UCLA
UCLA Electronic Theses and Dissertations

Title
Cloud-based Analysis and Integration of Proteomics and Metabolomics Datasets

Permalink
https://escholarship.org/uc/item/38z2r59g

Author
Choi, Howard

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Cloud-based Analysis and Integration of Proteomics and Metabolomics Datasets

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Bioinformatics

by

Jeong Ho (Howard) Choi

2019

© Copyright by

Jeong Ho (Howard) Choi

2019

ABSTRACT OF THE DISSERTATION

Cloud-based Analysis and Integration of Proteomics and Metabolomics Datasets

by

Jeong Ho (Howard) Choi

Doctor of Philosophy in Bioinformatics

University of California, Los Angeles, 2019

Professor Peipei Ping, Chair

Our capabilities to define cardiovascular health and disease using highly multivariate “omics” datasets have substantially increased in recent years. Advances in acquisition technologies as well as bioinformatics methods have paved the way for ultimately resolving every biomolecule comprising various human “omes”. Understanding how different “omes” change and interact with one another temporally will ultimately unveil multi-omic molecular signatures that inform pathologic mechanisms, indicate disease phenotypes, and identify new therapeutic targets. Herein we describe a thesis project that creates novel, contemporary data science methods and workflows to extract temporal molecular signatures of disease from multi-omics analyses, and develops integrated omics knowledgebases for the cardiovascular community at-large.

Chapter 1 provides an overview of cardiac physiology and pathophysiology involved in cardiac remodeling and heart failure (HF). A description of the systematic characterization of cardiac proteomes and metabolomes is included, including methodologies for multi-omics phenotyping. Finally, an overview of bioinformatics methods for driver molecule discovery is provided, discussing strategies for characterizing temporal patterns and conducting functional enrichment.

Chapter 2 describes computational approaches to discern the oxidative posttranslational modification (O-PTM) proteome, an important factor in cardiac remodeling. We developed a novel platform involving a customized, quantitative biotin switch pipeline and advanced analytic workflow to profile O-PTMs in an isoproterenol (ISO)-induced cardiac remodeling mouse model. We identified 1,655 proteins containing 3,324 oxidized sites, and unveiled temporal progression of O-PTM in disease. Chapter 3 discusses computational approaches for identifying temporal metabolomics fingerprints in HF treatment. Pathologic remodeling from a healthy to diseased heart involves a series of alterations over time. Mechanical circulatory support devices (MCSD) are a promising strategy for unloading the heart and reversing this process. We sought to identify molecular drivers of pathologic remodeling and reverse remodeling in the plasma metabolome; however, machine learning (ML)-empowered technological platforms required for these analyses are lacking. Thus, we established a Multiple Reaction Monitoring (MRM)-based MS quantitative platform and ML-based computational workflow to discern metabolomics fingerprints. We quantified 610 plasma metabolites and identified those exhibiting high correlation to cardiac phenotype, demonstrating a novel platform for biomarker discovery. Finally, Chapter 4 integrates all aforementioned innovations into one unified, cloud-based computational knowledgebase, MetProt, equipped to analyze, annotate, and integrate metabolomics and proteomics information. This pipeline fully characterizes the plasma metabolome in HF, unveils the interplay of proteomes and metabolomes, and derives new knowledge in cardiovascular medicine. Innovations include engineered features for addressing large-scale clinical datasets as well as algorithms to connect various types of molecules (e.g., proteins and metabolites). Chapter 4 is subdivided into 3 projects: Project 1 describes a computational pipeline to characterize the plasma metabolome using datasets from the ISO mouse model of HF and human HF; Project 2 develops bioinformatics strategies to integrate proteome and metabolome datasets from six genetically distinct mouse strains; and Project 3 establishes the cloud-based MetProt to disseminate the above computational pipelines for the

cardiovascular community at-large. Taken together, these innovations offer new approaches and workflows for integrated omics investigations that enable novel discovery and ultimately advance precision cardiovascular science and medicine.

The dissertation of Jeong Ho (Howard) Choi is approved

Alex Bui

Mario C. Deng

Wei Wang

Peipei Ping, Committee Chair

University of California, Los Angeles

2019

To my family.

TABLE OF CONTENTS

Abstract of the Dissertation …… ii
Table of Contents …… vii
List of Figures …….. viii
List of Tables …… ix
Recurring Notations and Abbreviations ……... x
Acknowledgements …… xi
Biographical Sketch …… xii

CHAPTER 1: INTRODUCTION AND OVERVIEW.

1.I. The Landscape of the Human Heart and Disease …….. 2
1.II. Systematic Characterization of the Cardiac Proteome and Metabolome …….. 5
1.III. Overview of Bioinformatic Methods for Driver Molecule Identification and Integration .. 6
1.IV. Overview and Rationale …… 12

CHAPTER 2: COMPUTATIONAL APPROACHES TO DISSECT THE CYSTEINE O-PTM PROTEOME DURING CARDIAC HYPERTROPHY.

2.I. Abstract …….. 15
2.II. Introduction …… 15
2.III.Methods and Materials ……... 17
2.IV.Results …… 26
2.V. Discussion …… 56
2.VI.Conclusions ……..... 58

CHAPTER 3: COMPUTATIONAL APPROACHES TO IDENTIFY METABOLOME FINGERPRINTS OF PATHOLOGICAL STAGES FOLLOWING HEART FAILURE TREATMENT.

3.I. Abstract …….. 60
3.II. Introduction …… 61
3.III. Methods and Materials ……... 64
3.IV.Results …… 68
3.V. Discussion …… 83

CHAPTER 4: CLOUD-BASED COMPUTATIONAL KNOWLEDGEBASE TO ANALYZE, ANNOTATE, AND INTEGRATE METABOLOMICS AND PROTEOMICS DATASETS

4.I. Abstract …….. 93
4.II. Introduction …… 94
4.III.Results …… 99
4.IV.Discussion …… 108
4.V. Conclusions ……..... 115

REFERENCES ..……..... 116

LIST OF FIGURES

FIGURE 1.1 Hierachical Structure of Hypothesis Tests …… 8

FIGURE 2.1 Quantitative cysteine O-PTM proteomics workflow ……. 25
FIGURE 2.2 Impact of ISO-induced cardiac hypertrophy on the reversible cysteine O-PTM profile …… 37
FIGURE 2.3 Temporal profiling of reversible cysteine O-PTM proteomes using cubic spline-based clustering. …… 43
FIGURE 2.4 Temporal profiling combining reversible cysteine O-PTM and total cysteine proteomes using cubic spline-based co-clustering …….. 45
FIGURE 2.5 Phenotypic alteration and corresponding cysteine site fingerprints during ISO-induced cardiac hypertrophy …….. 50
FIGURE 2.6 Signature pathways and proteins that contribute to cardiac hypertrophy …….. 52

FIGURE 3.1 Schematic Overview of Experimental Protocol for Global Plasma Metabolomics Profiling in ISO-stimulated Mice, Healthy Humans, and HF Patients. …….. 69
FIGURE 3.2 Metabolomics Dynamics in Mouse Plasma During ISO-induced Cardiac Remodeling.. …….. 71
FIGURE 3.3 Phenotypic Alteration and Corresponding Metabolic Fingerprints During ISO-induced Cardiac Remodeling in Mice …… 74
FIGURE 3.4 Healthy Human Subjects Plasma Metabolomics Stability and Their Distinction from HF Patients ……... 77
FIGURE 3.5 Metabolomics Dynamics in Plasma from HF Patients Before and After MCSD Implantation…… 78
FIGURE 3.6 Temporal Changes of Clinical Manifestation and Corresponding Metabolomic Dynamics in HF Patients Following MCSD Implantation…… 81

FIGURE 4.1 MetProt Integrated workflws …… 100
FIGURE 4.2 Intersections where metabolome meets proteome …….. 103
FIGURE 4.3 MetProt Distributed System Archiecture …… 106
FIGURE 4.4 MetProt User Interface ……107

LIST OF TABLES

TABLE 2.1 Summary information of raw file contents and statistics for all experimental groups …….. 23&24
TABLE 2.2 List of cysteine sites with alterations in modification abundance during cardiac hypertrophy ……. 28-32
TABLE 2.3 Enriched biological pathways of proteins bearing cysteine sites with significantly increased/decreased modification abundance …… 33-35
TABLE 2.4 List of cysteine sites significantly correlated with the hypertrophy phenotype ……..... 46-48
TABLE 2.5 Enriched biological pathways of proteins associated with cysteine sites significantly correlated with the hypertrophy phenotype …… 51

TABLE 3.1 Biological Calsses of 610 Targeted Metabolites in Mouse and Human Plasma ……..... .63
TABLE 3.2 Demographics of the Human Cohorts …… 65-66

RECURRING NOTATIONS AND ABBREVIATIONS

ATP	Adenosine triphosphate
HF	Heart Failure
ISO	Isoproterenol
k	Protein turnover rate constant
kp	Precursor turnover rate constant
LC	Liquid chromatography
LMM	Linear Mixed Model
MCSD	Mechanical Circulatory Support Device
ML	Machine Learning
MRM	Multiple-Reaction-Monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
O-PTM	Oxidative stress-sensitive Post-translational modifications
PTM	Post-translational modifications
ROS	Reactive oxygen species

ACKNOWLEDGEMENTS

I am particularly indebted to my mentor, Dr. Peipei Ping, for not only giving me the opportunity to conduct my research in her laboratory, but also providing me with the necessary resources. From her guidance, support, and patience, I have grown tremendously both intellectually and as an individual. I would also like to give my sincere appreciations to the other members of my committee, Drs. Wei Wang, Alex Bui, and Mario C. Deng for their enduring guidances and supports.

I would also like to thank my dear friends and colleagues, Dr. Jie Wang, Dr. David Liem, Dr. Ding Wang, Dr. Dominic Ng, Dr. Bilal Mirza, Dr. Harry Caufield, Mr. Anders Garlid (soon to be a doctor!), Dr. Dibakar Sigdel, Mr. Vincent Kyi, Dr. X'avia Chan, Dr. Sarah Scruggs, Dr. Dibakar Sigdel, Dr. Travis Cao, Dr. Nobel Zong, Dr. Haomin Li and the rest of the Ping lab members and our collaborators globally. To Henning Hermjakob, Antonio Febregat Mundo, and Juan Antonio Vizcaino at EMBL-EBI, I could not accomplish this without you.

Finally, I am most grateful to my family; to my parents whom taught me to say my first sentence and write my first equation, whom hold my hands through my educations in kindergarten, elementary schools, until now the completion of my PhD thesis. I am also tremendously grateful to my parents in law for their devoted support and love. To the joy of my life, my son and my wife for your enormous love and strong support, this thesis is for you.

BIOGRAPHICAL SKETCH

Education and employment:

M.S.	Bioinformatics and Medical Informatics, San Diego State University, CA, 2012
B.S.	Biological Resources and Technology, Yonsei University, South Korea, 2007
2010 - 2012	Research Programmer, The Scripps Research Institute, La Jolla, CA,
2010	Summer Internship, The Scripps Research Institute, La Jolla, CA

Peer reviewed publications:

1. Chung, N. C.; Mirza, B.; Choi, H.; Wang, J.; Wang, D.; Ping, P.; Wang, W. (2019). Unsupervised Classification of Multi-Omics Data during Cardiac Remodeling using Deep Learning. Methods., pii: S1046-2023(18)30359-1. doi: 10.1016/j.ymeth.2019.03.004. PMID: 30853547.

