $^3\text{GSH}$  at  $m/z$  309.0803. Regions of interest area analyzed with PIE tool are outlined. Scale bar = 1 mm.



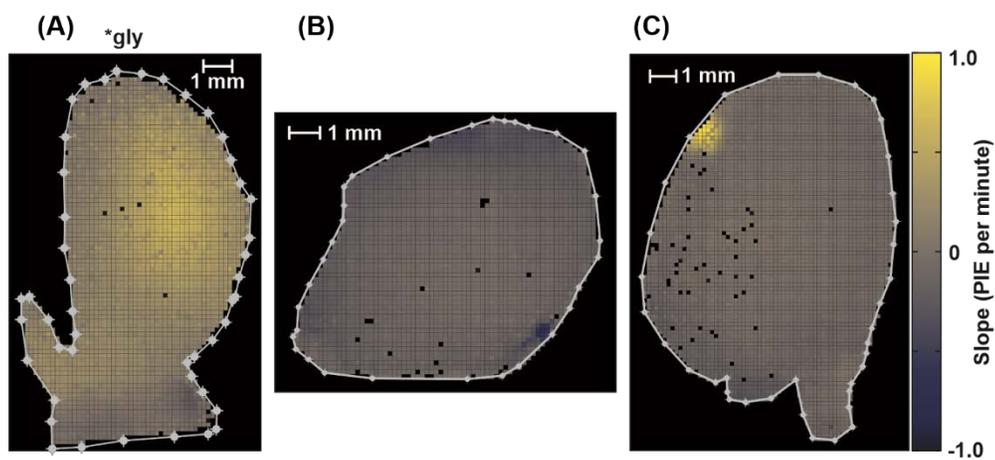
**Figure D.3.** Negative control (sham-labeled) tumor section MSI heatmaps depicting steady state abundances of **A**) glutathione (GSH), **B**)  $^1\text{GSH}$  at  $m/z$  308.0769 **C**)  $^2\text{GSH}$  at  $m/z$  308.0832 and **D**)  $^3\text{GSH}$  at  $m/z$  309.0803. All heatmaps are normalized to homoglutathione standard. PIE heatmaps of **E**) unlabeled glutathione,  $^0\text{GSH}$ , **F**)  $^1\text{GSH}$ , **G**)  $^2\text{GSH}$  and **H**)  $^3\text{GSH}$  detected in a thin section of the 4T1 tumor. Region of interest analyzed with PIE tool is outlined. Scale bar = 1 mm.



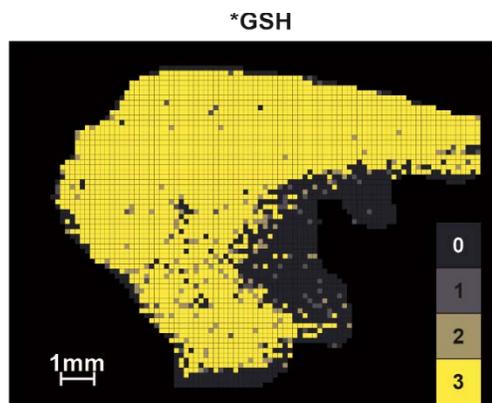
**Figure D.4** Each row shows positive mode data from the additional 4T1 tumor after 120-minute MIST protocol (same tumors but adjacent sections from those shown in **Figure D.1/D.2**, in same order) PIE heatmaps of **A**) remaining unlabeled glycine ( $^0\text{gly}$ ); and the injected glycine isotopologues: **B**)  $^1\text{gly}$ , **C**)  $^2\text{gly}$ , and **D**)  $^3\text{gly}$ . Regions of interest area analyzed with PIE tool are outlined. Heatmaps in Column **E**) show abundance of  $^*\text{urea}$  detected with tissue same regions of interest outlined in pink. Scale bar = 1mm.



**Figure D.5** A-C) Heat maps showing the slopes of the lines fitted to **\*GSH** PIE data from all voxels containing  $\geq 2$  data points for three 4T1 tumors (**Figures D.1/D.2.**) Regions of interest analyzed with PIE tool are outlined.



**Figure D.6.** Heat maps showing the slopes of the lines fitted to **\*gly** PIE data from all voxels containing  $\geq 2$  data points for three 4T1 tumors (**Figure D.4.**) Regions of interest analyzed with PIE tool are outlined.



**Figure D.7** Discrete heatmap showing the number of \*GSH isotopologues detected at each tissue voxel in a thin liver section.