2. Mirza, B.; Wang, W.; Wang, J.; Choi, H.; Chung, N. C.; Ping, P. (2019). Machine Learning and Integrative Analysis of Biomedical Big Data. Genes (Basel)., 10(2), 87. doi: 10.3390/genes10020087. PMID: 30696086.

3. Choi, H.; Wang, J.; Chung, N. C.; Cao, Q.; Ng, D. C. M.; Mirza, B.; Scruggs, S. B.; Wang, D.; Garlid, A. O.; Ping, P. (2018). Integrated Dissection of the Cysteine Oxidative Post-translational Modification Proteome During Cardiac Hypertrophy. J Proteome Res., 17(12), 4243-4257. doi: 10.1021/acs.jproteome.8b00372. PMID: 30141336.

4. Liem, D. A.; Murali, S.; Sigdel, D.; Shi, Y.; Wang, X.; Shen, J.; Choi, H.; Caufield, J. H.; Wang, W.; Ping, P.; Han, J. (2018). Phrase Mining of Textual Data to Analyze Extracellular Matrix Protein Patterns Across Cardiovascular Disease. Am J Physiol Heart Circ Physiol., 315(4), H910-H924. doi: 10.1152/ajpheart.00175.2018. PMID: 29775406.

5. Baek, K. I.; Li, R.; Jen, N.; Choi, H.; Kaboodrangi, A.; Ping, P.; Liem, D.; Beebe, T.; Hsiai, T. K. (2018). Flow-Responsive Vascular Endothelial Growth Factor Receptor-Protein Kinase C Isoform Epsilon Signaling Mediates Glycolytic Metabolites for Vascular Repair. Antioxid Redox Signal., 28(1), 31-43. doi: 10.1089/ars.2017.7044. PMID: 28762754.

6. Zhang, Y.; Xu, T.; Shan, B.; Hart, J.; Aslanian, A.; Han, X.; Zong, N.; Li, H.; Choi, H.; Wang, D.; Acharya, L.; Du, L.; Vogt, P. K.; Ping, P.; Yates, J. R., 3rd. (2015). ProteinInferencer: Confident protein identification and multiple experiment comparison for large scale proteomics projects. J Proteomics., 129, 25-32. doi: 10.1016/j.jprot.2015.07.006. PMID: 26196237.

7. Villaveces, J. M.; Jimenez, R. C.; Porras, P.; Del-Toro, N.; Duesbury, M.; Dumousseau, M.; Orchard, S.; Choi, H.; Ping, P.; Zong, N. C.; Askenazi, M.; Habermann, B. H.; Hermjakob, H. (2015). Merging and scoring molecular interactions utilising existing community standards: tools, use-cases and a case study. Database (Oxdord)., 2015, bau131. doi: 10.1093/database/bau131. PMID: 25652942.

8. Lindskog, C.; Linne, J.; Fagerberg, L.; Hallstrom, B. M.; Sundberg, C. J.; Lindholm, M.; Huss, M.; Kampf, C.; Choi, H.; Liem, D. A.; Ping, P.; Varemo, L.; Mardinoglu, A.; Nielsen, J.; Larsson, E.; Ponten, F.; Uhlen, M. (2015). The human cardiac and skeletal muscle

proteomes defined by transcriptomics and antibody-based profiling. BMC Genomics., 16(1), 475. doi: 10.1186/s12864-015-1686-y. PMID: 26109061.

9. Zong, N.; Ping, P.; Lau, E.; Choi, H. J.; Ng, D. C.; Meyer, D.; Fang, C.; Li, H.; Wang, D.; Zelaya, I. M.; Yates, J. R., 3rd; Lam, M. P. (2014). Lysine ubiquitination and acetylation of human cardiac 20S proteasomes. Proteomics Clin Appl., 8(7-8), 590-594. doi: 10.1002/prca.201400029. PMID: 24957502.

10. Lotz, C.; Lin, A. J.; Black, C. M.; Zhang, J.; Lau, E.; Deng, N.; Wang, Y.; Zong, N. C.; Choi, J. H.; Xu, T.; Liem, D. A.; Korge, P.; Weiss, J. N.; Hermjakob, H.; Yates, J. R., 3rd; Apweiler, R.; Ping, P. (2014). Characterization, design, and function of the mitochondrial proteome: from organs to organisms. J Proteome Res., 13(2), 433-446. doi: 10.1021/pr400539j. PMID: 24070373.

11. Zong, N. C.; Li, H.; Li, H.; Lam, M. P.; Jimenez, R. C.; Kim, C. S.; Deng, N.; Kim, A. K.; Choi, J. H.; Zelaya, I.; Liem, D.; Meyer, D.; Odeberg, J.; Fang, C.; Lu, H. J.; Xu, T.; Weiss, J.; Duan, H.; Uhlen, M.; Yates, J. R., 3rd; Apweiler, R.; Ge, J.; Hermjakob, H.; Ping, P. (2013). Integration of cardiac proteome biology and medicine by a specialized knowledgebase. Circ Res., 113(9), 1043-1053. doi: 10.1161/CIRCRESAHA.113.301151. PMID: 23965338.

12. Li, H.; Zong, N. C.; Liang, X.; Kim, A. K.; Choi, J. H.; Deng, N.; Zelaya, I.; Lam, M.; Duan, H.; Ping, P. (2013). A novel spectral library workflow to enhance protein identifications. J Proteomics., 81, 173-184. doi: 10.1016/j.jprot.2013.01.026. PMID: 23391412.

13. Fonslow, B. R.; Stein, B. D.; Webb, K. J.; Xu, T.; Choi, J.; Park, S. K.; Yates, J. R., 3rd. (2013). Digestion and depletion of abundant proteins improves proteomic coverage. Nat Methods., 10(1), 54-56. doi: 10.1038/nmeth.2250. PMID: 23160281.

14. Chan, C. X.; Khan, A. A.; Choi, J. H.; Ng, C. D.; Cadeiras, M.; Deng, M.; Ping, P. (2013). Technology platform development for targeted plasma metabolites in human heart failure. Clin Proteomics., 10(1), 7. doi: 10.1186/1559-0275-10-7. PMID: 23826926.

15. Fonslow, B. R.; Niessen, S. M.; Singh, M.; Wong, C. C.; Xu, T.; Carvalho, P. C.; Choi, J.; Park, S. K.; Yates, J. R., 3rd. (2012). Single-step inline hydroxyapatite enrichment facilitates identification and quantitation of phosphopeptides from mass-limited proteomes with MudPIT. J Proteome Res., 11 (5), 2697-2709. doi: 10.1021/pr300200x. PMID: 22509746.

16. Culver, B. P.; Savas, J. N.; Park, S. K.; Choi, J. H.; Zheng, S.; Zeitlin, S. O.; Yates, J. R., 3rd. (2012). Tanese, N., Proteomic analysis of wild-type and mutant huntingtin-associated proteins in mouse brains identifies unique interactions and involvement in protein synthesis. J Biol Chem., 287(26), 21599-21614. doi: 10.1074/jbc.M112.359307. PMID: 22556411.

CHAPTER 1: INTRODUCTION AND OVERVIEW.

1.I. The Landscape of the Human Heart and Disease.

1.I.A. Cardiac Structure and Physiology.

The heart is a muscular organ that pumps blood to the body via the vessels of the circulatory system (Venes 2009). The circulatory system delivers the blood to all tissues in order to provide the body with oxygen and nutrients, and carries away metabolic wastes such as nitrogen compounds or sulphates for excretion (Hall and Guyton 2011). In humans and other mammals (e.g., mouse), the heart consists of four chambers including left and right atria and left and right ventricles, as well as four valves including two atrioventricular valves between atria and ventricles and two semilunar valves at the exit of each ventricle that prevent backflow (Starr, Evers, and Starr 2009; Kaplan 2008). The right atrium and right ventricle, commonly referred to as the right heart, collects deoxygenated blood from superior and inferior venae cavae and pumps into the pulmonary circulation in order to oxygenate the blood via the lungs. The left atrium and the left ventricle, referred to as the left heart, receives oxygenated blood via pulmonary veins and pumps into the systemic circulation. The heart is primarily made up with two types of cells: cardiomyocytes (~99%) which are contractile cells that enable the heart to pump, and cardiac pacemaker cells (~1%) that spontaneously depolarize and rapidly spread the impulse from cell to cell to trigger the autorhythmic contraction. Cardiomyocytes have a high mitochondrial density, which occupy at least 30% of cell volume (Piquereau et al. 2013), enabling them to produce sufficient amounts of adenosine triphosphate (ATP) to fuel contraction. This high density of the power plant of the cell combined with its rhythmic electrical impulse enables the heart to beat approximately to 72 beats per minute (bpm) at rest in humans, and more than 31 million times per year or 2 billion beats over an average lifetime.

1.I.B. Cardiac Remodeling and Heart Failure.

Cardiac remodeling, also referred to as ventricular remodeling, is defined as an alteration in the structure (dimensions, mass, shape) of the heart in response to hemodynamic load and/or cardiac injury in parallel with neurohormonal activation (Cohn, Ferrari, and Sharpe 2000). Remodeling can be either adaptive or maladaptive, with maladaptive cardiac remodeling leading to heart failure (HF) (Dorn, Robbins, and Sugden 2003). HF is a condition in which the heart is unable to pump a sufficient amount of blood to meet the body's demand. HF is a convergent end result of many etiologies of heart disease, including cardiomyopathies. HF is a common, costly, and potentially fatal disease responsible for >300,000 deaths a year in the US, with 5.8 million Americans currently afflicted (Mancini and Colombo 2015; Roger 2013; Braunwald 2015; Goldberg et al. 2007). A major reason why this common disease is so intractable is its complex, multifactorial nature where disease outcome is often the combinatorial effect of multiple genetic and environmental causes.

1.I.C. Biomarkers of Heart Failure Progression.

B-type natriuretic peptide (BNP) and its biologically inert, amino-terminal pro-peptide counterpart (NT-proBNP) are the gold standard biomarkers in determining the diagnosis and prognosis of HF (Hanna K. Gaggin and Januzzi 2013). An array of additional biomarkers has emerged and can be arranged into the following categories: 1) myocardial stress/injury (e.g., myocyte stretch markers: MR-proANP; myocardial injury markers: troponin T, troponin I, myosin light-chain I, heart-type fatty-acid protein, CKMB; oxidative stress markers: myeloperoxidase, uric acid, oxidized low-density lipoproteins, urinary biopyrrins, urinary and plasma isoprostanes, plasma malodialdehyde), 2) neurohormonal activation (e.g., renin angiotensin system markers: renin, angiotensin II, aldosterone; sympathetic nervous system: norepinephrine, chromogranin A, MR-proADM; arginine vasopressin system: arginine vasopressin, copeptin; endothelins: ET-1, big proET-1), 3) remodeling (e.g., inflammation: C-reactive protein, TNF-α, soluble TNF

receptors, Fas, interleukins [I, 6 and 18], osteoprotegerin, adiponectin; hypertrophy/fibrosis: matrix metalloproteinases, collagen propeptides, galectin-3, soluble ST2; apoptosis: GDF-15; misc: MicroRNA, quiescin Q6, VEGFR-1) and 4) comorbidities (e.g., renal biomarkers: creatinine, BUN, eGFR, cystatin C, β-trace protein; renal injury markers: NGAL, KIM-1, NAG, liver-type fatty acid binding protein, IL-18; hematologic biomarkers: hemoglobin, RDW, iron deficiency [ferritin, transferrin sat], albumin) (Hanna Kim Gaggin et al. 2017).

The majority of key biomarkers mentioned above are comprised of proteins, post-translationally modified proteins, and metabolites. Notably, oxidative stress-related molecules comprise a significant population of myocardial stress/injury markers, suggesting a prominent role for oxidative stress in cardiac remodeling and heart failure. Although biomarkers such as BNP and troponin T are helpful in the diagnosis and management of irreversible myocardial injury, we currently have no satisfactory diagnostic markers of many cardiovascular complications that possibly leads to heart failure such as reversible myocardial ischemia, i.e. either stable or unstable angina (Gerszten, Asnani, and Carr 2011; Morrow et al. 2003). Despite increasing demand for cardiovascular biomarkers, few new markers have been FDA approved for the diagnosis of cardiovascular diseases (Morrow et al. 2003). Mass spectrometry (MS)-based proteomics and metabolomic technologies are capable of identifying hundreds to thousands of proteins, metabolites and oxidative stress-related molecules in cells, tissues and biofluids. Thus, a systematic approach integrating multi-omics analysis (i.e., proteomics, O-PTM proteomics, and metabolomics) would provide deep phenotyping to reveal new biomarkers, prioritize biomarker validation, and monitor biomarker panels for heart failure progression and drug response.

1.II. Systematic Characterization of the Cardiac Proteome and Metabolome.

1.II.A. Systematic Profiling of The Cardiac Proteome

As molecules that perform most mechanical and biochemical functions in the heart, proteins provide critical information to systems biology studies that aim to uncover the mechanisms of disease susceptibility and cardiac plasticity. Technologies now exist that can easily identify and quantify large numbers of proteins in a single experiment (Lotz et al. 2014; Mann 2006). These studies have advanced our understanding of the components constructing protein interaction networks (Orchard et al. 2014), the role of post-translational modifications in modulating cellular signals (Streng et al. 2013), the dynamics of individual protein half-life in response to perturbation (Lam et al. 2014), and the totality of proteins contributing to a given cellular process or organelle (Kim et al. 2014). One of the principal lessons from the proteomics revolution is that gene expression is no surrogate for protein function, e.g., following ISO stimulation, protein abundance changes independently of mRNA (Lam et al. 2014).

1.II.B. Profiling of the Cardiac Metabolomic Landscape.

The persistent contraction of cardiomyocytes demands tremendous amounts of energy; thus, maintaining metabolic homeostasis is essential for heart function. Accumulating evidence suggests that perturbation of cardiac metabolism plays an important role in the pathological progression of HF (Neubauer 2007). Comprehensive quantification of metabolite abundance in blood plasma, also known as plasma metabolomics profiling, provides mechanistic insights into the molecular alterations underlying HF (Cheng et al. 2015). The large-scale quantification of circulating metabolites also identifies metabolic changes for clinical classification, which may assess the outcomes of HF patients before and after optimal medical therapy and surgical interventions (Lloyd-Jones 2010). These metabolomic approaches can be established as

important avenues for biomarker and drug target development, as well as the validation of the efficacy of patient treatments.

1.II.C. Multi-omics Phenotyping of HF.

Along with advances in high-throughput technologies, recent studies have integrated gene expression profiling with proteomics, and have facilitated our understanding of the pathophysiology and the molecular mechanisms of HF (Hou et al. 2015; Dos Remedios et al. 2003). However, there are no systematic approaches for the integration of large proteomic datasets with metabolome datasets due to a glaring lack of bioinformatics strategies that can connect proteins to metabolites and elucidate their relationships.

1.III. Overview of Bioinformatic Methods for Driver Molecule Identification and Integration.

The availability of proteomics and metabolomics data analysis tools is currently limited compared to genomics analysis tools. For example, the Aztec resource discovery index (http://aztec.bio/) lists 778 proteomics tools, and 147 metabolomics tools while it lists thousands of genomics tools. To our knowledge, there is no application or computational pipeline that comprehensively analyzes, annotates, and integrates large-scale proteomic and metabolomic datasets. This limitation significantly hinders our capability to identify patterns of biomolecules relevant to HF.

1.III.A. Computational Approaches to Identify Driver Molecules.

Differential Expression Analysis: Differential Expression Analysis (DEA) is a univariate method widely used in omics studies for targeting molecules that are statistically significantly altered (i.e., up-regulated or down-regulated) after a medical treatment or physical exercise (Xia et al. 2015, 2012, 2009; Xia and Wishart 2011b). This method mainly employs two criteria: the fold-

change of the molecule abundance in two sample groups and the adjusted p-value from statistical hypothesis testing followed by multiple testing correction. Recent studies have shown that if the sample size is sufficient (i.e., n > 12), the univariate test such as t test followed by multiple testing correction can outperform the multivariate methods (e.g., linear support vector machine-recursive features elimination or principal component discriminant analysis) in terms of not missing the true positives (Christin et al. 2013). Alternatively to the t test, a non-parametric test such as Wilcoxon signed-rank test can be applied if it is difficult to test the normality of the dataset (Figure 1.1.). Following the statistical hypothesis test, multiple testing correction to the p-values can be applied using the Benjamini-Hochberg method to reduce the false positives rate. Subsequently, the identified molecules can be validated by classification/clustering methods, such as Principal Component Analysis (PCA) or Unsupervised Hierarchical Clustering. PCA is a multivariate method for dimensional reduction of large datasets with many variables or observations. This facilitates visualization, clustering, pattern recognition, and identification of the key variables that vary most significantly across a population. PCA can aid investigators in understanding complex, higher-dimensional datasets by projecting them into a 2- or 3-dimensional space, where they can be more easily visualized without sacrificing fidelity (Xia et al. 2015, 2012, 2009; Xia and Wishart 2011b, [a] 2011). Moreover, Unsupervised Hierarchical Clustering is a widely used data analysis method to build a binary tree from data by merging similar groups of points based on the calculated distances across samples or molecules. Usually, the distances are displayed as dendrograms (Xia et al. 2015, 2012, 2009).

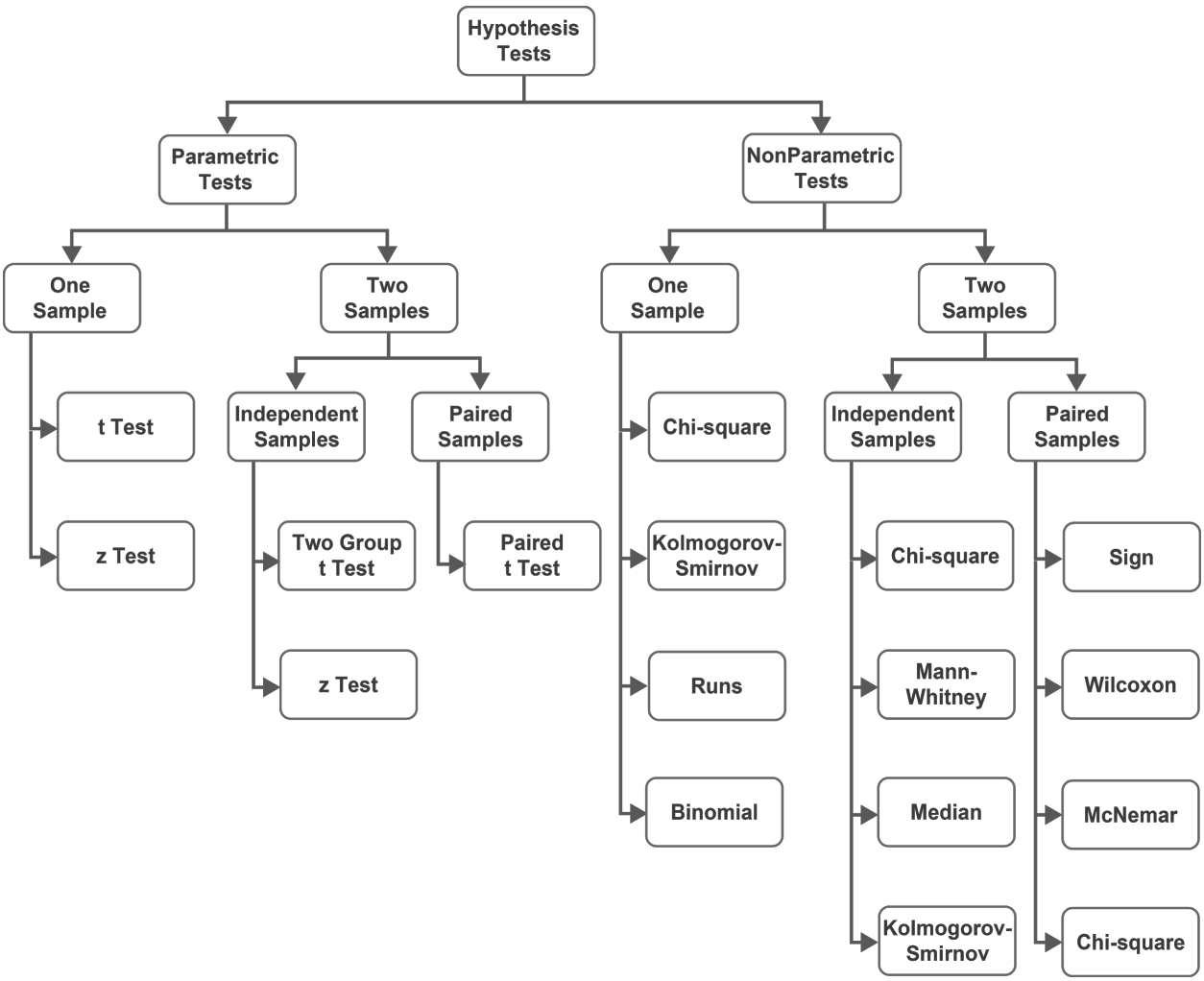


Figure 1.1. Hierachical Structure of Hypothesis Tests. Hypothetical testing related to statistical differences are classified into two categories: parametric and nonparametric tests, depending on prior knowledge about the population parameter. For each category, statistical approaches suitable for one sample or two samples tests are as illustrated in the hierarchical graph. For two sample test, statistical methods are further divided into independent and paired samples depend on the structure of the data. Among 11 statistical methods, t test and z test are applied in both one sample and two-sample tests of the parametric test, whereas Chi-square and Kolmogorov-Smirnov are applied in both subcategories of the non-parametric test. In addition, Runs and Binomial are specific for one sample non-parametric test; Mann-Whitney and Median are specific for independent two-sample non-parametric test; Sign, Wilcoxon, and McNemar are specific for paired two-sample non-parametric test. This hypothesis test hierarchy serves as the guiding principle for the implementation of statistical approaches throughout this thesis.

Machine Learning-based Feature Extraction: Machine Learning (ML)-based feature selection algorithm aims to identify key molecules discriminating one condition from the other. The filtered and scaled data from a mouse model is divided into non-overlapping training and test datasets. A two-stage feature selection and validation workflow are created to extract signature

metabolites. On the training set, randomized least absolute shrinkage and selection operator (LASSO)-based stability selection is applied in the first stage to screen for metabolites of interest and support vector machine (SVM)-based recursive feature elimination (RFE) is applied in the second stage to identify a subset of accurate metabolites. Subsequently, the identified metabolites are validated on the test sets using predictive models and then local interpretable model-agnostic explanations (LIME) (Ribeiro, Singh, and Guestrin 2016) is applied to understand the predictions made by ML models. Finally, a Bayesian networks-based graphical model is created to discover causal relationships between identified metabolites and phenotypic response.

Mixed Model-based Association Study: Linear Mixed Model (LMM) with Sample Relatedness Correction analyzes time-series omics data generated from heterogeneous samples to identify driver molecules that are highly associated with phenotypic characteristics. LMM correlates two variables and incorporates the samples' relatedness, represented by omics expression or demographic factors, to improve statistical accuracy. We have applied LMM with Sample Relatedness Correction in our preliminary metabolomics study to find fingerprints of pathological stages following induction of HF using chronic infusion of the beta-adrenergic agonist, isoproterenol (ISO) HF treatment (H. M. Kang et al. 2008; Kirby et al. 2010).

1.III.B. Bioinformatic Stretage to Characterize Dynamic Patterns of Molecules

Cubic Spline-based Temporal Clustering: Cubic spline-based temporal clustering is a technique where first computes a low-degree polynomial pieces that fit smoothly on data, and then clusters the similar spline curves using clustering methods such as K-mean clustering. We used this technique in our studies to analyze metabolite concentrations for six genetically heterogeneous strains of mich that exhibit varying phenotypes including different susceptibilities to cardiac dysfunction. We subjected these to chronic ISO treatment over 14 days. Both CTRL

and ISO groups were selected for cubic spline-based temporal clustering. Missing concentration values were replaced with the average occupancy of the available timepoints for every strain. The log2-transformed ratio of concentration values of ISO to CTRL was calculated per site per strain. After scaling and centering these ratios, cubic splines were fitted to them across the timepoints in R. The degree of smoothing was automatically determined by leave-one-out cross-validation applied to each site per strain. The predicted spline values were used for K-mean clustering (kmean in R), following identification of the cluster numbers (mclust package in R).

1.III.C. Functional Enrichment Analysis and Annotation

Pathway Enrichment Analysis: Significantly altered metabolites in our studies may be searched against several metabolomics databases, including Chemical Entities of Biological Interest (ChEBI, UK) (Hastings et al. 2013), Human Metabolome Database (HMDB, Canada) (Wishart et al. 2013, 2009, 2007), and LIPID Metabolites And Pathways Strategy (LIPID MAPS) (Fahy et al. 2009, 2007) to cross-reference identifiers that are commonly used by pathway enrichment search engines and databases. The pathways enriched with identified metabolites may be further annotated with resources including MetaboLights (Haug et al. 2013) and Reactome (EMBL-EBI, UK) (Fabregat et al. 2016; Milacic et al. 2012).

Cloud-based Computational Knowledgebase: To extract biological meaning from multi-omics datasets, researchers rely heavily on computational resources that analyze and annotate the molecules of interest with known information. Currently, access to tools/annotations is not straightforward because they reside in fragmented and incomplete repositories. To address this challenge, we created a novel distributed query system and cloud-based infrastructure, MetProt, that is capable of providing unified access to protein and metabolite datasets, allowing users to submit a single query to access multiple resources including Reactome (Fabregat et al. 2016; Milacic et al. 2012), UniProt (UniProt Consortium 2015), MetaboLights (Haug et al. 2013),

BioGPS (Wu et al. 2016; Wu, Macleod, and Su 2013), Gene Wiki (Tsueng et al. 2016), and COPaKB (N. C. Zong et al. 2013; H. Li et al. 2013). We also engineered customized application programming interfaces (APIs) to provide direct access for information in MetProt. These resources are described below.

COPaKB: COPaKB (N. C. Zong et al. 2013; H. Li et al. 2013; N. Zong et al. 2014) (http://heartproteome.org) is an omics analysis platform with 2 key components: (1) a peptide spectral search engine; and (2) spectral library modules for knowledge annotation.

Reactome: The Reactome (Milacic et al. 2012; Croft et al. 2014) (http://reactome.org) platform is a suite of network analysis tools for performing topology analysis and over-representation analysis of gene/protein networks. It comprises a manually curated database of human pathways with views of protein and metabolite structures overlaid with expression data. We interfaced with Reactome to identify key pathways of interest in our omics data.

MetaboLights: MetaboLights (http://ebi.ac.uk/metabolights/) is a public database dedicated to the submission and sharing of metabolomics data, mass spectra, annotated biological roles, and other derived information (Haug et al. 2013). Based on spectral similarities and chemical structures, we employed its search services to analyze and interpret metabolite data we collected from our studies.

BioGPS/Gene Wiki: BioGPS (http://biogps.org) is an interface for omics research (Wu et al. 2016). It provides a user-customizable portal with aggregated information on protein annotations and target list analysis. Gene Wiki translates molecular information into structured knowledge; relevant pages representing molecular transducers will be aggregated and recruited into MetProt.

1.IV. Overview and Rationale.

Heart failure (HF), which afflicts an estimated 6.5 million Americans and has incurred $6.5 billion in medical costs that are projected to grow within the next 20 years, remains a significant challenge to diagnosis and treat due to its heterogeneous clinical manifestations and multifactorial risk drivers (Benjamin et al. 2018; Dunbar et al. 2018). Maladaptive cardiac remodeling and heart failure (HF) are common stages of many heart diseases and pose major public health problems in the United States (McMurray 2010). To date, advanced HF patients are clinically treated as a homogenous group with similar standard therapies (WRITING COMMITTEE MEMBERS et al. 2013). However, HF is a multifactorial disease and results not only from cardiac overload or injury, but also from a complex interplay among genetic, neurohormonal, inflammatory, and biochemical alterations, requiring more sophisticated means to stratify patient groups for divergent therapies (S. J. Shah et al. 2015; Taylor et al. 2004). Over the past years, racial disparities in HF are well documented (East et al. 2004; Mensah et al. 2005) along with common comorbidities such as hypertension, obesity, kidney disease, and diabetes, which before the age of 40 can predict the severity and outcome of HF (Husaini et al. 2011).

Research efforts have been spurred to understand its underlying molecular mechanisms for developing targeted therapeutic strategies (A. M. Shah et al. 2014). Towards this goal, the advancement of analytical technologies has drastically increased the depth, quality and volume of data generated by both proteomics and metabolomics experiments. Both proteins and metabolites are essential building blocks of cellular processes that collaboratively formulate biological pathways. In addition, oxidative stress is becoming increasingly recognized as a key signaling event in the progression of heart failure (Maulik and Kumar 2012). In response to oxidative stress, O-PTMs of proteins have been shown to alter the 3-dimensional structure and therefore functional activities of proteins in various biological pathways contributing to

hypertrophy development (Qin et al. 2014; Birk et al. 2015; Bak and Weerapana 2015). However, many large datasets of either proteins, protein O-PTM, or metabolites are often fragmented and rarely integrated, hindering the understanding of mechanistic details of heart failure progression and the extraction of statistically significant and biologically relevant molecular signatures.

Here, we will deep phenotype heart failure progression via newly developed machine learning and statistical workflows and platforms for collective analyses of the landscape of oxidative stress-modified proteomics, metabolomics, and protein-metabolite networks. In Chapter 2, we performed a multi-faceted proteomics study, combined with a state-of-the-art computational pipeline, to elucidate key O-PTMs of molecular signatures and pathways in cardiac remodeling and heart failure progression. In Chapter 3, we established a Multiple Reaction Monitoring (MRM)-based MS quantitative platform and a corresponding computational workflow to unravel the metabolomics fingerprints of healthy humans, HF patients, and a mouse model. Accordingly, we quantified the plasma levels of 610 metabolites, and identified metabolites highly correlated to phenotypic changes. Our study highlights an established computational workflow that, in conjunction with a benchmark dataset, will facilitate future biomarker discovery. In chapter 4, we built a cloud-based platform, MetProt, for quantifying, triaging and analyzing omics datasets, and in doing so elucidating novel connections between proteins and metabolites, annotating molecular functions, and providing biomedical insights.

CHAPTER 2: COMPUTATIONAL APPROACHES TO DISSECT THE CYSTEINE O-PTM PROTEOME DURING CARDIAC HYPERTROPHY

2.I. Abstract.

Cysteine oxidative modification of cellular proteins is crucial for many aspects of cardiac hypertrophy development. However, integrated dissection of multiple types of cysteine oxidative post-translational modifications (O-PTM) of proteomes in cardiac hypertrophy is currently missing. Here we developed a novel discovery platform encompassing a customized biotin switch-based quantitative proteomics pipeline and an advanced analytic workflow to comprehensively profile the landscape of cysteine O-PTM in an ISO-induced cardiac hypertrophy mouse model. Specifically, we identified a total of 1,655 proteins containing 3,324 oxidized cysteine sites by at least one of the following three modifications: reversible cysteine O-PTM, cysteine sulfinylation (CysSO2H), and cysteine sulfonylation (CysSO3H). Analyzing the hypertrophy signatures that are reproducibly discovered from this computational workflow unveiled 4 biological processes with increased cysteine O-PTM. Among them, protein phosphorylation, creatine metabolism, and response to elevated Ca2+ pathways exhibited an elevation of cysteine O-PTM in early stages whereas glucose metabolism enzymes were increasingly modified in later stages, illustrating a temporal regulatory map in cardiac hypertrophy. Our cysteine O-PTM platform depicts a dynamic and integrated landscape of the cysteine oxidative proteome, through the extracted molecular signatures, and provides critical mechanistic insights in cardiac hypertrophy. Data are available via ProteomeXchange with identifier PXD010336.

2.II. Introduction.

Cardiac hypertrophy is an adaptive response of the heart to pressure overload and a risk factor for heart failure and sudden cardiac death (Shimizu and Minamino 2016; Maulik and Kumar 2012; Souders et al. 2012). The complex and dynamic pathophysiological mechanisms surrounding cardiac hypertrophy have been the focus of many investigations seeking therapeutic strategies. Reactive oxygen and nitrogen species (RO/NS) have been recognized

as second messengers that mediate important biological processes during the development of cardiac hypertrophy (Maulik and Kumar 2012; Murray and Van Eyk 2012). As a major target of RO/NS, protein cysteine residues can act as molecular switches and undergo various types of O-PTM (Lim et al. 2003), which can alter protein 3D structure and activity in cellular signaling, adaptive and maladaptive cardiac responses (Murray and Van Eyk 2012; Pastore and Piemonte 2013; Sag, Santos, and Shah 2014). Major cysteine O-PTM types observed in cardiac muscle include reversible (S-nitrosylation, S-glutathionylation, sulfenic acid, inter- and intramolecular disulfide bonds and S-sulfhydration) and irreversible sulfinic acid ($CysSO_2H$) or sulfonic acid ($CysSO_3H$) modifications (Murray and Van Eyk 2012; Alcock, Perkins, and Chalker 2018; Forrester and Stamler 2007).

As dysregulation of cysteine O-PTM is directly associated with disease pathology (Bechtel and Weerapana 2017; Murray and Van Eyk 2012), identification and characterization of cysteine O-PTM could yield a wider range of biomarkers and therapeutic targets for diseases characterized by oxidative stress. Due to the complexity of the cysteine oxidative proteome, elucidation of these molecular signatures during hypertrophy progression is currently lacking. Biotin switch assays and numerous iterations of this technique provide validated toolkits to customize the proteomic discovery platform and maximize the detection and quantification of the labile reversible cysteine O-PTM (Murray and Van Eyk 2012; Jaffrey and Snyder 2001). For example, work by Jennifer van Eyk's laboratory has advanced methodologies to specifically detect S-nitrosylation and measure individual protein thiol-reactivity, thereby discriminating artifacts and ambiguity surrounding site assignment of oxidation (Murray et al. 2013, 2012). Meanwhile, spline-based computational analysis enables extraction of temporal patterns (T.-Y. Liu et al. 2017; Bhasi, Forrest, and Ramanathan 2005).

Here, we present the first proteome-wide study of major cysteine modification types in cardiac hypertrophy, including reversible O-PTM, $CysSO_2H$, and $CysSO_3H$. We devised a biotin switch-based cysteine O-PTM discovery platform specific for multi-time point quantitative proteomic analysis at different disease stages of hypertrophic progression. To limit non-biologically oxidized peptides, we selected a maleimide-based reagent for enrichment labeling and combined dimethyl labeling for quantification. Comparing to IAM based labeling, maleimide provides higher specificity and reactivity with free thiols (Kramer et al. 2015; Reisz et al. 2013). In addition, dimethyl-based quantitative labeling is highly-efficient and cost-effective in studies with large sample size. We implemented the cubic spline-based computational workflow to capture the cysteine O-PTM clusters with distinct temporal profiles and high association with a hypertrophic phenotypic temporal profile. We define proteins in these clusters as molecular signatures. The novelty lies in this being the first study to link global oxidative proteomic signatures to distinct time points of hypertrophic progression, thus showcasing multi-factor, temporal fingerprints rather than a singular protein or snapshot at one specific time point.

2.III. Methods and Materials.

2.III.A. Experimental Animal Models.

Male C57BL/6J mice, 9–12 weeks of age (Jackson Laboratories), were housed in a 12-hour light/12-hour dark cycle with controlled temperature, humidity, and access to standard chow and water ad libitum. Mice were surgically implanted with a subcutaneous micro-osmotic pump (ALZET) delivering 15mg/kg/day isoproterenol (ISO) (Sigma) or saline vehicle (Lau et al. 2018, 2016). In this treatment protocol, mice develop a gradual cardiac hypertrophy phenotype characterized by significantly increased ejection fraction and HW/BW (Lau et al. 2016; Drews et al. 2010). Independent groups of 3 mice from each treatment condition were euthanized for sample collection at 1, 3, 5, 7, 10, and 14 days post-implantation. Four technical replicates were

performed for each treatment condition. All animal procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals by the National Research Council and approved by the Animal Research Committee at UCLA.

2.III.B. Biotin Switch-based Sample Processing

Protein extraction. Left ventricles were collected from mouse hearts and placed in 1mL NP-40 lysis buffer (50mM Tris–HCl, pH 8, 137mM NaCl, 10% glycerol, 1% NP-40, 2mM EDTA) containing 100mM N-ethylmaleimide (NEM) and Halt™ Protease Inhibitor Cocktail (100X) (ThermoFisher Scientific). Tissue was homogenized using a glass hand homogenizer and mixed by rotation for 2hr at 4°C before centrifugation at 13,800g for 20 min at 4°C. The protein concentration of the supernatant was measured using the DC™ protein assay (Bio-Rad).

Biotin switch-based labeling of reversible cysteine O-PTM. Aliquots containing 2mg protein were prepared, then precipitated by 10% trichloroacetic acid (TCA) and centrifuged at 20,000g for 15min at 4°C. One wash with ice-cold 5% TCA and two washes with 95% ethanol were performed to get rid of small molecules. Pellets were re-suspended and alkylated in 1mL of urea-containing cysteine modification buffer (CMBU) (0.1M HEPES-NaOH, pH7.4, 1% SDS, 10mM diethylenetriaminepentaacetic acid [DTPA], 6M urea) with 0.1M NEM. After 30min rotation at room temperature, the samples were reduced by 0.12M dithiothreitol (DTT) and rotated for another 60min. To quantify the total abundance of cysteine sites and to preserve cysteine sulfinylation ($CysSO_2H$) and sulfonylation ($CysSO_3H$), 10% of the lysate was reserved as an unlabeled portion and underwent acetone precipitation followed by digestion. For the remaining 90%, proteins were separated from small molecules by centrifugation with 10% TCA, followed by one wash with 5% TCA and two washes with 95% ethanol. Pellets were suspended in 300μL CMBU with 0.1mM maleimide-biotin (Mal-Biotin) (Sigma-Aldrich). After a 30min rotation at room temperature, unreacted NEM was quenched with 10mM DTT for an additional

30min. Small molecules and proteins were separated by TCA and ethanol washes as described above (J. Wang and Sevier 2016; J. Wang et al. 2014).

Digestion and dimethyl labeling. Pellets from both labeled and unlabeled portions were solubilized in 0.1M triethylammonium bicarbonate buffer (TAEB) with 0.1% Rapigest (Waters) and heated at 60°C for 45min. Solubilized proteins were alkylated with 9mM iodoacetamide (IAM) incubation in the dark at room temperature for 30min. The alkylated lysate underwent trypsin digestion overnight (16hr) at 37°C with a 1:100 ratio of trypsin to protein. A final concentration of 0.16% (vol/vol) CH_2O or C_2H_2O (Sigma-Aldrich) and 24mM sodium cyanoborohydride ($NaBH_3CN$) (Sigma-Aldrich) were added to the designated samples with light or medium labeling, respectively. Reciprocal labeling was performed on two out of four technical replicates to minimize the technical bias from dimethyl labeling. After a 1 h incubation at room temperature, 0.16% (vol/vol) ammonium solution was added and mixed for 15 min to quench the reaction (García-Santamarina et al. 2014; Gu and Robinson 2016; Boersema et al. 2009).

Mal-biotin enrichment. Mal-Biotin and dimethyl labeled peptides were diluted in 1.2mL PBS with 200µL pre-washed High-Capacity NeutrAvidin slurry (Thermo Scientific). After overnight incubation, the sample was centrifuged, washed twice with 1mL PBS, once with 50mM ammonium bicarbonate w/ 20% methanol, and eluted with 50% Pierce™ acetonitrile (ACN) w/ 0.4% trifluoroacetic acid (TFA) (Chung et al. 2015).

C18 column cleanup. All samples were subjected to 30min incubation at 37°C with 1% TFA and centrifugation (13,000g for 15min) to remove remaining Rapigest. Samples were cleaned with PierceTM C18 Spin Columns (Thermo Scientific) to remove any interfering substances prior to LC-MS/MS analysis.

2.III.C. LC-MS/MS Analysis.

LC-MS/MS was performed on digested peptides as previously described (Lau et al. 2018, 2016). To reduce sample complexity and increase protein coverage, we fractionated peptide samples using high-pH/low–pH two-dimensional reversed-phase chromatography prior to MS/MS. Fifty micrograms of peptides were injected into a Phenomenex C18 column (Jupiter Proteo C12, 4-μm particle, 90-Å pore, 100mm×1mm dimension) using a Finnigan Surveyor LC system (Thermo Scientific) for the first-dimension (high-pH) separation. We established a gradient between solvent A (20mM ammonium formate, pH 10) and solvent B (20mM ammonium formate, 90% acetonitrile, pH10) at a 50μL·min^{-1} flow-rate with the following timing and solvent proportions: 0–5% solvent B in solvent A from 0–2min; 5–35% solvent B in solvent A from 3–32min; and, finally, 80% solvent B in solvent A from 32–37min. Six fractions of peptides were collected from 16–40min, lyophilized, and re-dissolved in 20μL 0.5% formic acid with 2% acetonitrile. Each high-pH fraction was injected (10μL) to an EasySpray C18 column (PepMap, 3-μm particle, 100-Å pore; 75μm×150mm dimension; Thermo Scientific) using an auto-sampler on a single Easy-nLC 1000 nano-UPLC system (Thermo Scientific) for second-dimension (low-pH) reversed-phase chromatography analysis. We established a gradient between solvent A (0.1% formic acid, 2% acetonitrile) and solvent B (0.1% formic acid, 80% acetonitrile) at a flow rate of 300nL·min−1 with the following timing and solvent proportions: 0–40% solvent B from 0–110 min; 40–80% B from 110–117min; and 80% B from 117–120min. Column pressure was maintained below 150bar. High-resolution LC-MS/MS was performed on a single LTQ Orbitrap Elite instrument (Thermo Scientific) through a Thermo EasySpray interface. MS signals were acquired in Fourier-Transform/Ion-Trap (FT/IT) mode: each FT MS1 survey scan was analyzed at 400 to 2000 m/z mass range and 60,000 resolving power in profile mode, followed by rapid IT

MS2 scans on the top 15 ions with monoisotopic peak selection and 3000 intensity threshold. MS2 precursor isolation width was set to 2 m/z, normalized collision energy was 35, and charge state 1 and unassigned charge state were excluded. MS1 and MS2 target ion accumulation targets were 10^4 and 10^6, respectively. MS1 lock mass (m/z 425.120025) and dynamic exclusion (90s) were used. Throughout the LC-MS/MS experiment, column temperature was held at a constant 50°C.

2.III.D. Quantification of Cysteine O-PTM Abundance

The acquired raw mass spectra were processed with MaxQuant software (Cox and Mann 2008) version 1.5.6.0 as described (García-Santamarina et al. 2014). Peptide identification was performed using the Andromeda search engine (Cox et al. 2011), against a reverse-decoyed protein sequence database (UniProt Reference Proteome, reviewed, accessed June-12–2016). This *Mus musculus* proteome (taxonomy ID: 10090) database contains 33588 canonical sequences and does not include isoform information. Common contaminants were included in the database search. First and main searches were performed with precursor mass tolerances of 20ppm and 4.5ppm. Product ion tolerance was set to 0.5Da. Specificity for trypsin cleavage was required, allowing up to two missed cleavage sites (van der Reest et al. 2018). Dimethylated peptide labels were identified using the “multiplicity” query, including “DimethLys0” and “DimethN-term0” as light labels as well as “DimethLys4” and “DimethN4” as medium labels with a maximum four modified sites for each identified peptide. Variable modification types for enriched samples include Mal-biotin labeled cysteine (451.1889 Da), NEM-labeled cysteine (125.0477 Da), IAM-labeled cysteine (57.0215 Da), and methionine sulfoxidation (15.9949 Da) were queried with a maximum five modified sites for each identified peptide. Variable modification types for unlabeled whole tissue lysate samples include NEM-labeled cysteine (125.0477 Da), IAM-labeled cysteine (57.0215 Da), methionine sulfoxidation (15.9949 Da),

$CysSO_2H$ (31.9898 Da), and $CysSO_3H$ (47.9847 Da) were queried with a maximum five modified sites for each identified peptide. Tryptic, semi-tryptic, and non-tryptic peptides within a 20ppm parent mass window surrounding the candidate precursor mass were searched. Peptide ions from up to 3 isotopic peaks with fragment mass tolerance of 600ppm were allowed. Protein inference required ≤5% peptide spectra matching (PSM), posterior error probability (PEP), and ≤1% global level protein false discovery rate (FDR) (Q-value ≤1%), as well as a minimum of 2 ratio counts. Peptides with a cysteine count lower than one were excluded, along with reverse and potential contaminant flagged peptides. Modified peptide identifications with an Andromeda search score greater than 40, a delta score greater than 6, and a localization probability >0.8 were allowed (Bogdanow, Zauber, and Selbach 2016). All searches for a given data set were based on one set of Andromeda peak list files (apl-files). Each of the cysteine modifications (i.e. Mal-biotin, NEM, IAM, $CysSO_2H$, and $CysSO_3H$) was generated as a separate output file with identified cysteine sites, their extracted ion chromatogram (XIC) values, and normalized ratios of ISO vs. Vehicle conditions calculated from the differential dimethyl labeled peptides. To ensure data quality, technical replicates with significant change (p-value <0.05) in overall normalized ratio distribution among all four replicates were excluded. The total abundance of one cysteine site is quantified as the sum of XIC values from both modified and unmodified forms of that particular cysteine. Detailed information on number of spectra for identification and quantification, as well as number of unique peptide identified per experimental group are listed in Table 2.1. Furthermore, number of unique peptides, percent sequence coverage, heavy over light (H/L) ratio of mean peak area (± percent ratio H/L variability). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Deutsch et al. 2017) via the PRIDE (Vizcaíno et al. 2016) partner repository with the dataset identifier PXD010336.

Experimental Group	# Spectra for Identification	# Spectra for Quantification	# Unique Peptide Identified
Total cysteine_Day1_Rep1	545427	9521	4935
Total cysteine_Day1_Rep2	435853	19484	8035
Total cysteine_Day1_Rep3	613730	18328	7422
Total cysteine_Day1_Rep4	494484	18236	7156
Total cysteine_Day3_Rep1	493264	13098	6039
Total cysteine_Day3_Rep2	478093	19376	8885
Total cysteine_Day3_Rep3	572188	19778	8007
Total cysteine_Day3_Rep4	496781	19996	7752
Total cysteine_Day5_Rep1	513134	6201	3293
Total cysteine_Day5_Rep2	327835	10759	5743
Total cysteine_Day5_Rep3	626678	20711	8109
Total cysteine_Day5_Rep4	521663	16660	7057
Total cysteine_Day7_Rep1	533963	16200	6780
Total cysteine_Day7_Rep2	597921	15568	7120
Total cysteine_Day7_Rep3	609493	24065	10116
Total cysteine_Day7_Rep4	505500	20557	8011
Total cysteine_Day10_Rep1	614085	18505	6755
Total cysteine_Day10_Rep2	456761	21063	8392
Total cysteine_Day10_Rep3	602423	18246	7832
Total cysteine_Day10_Rep4	503221	18106	7063
Total cysteine_Day14_Rep1	586878	11614	5019
Total cysteine_Day14_Rep2	554245	18840	8224
Total cysteine_Day14_Rep3	604953	20311	8381
Total cysteine_Day14_Rep4	472933	17988	7126
Reversible cysteine O-PTM_Day1_Rep1	227533	5012	2048
Reversible cysteine O-PTM_Day1_Rep2	314860	5868	2080
Reversible cysteine O-PTM_Day1_Rep3	357770	4044	1688
Reversible cysteine O-PTM_Day1_Rep4	274523	2887	1397
Reversible cysteine O-PTM_Day3_Rep1	262151	4290	1891
Reversible cysteine O-PTM_Day3_Rep2	326416	6133	2061
Reversible cysteine O-PTM_Day3_Rep3	284474	3457	1579
Reversible cysteine O-PTM_Day3_Rep4	252848	3052	1730
Reversible cysteine O-PTM_Day5_Rep1	213077	5751	2401
Reversible cysteine O-PTM_Day5_Rep2	337457	6302	2123
Reversible cysteine O-PTM_Day5_Rep3	296146	3761	1690
Reversible cysteine O-PTM_Day5_Rep4	315913	4370	1862
Reversible cysteine O-PTM_Day7_Rep1	312953	7326	2797
Reversible cysteine O-PTM_Day7_Rep2	238772	3392	1568
Reversible cysteine O-PTM_Day7_Rep3	278035	2659	1346
Reversible cysteine O-PTM_Day7_Rep4	235911	1831	1280

Experimental Group	# Spectra for Identification	# Spectra for Quantification	# Unique Peptide Identified
Reversible cysteine O-PTM_Day10_Rep1	177615	3464	1588
Reversible cysteine O-PTM_Day10_Rep2	336361	5053	1733
Reversible cysteine O-PTM_Day10_Rep3	273030	2453	1332
Reversible cysteine O-PTM_Day10_Rep4	252848	3052	1730
Reversible cysteine O-PTM_Day14_Rep1	179144	3327	1584
Reversible cysteine O-PTM_Day14_Rep2	339793	6088	2007
Reversible cysteine O-PTM_Day14_Rep3	261463	2254	1136
Reversible cysteine O-PTM_Day14_Rep4	238437	3850	2067

Table 2.1. Summary information of raw file contents and statistics for all experimental groups.

2.III.E. Data Analysis.

Cubic spline-based temporal clustering. Mal-biotin labeled cysteine sites or total cysteine sites with abundance values represented in at least 4 out of 6 time points in both ISO and Vehicle groups were selected for cubic spline-based temporal clustering. The averaged ratio of abundance in ISO to Vehicle across replicates was calculated per site. Missing abundance values for a modification site were imputed using average abundance of remaining time-points. After scaling and centering, cubic splines were fitted to ratios across the time points in the R statistical programming language (v3.4.3). The predicted abundance ratios from cubic spline were used for K-mean clustering (kmean package in R) following identification of the cluster numbers (mclust package in R). All the codes used for the analyses were deposited to the Github public repository (https://github.com/UCLA-BD2K/Cubic-Spline-based-Temporal-Analysis-Workflow). The statistical significance of abundance change in ISO and Vehicle conditions was evaluated by a paired two-sided t-test. Pearson correlation was performed to analyze the association between phenotype and abundance of Mal-biotin labeled cysteines.

Functional annotations. Cellular pathway information was retrieved from Reactome (release v63, 2017_12) (https://reactome.org/) (Fabregat et al. 2018). To exclude single entity function

and non-specific/generalized pathway terms (e.g. metabolism), enriched pathways are required to house a minimum 2 found entities and a maximum of 200 total entities. Significantly enriched pathways are defined by FDR <0.05 calculated from Fisher's exact test with FDR multiple test correction. NeXtProt (release 2018-02) was implemented to annotate the biological processes and disease association of key proteins identified in the analysis (Gaudet et al. 2017).

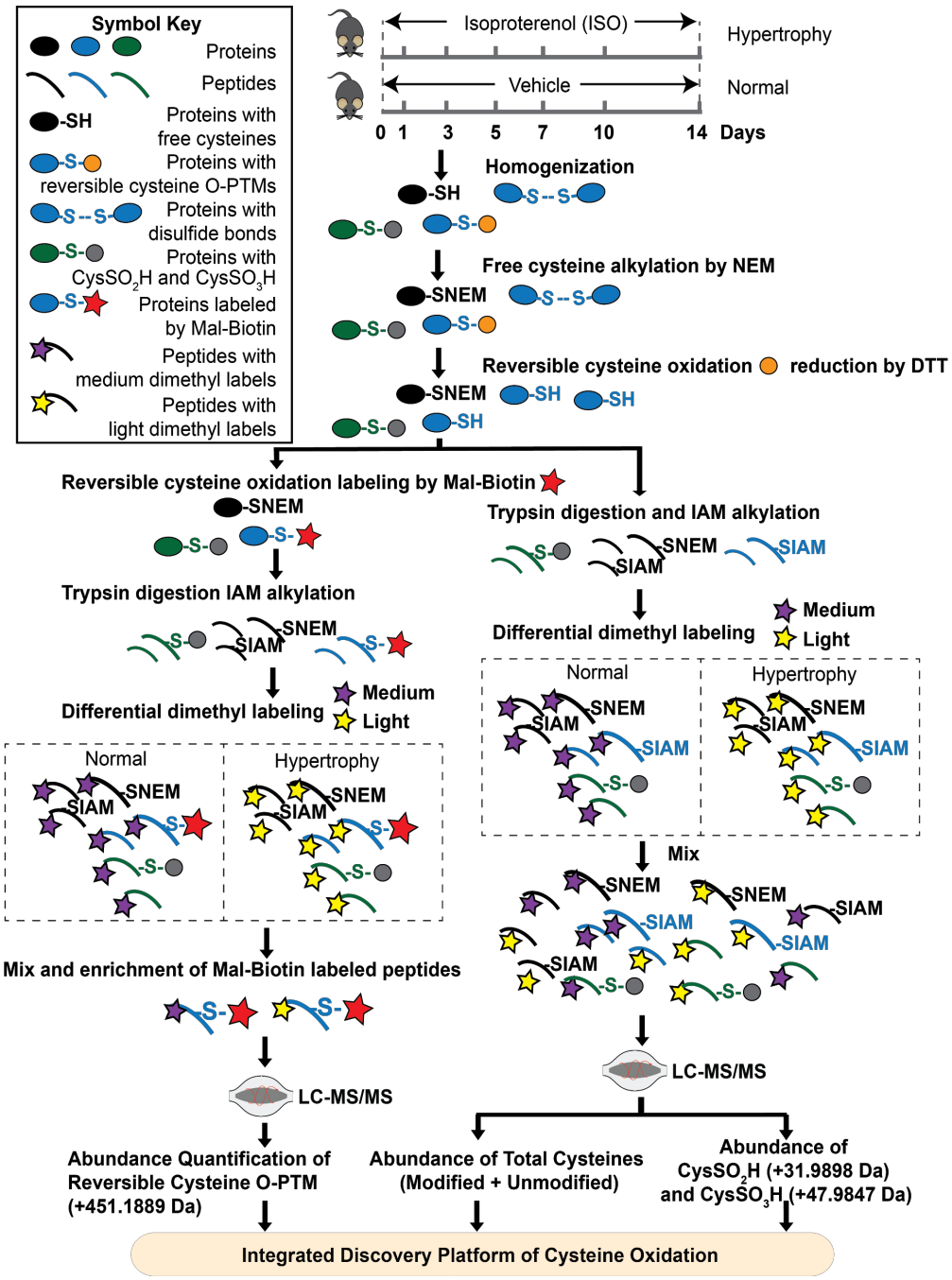


Figure 2.1. Quantitative cysteine O-PTM proteomics workflow. Left ventricular tissue samples originated from C57BL/6J were subjected to ISO (Hypertrophy) conditions or saline vehicle (Normal) for 1, 3, 5, 7, 10, and 14 days, with n = 3 mice per group. Extracted proteins underwent biotin switch for reversible cysteine O-PTM labeling. First, free cysteines in the cell lysate were alkylated with NEM to prevent nonspecific labeling of reversible cysteine O-PTM. Subsequently, reversibly oxidized cysteine were reduced with DTT. At this point, 10% of the lysate was separated out for the quantification of total cysteine site abundance and DTT-irreversible cysteine O-PTMs (i.e., CysSO2H and CysSO3H) (right side of flowchart). The remaining 90% (left side of flowchart) was treated with Mal-Biotin, which labels newly-reduced, free cysteine residues. Following trypsin digestion and IAM alkylation of newly exposed free cysteines, lysates from Normal and Hypertrophy groups were differentially labeled with medium/light dimethyl in two of the four technical replicates, whereas the other two received reciprocol labeling. The dimethyl-labeled peptides were quenched by ammonium and equally mixed. The Mal-Biotin labeled portion was enriched by avidin agarose and the abundance of reversible cysteine O-PTM was quantified by LC-MS/MS. Following differential dimethyl labeling, the unlabeled 10% (right side) underwent LC-MS/MS to quantify the abundance of total cysteines and DTT-irreversible cysteine O-PTM. The identification of modified peptides is based on their mass changes by corresponding modifications, i.e., +451.1889 Da for reversible cysteine O-PTM, +31.9898 Da for CysSO2H, and +47.9847 Da for CysSO3H. Quantification of modifications is based on XIC values. Integrative analysis was performed to characterize the impact of ISO-induced cardiac hypertrophy on the cysteine O-PTM profile. Abbreviations: oxidative post-translational modification (O-PTM); isoproterenol (ISO); N-ethylmaleimide (NEM); dithiothreitol (DTT); cystein sulfinylation (CysSO2H); cystein sulfonylation (CysSO3H); maleimide biotin (Mal-Biotin); extracted ion chromatogram (XIC); iodoacetamide (IAM).

2.IV. Results.

2.IV.A. Overview.

We developed a cysteine O-PTM discovery platform by integrating a biotin switch-based quantitative proteomics approach with advanced computational analysis, and we applied the novel platform on an ISO-induced cardiac hypertrophic mouse model. First, we defined the abundance distribution of three types of cysteine O-PTM, including reversible $CysSO_2H$, and $CysSO_3H$ during the progression of hypertrophy. Secondly, we applied statistical analysis to identify key proteins with significantly increased or decreased cysteine O-PTM abundance, followed by cubic spline-based K-mean clustering to dissect the temporal profiles of cysteine O-PTM together with their total cysteine abundance. Thirdly, temporal signatures along with their enriched hypertrophic pathways were identified. And finally, we established the association between cysteine O-PTM profiles and hypertrophic phenotypes, keying in on cysteine sites that were significantly correlated with the phenotype dynamics. Integrated analysis of key sites extracted from our computational approaches was performed to comprehensively identify

molecular signatures and provide mechanistic insights during the development of cardiac hypertrophy.

2.IV.B. Impact of ISO-induced hypertrophy on cysteine O-PTM.

Using our cysteine O-PTM discovery pipeline on mouse LV tissue from ISO and Vehicle groups (Figure 2.1.), we identified a total of 2,505 proteins containing 6,818 cysteine sites; with 1,655 proteins containing 3,324 sites modified by at least one of three cysteine O-PTM types. Specifically, 1,095 proteins containing 2,162 sites, 484 proteins containing 671 sites, and 447 proteins containing 613 sites were modified by reversible modification, CysSO2H, and CysSO3H, respectively. According to Uniprot, among the reversibly modified proteins and cysteine sites only 167 proteins containing 470 sites and 8 proteins containing 9 sites are known sites of disulfide-bonds and nitrosylation, respectively, revealing that the majority of identified cysteine sites are novel. Figure 2.2.A illustrates the overall change in abundance distribution of reversible cysteine O-PTM over the progression of hypertrophy, with the average abundance of reversible cysteine O-PTM sites across 4 technical replicates plotted for each time point. All histograms exhibit a J-shaped curve, demonstrating the higher frequency of modified cysteine sites concentrated in the low abundance XIC range. Notably, the distribution of O-PTM abundance in ISO demonstrates a sizeable reduction in the lower abundance range, indicating a global enrichment of O-PTM occurring as early as Day 3. Figure 2.2.B shows a Venn diagram to depict the number of proteins associated with each individual type of cysteine modification. To summarize, 1,057 sites within 377 proteins are unique to an enriched reversible O-PTM proteome, demonstrating the advantage of using enrichment to concentrate less abundant but highly oxidized proteins. The overlapping cysteine sites among three modification types are limited (5 sites), suggesting that this arrangement of cysteine O-PTM is highly site-specific. Comparatively, the numbers of proteins co-modified with all three modification types on separate residues are significantly higher, suggesting that diverse cysteine O-PTM patterns

occur on the same protein albeit in a site-specific manner. Overall, these data indicate that the oxidative regulation of proteins requires complex orchestration of myriad oxidative chemistries in a site-specific manner.

Accession	Cysteine	Fold Change	P Value	Signature	Oxidation Type	Entry Name	Gene Name	Protein Name	Molecular Weight
A2ASQ1	A2ASQ1_C441	0.842	0.026	Decreased Reversible Modification	Reversible cysteine O-PTM	AGRIN_MOUSE	Agrn Agrin	Agrin [Cleaved into: Agrin N-terminal 110 kDa subunit; Agrin C-terminal 110 kDa subunit; Agrin C-terminal 90 kDa fragment (C90); Agrin C-terminal 22 kDa fragment (C22)]	207,539
A2AX52	A2AX52_C211	0.759	0.028	Decreased Reversible Modification	Reversible cysteine O-PTM	CO6A4_MOUSE	Col6a4 Dvwa	Collagen alpha-4(VI) chain	250,798
B2RXS4	B2RXS4_C617	0.914	0.04	Decreased Reversible Modification	Reversible cysteine O-PTM	PLXB2_MOUSE	Plxnb2	Plexin-B2	206,230
O08528	O08528_C368	0.715	0.029	Decreased Reversible Modification	Reversible cysteine O-PTM	HXK2_MOUSE	Hk2	Hexokinase-2 (EC 2.7.1.1) (Hexokinase type II) (HK II)	102,535
O08796	O08796_C145	0.876	0	Decreased Reversible Modification	Reversible cysteine O-PTM	EF2K_MOUSE	Eef2k	Eukaryotic elongation factor 2 kinase (eEF-2 kinase) (eEF-2K) (EC 2.7.11.20) (Calcium/calmodulin-dependent eukaryotic elongation factor 2 kinase)	81,739
O35459	O35459_C91	0.921	0.011	Decreased Reversible Modification	Reversible cysteine O-PTM	ECH1_MOUSE	Ech1	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial (EC 5.3.3.-)	36,118
O35855	O35855_C148	0.805	0.004	Decreased Reversible Modification	Reversible cysteine O-PTM	BCAT2_MOUSE	Bcat2 Bcatm Eca40	Branched-chain-amino-acid aminotransferase, mitochondrial (BCAT(m)) (EC 2.6.1.42)	44,127
O55143	O55143_C349	0.924	0.031	Decreased Reversible Modification	Reversible cysteine O-PTM	AT2A2_MOUSE	Atp2a2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (SERCA2) (SR Ca(2+)-ATPase 2) (EC 3.6.3.8)	114,858
O55143	O55143_C344	0.833	0.016	Decreased Reversible Modification	Reversible cysteine O-PTM	AT2A2_MOUSE	Atp2a2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (SERCA2) (SR Ca(2+)-ATPase 2) (EC 3.6.3.8)	114,858
P00397	P00397_C498	0.932	0.046	Decreased Reversible Modification	Reversible cysteine O-PTM	COX1_MOUSE	Mtco1 COI mt-Co1	Cytochrome c oxidase subunit 1 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I)	56,910

Accession	Cysteine	Fold Change	P Value	Signature	Oxidation Type	Entry Name	Gene Name	Protein Name	Molecular Weight
P02469	P02469_C190	0.894	0.043	Decreased Reversible Modification	Reversible cysteine O-PTM	LAMB1_MOUSE	Lamb1 Lamb-1 Lamb1-1	Laminin subunit beta-1 (Laminin B1 chain) (Laminin-1 subunit beta) (Laminin-10 subunit beta) (Laminin-12 subunit beta) (Laminin-2 subunit beta) (Laminin-6 subunit beta) (Laminin-8 subunit beta)	197,090
P10493	P10493_C616	0.869	0.037	Decreased Reversible Modification	Reversible cysteine O-PTM	NID1_MOUSE	Nid1 Ent	Nidogen-1 (NID-1) (Entactin)	136,538
P14094	P14094_C214	0.915	0.049	Decreased Reversible Modification	Reversible cysteine O-PTM	AT1B1_MOUSE	Atp1b1 Atp4b	Sodium/potassium-transporting ATPase subunit beta-1 (Sodium/potassium-dependent ATPase subunit beta-1)	35,195
P18242	P18242_C327	0.848	0.018	Decreased Reversible Modification	Reversible cysteine O-PTM	CATD_MOUSE	Ctsd	Cathepsin D (EC 3.4.23.5)	44,954
P23242	P23242_C198	0.895	0.004	Decreased Reversible Modification	Reversible cysteine O-PTM	CXA1_MOUSE	Gja1 Cxn-43	Gap junction alpha-1 protein (Connexin-43) (Cx43) (Gap junction 43 kDa heart protein)	43,004
P32261	P32261_C54	0.939	0.043	Decreased Reversible Modification	Reversible cysteine O-PTM	ANT3_MOUSE	Serpinc1 At3	Antithrombin-III (ATIII) (Serpin C1)	52,004
P41216	P41216_C298	0.868	0.006	Decreased Reversible Modification	Reversible cysteine O-PTM	ACSL1_MOUSE	Acsl1 Acsl2 Facl2	Long-chain-fatty-acid--CoA ligase 1 (EC 6.2.1.3) (Long-chain acyl-CoA synthetase 1) (LACS 1)	77,951
P41216	P41216_C510	0.81	0	Decreased Reversible Modification	Reversible cysteine O-PTM	ACSL1_MOUSE	Acsl1 Acsl2 Facl2	Long-chain-fatty-acid--CoA ligase 1 (EC 6.2.1.3) (Long-chain acyl-CoA synthetase 1) (LACS 1)	77,951
P41216	P41216_C626	0.874	0.015	Decreased Reversible Modification	Reversible cysteine O-PTM	ACSL1_MOUSE	Acsl1 Acsl2 Facl2	Long-chain-fatty-acid--CoA ligase 1 (EC 6.2.1.3) (Long-chain acyl-CoA synthetase 1) (LACS 1)	77,951
P58281	P58281_C856	0.772	0.022	Decreased Reversible Modification	Reversible cysteine O-PTM	OPA1_MOUSE	Opa1	Dynamin-like 120 kDa protein, mitochondrial (EC 3.6.5.5) (Large GTP-binding protein) (LargeG) (Optic atrophy protein 1 homolog) [Cleaved into: Dynamin-like 120 kDa protein, form S1]	111,339
P58281	P58281_C786	0.91	0.012	Decreased Reversible Modification	Reversible cysteine O-PTM	OPA1_MOUSE	Opa1	Dynamin-like 120 kDa protein, mitochondrial (EC 3.6.5.5) (Large GTP-binding protein) (LargeG) (Optic atrophy protein 1 homolog) [Cleaved into: Dynamin-like 120 kDa protein, form S1]	111,339

Accession	Cysteine	Fold Change	P Value	Signature	Oxidation Type	Entry Name	Gene Name	Protein Name	Molecular Weight
P97370	P97370_C249	0.816	0.024	Decreased Reversible Modification	Reversible cysteine O-PTM	AT1B3_MOUSE	Atp1b3	Sodium/potassium-transporting ATPase subunit beta-3 (Sodium/potassium-dependent ATPase subunit beta-3) (ATPB-3) (CD antigen CD298)	31,776
P97443	P97443_C125	0.886	0.024	Decreased Reversible Modification	Reversible cysteine O-PTM	SMYD1_MOUSE	Smyd1 Bop	Histone-lysine N-methyltransferase Smyd1 (EC 2.1.1.43) (CD8b-opposite) (SET and MYND domain-containing protein 1) (Zinc finger protein BOP) (m-BOP)	56,496
Q02566	Q02566_C949	0.856	0.012	Decreased Reversible Modification	Reversible cysteine O-PTM	MYH6_MOUSE	Myh6 Myhca	Myosin-6 (Myosin heavy chain 6) (Myosin heavy chain, cardiac muscle alpha isoform) (MyHC-alpha)	223,565
Q04592	Q04592_C1360	0.94	0.01	Decreased Reversible Modification	Reversible cysteine O-PTM	PCSK5_MOUSE	Pcsk5	Proprotein convertase subtilisin/kexin type 5 (EC 3.4.21.-) (Proprotein convertase 5) (PC5) (Proprotein convertase 6) (PC6) (Subtilisin-like proprotein convertase 6) (SPC6) (Subtilisin/kexin-like protease PC5)	209,257
Q60675	Q60675_C1447	0.676	0	Decreased Reversible Modification	Reversible cysteine O-PTM	LAMA2_MOUSE	Lama2	Laminin subunit alpha-2 (Laminin M chain) (Laminin-12 subunit alpha) (Laminin-2 subunit alpha) (Laminin-4 subunit alpha) (Merosin heavy chain)	343,815
Q60936	Q60936_C403	0.894	0.02	Decreased Reversible Modification	Reversible cysteine O-PTM	COQ8A_MOUSE	Coq8a Adck3 Cabc1	Atypical kinase COQ8A, mitochondrial (EC 2.7.-.-) (Chaperone activity of bc1 complex-like) (Chaperone-ABC1-like) (Coenzyme Q protein 8A) (aarF domain-containing protein kinase 3)	71,743
Q61292	Q61292_C505	0.894	0.027	Decreased Reversible Modification	Reversible cysteine O-PTM	LAMB2_MOUSE	Lamb2 Lams	Laminin subunit beta-2 (Laminin-11 subunit beta) (Laminin-14 subunit beta) (Laminin-15 subunit beta) (Laminin-3 subunit beta) (Laminin-4 subunit beta) (Laminin-7 subunit beta) (Laminin-9 subunit beta) (S-laminin subunit beta) (S-LAM beta)	196,579
Q61738	Q61738_C103	0.556	0.023	Decreased Reversible Modification	Reversible cysteine O-PTM	ITA7_MOUSE	Itga7	Integrin alpha-7 [Cleaved into: Integrin alpha-7 heavy chain; Integrin alpha-7 light chain]	129,329
Q7TQ48	Q7TQ48_C899	0.671	0.002	Decreased Reversible Modification	Reversible cysteine O-PTM	SRCA_MOUSE	Srl Sar	Sarcalumenin	99,184
Q80TZ3	Q80TZ3_C372	0.769	0	Decreased Reversible Modification	Reversible cysteine O-PTM	AUXI_MOUSE	Dnajc6 Kiaa0473	Putative tyrosine-protein phosphatase auxilin (EC 3.1.3.48) (DnaJ homolog subfamily C member 6)	102,299

Accession	Cysteine	Fold Change	P Value	Signature	Oxidation Type	Entry Name	Gene Name	Protein Name	Molecular Weight
Q8BGK2	Q8BGK2_C206	0.674	0.027	Decreased Reversible Modification	Reversible cysteine O-PTM	ARHL1_MOUSE	Adprhl1 Arh2	[Protein ADP-ribosylarginine] hydrolase-like protein 1 (EC 3.2.-.-) (ADP-ribosylhydrolase 2)	39,885
Q8CC88	Q8CC88_C415	0.868	0.019	Decreased Reversible Modification	Reversible cysteine O-PTM	VWA8_MOUSE	Vwa8 Kiaa0564	von Willebrand factor A domain-containing protein 8	213,421
Q8CFX1	Q8CFX1_C621	0.765	0.025	Decreased Reversible Modification	Reversible cysteine O-PTM	G6PE_MOUSE	H6pd	GDH/6PGL endoplasmic bifunctional protein [Includes: Glucose 1-dehydrogenase (EC 1.1.1.47) (Glucose-6-phosphate dehydrogenase) (EC 1.1.1.363); 6-phosphogluconolactonase (6PGL) (EC 3.1.1.31)]	88,928
Q8K182	Q8K182_C140	0.812	0	Decreased Reversible Modification	Reversible cysteine O-PTM	CO8A_MOUSE	C8a	Complement component C8 alpha chain (Complement component 8 subunit alpha)	66,080
Q8K182	Q8K182_C194	0.749	0.001	Decreased Reversible Modification	Reversible cysteine O-PTM	CO8A_MOUSE	C8a	Complement component C8 alpha chain (Complement component 8 subunit alpha)	66,080
Q8K2B3	Q8K2B3_C536	0.88	0.002	Decreased Reversible Modification	Reversible cysteine O-PTM	SDHA_MOUSE	Sdha	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (EC 1.3.5.1) (Flavoprotein subunit of complex II) (Fp)	72,585
Q91VD9	Q91VD9_C367	0.935	0.016	Decreased Reversible Modification	Reversible cysteine O-PTM	NDUS1_MOUSE	Ndufs1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-75kD) (CI-75kD)	79,777
Q91VD9	Q91VD9_C727	0.924	0.004	Decreased Reversible Modification	Reversible cysteine O-PTM	NDUS1_MOUSE	Ndufs1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-75kD) (CI-75kD)	79,777
Q91WD5	Q91WD5_C347	0.928	0.022	Decreased Reversible Modification	Reversible cysteine O-PTM	NDUS2_MOUSE	Ndufs2	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-49kD) (CI-49kD) (NADH-ubiquinone oxidoreductase 49 kDa subunit)	52,626
Q99J39	Q99J39_C359	0.786	0.037	Decreased Reversible Modification	Reversible cysteine O-PTM	DCMC_MOUSE	Mlycd	Malonyl-CoA decarboxylase, mitochondrial (MCD) (EC 4.1.1.9)	54,736
Q99JY8	Q99JY8_C68	0.913	0.012	Decreased Reversible Modification	Reversible cysteine O-PTM	PLPP3_MOUSE	Plpp3 Lpp3 Ppap2b	Phospholipid phosphatase 3 (EC 3.1.3.4) (Lipid phosphate phosphohydrolase 3) (PAP2-beta) (Phosphatidate phosphohydrolase type 2b)	35,216

Accession	Cysteine	Fold Change	P Value	Signature	Oxidation Type	Entry Name	Gene Name	Protein Name	Molecular Weight
Q9CQN1	Q9CQN1_C503	0.806	0.016	Decreased Reversible Modification	Reversible cysteine O-PTM	TRAP1_MOUSE	Trap1 Hsp75	Heat shock protein 75 kDa, mitochondrial (HSP 75) (TNFR-associated protein 1) (Tumor necrosis factor type 1 receptor-associated protein) (TRAP-1)	80,209
Q9CRB9	Q9CRB9_C193	0.79	0.007	Decreased Reversible Modification	Reversible cysteine O-PTM	MIC19_MOUSE	Chchd3 Mic19	MICOS complex subunit Mic19 (Coiled-coil-helix-coiled-coil-helix domain-containing protein 3)	26,335
Q9CRB9	Q9CRB9_C183	0.893	0.006	Decreased Reversible Modification	Reversible cysteine O-PTM	MIC19_MOUSE	Chchd3 Mic19	MICOS complex subunit Mic19 (Coiled-coil-helix-coiled-coil-helix domain-containing protein 3)	26,335
Q9Z2I9	Q9Z2I9_C430	0.9	0.016	Decreased Reversible Modification	Reversible cysteine O-PTM	SUCB1_MOUSE	Sucla2	Succinate--CoA ligase [ADP-forming] subunit beta, mitochondrial (EC 6.2.1.5) (ATP-specific succinyl-CoA synthetase subunit beta) (A-SCS) (Succinyl-CoA synthetase beta-A chain) (SCS-betaA)	50,114

Table 2.2. List of cysteine sites with alterations in modification abundance during cardiac hypertrophy.

We mapped the distribution of O-PTM abundance, expressed as a ratio of ISO over Vehicle across 6 time points in violin plots for the three modification types (Figure 2.2.C). Consistent with 2.2.A, the ratio means of reversible O-PTM on Day 3, 5, 10, and 14 in ISO over Vehicle are slightly above zero, demonstrating an increase in reversible O-PTM abundance. CysSO2H oxidized sites exhibit no distribution change whereas CysSO3H oxidized sites have upshifts on Day 5 and 14. The distinct patterns of abundance distribution among the three modification types indicate O-PTM type-specific proteomic regulation during ISO-induced hypertrophy and oxidative stress.

To identify sites of cysteine O-PTM that are significantly altered by ISO treatment, we applied a t-test on the log2-transformed fold-change of the reversible cysteine O-PTM (ISO/Vehicle) with at least 3 time points, where a p-value <0.05 and fold-change within the top or bottom 20th

percentile were deemed significant. For reversible cysteine O-PTM (Figure 2.2.D, top panel), we found 128 sites within 74 proteins with significantly increased abundance and 46 sites within 39 proteins with significantly decreased abundance during ISO treatment, marking their potential to be signatures of cardiac hypertrophy. In addition, one and two CysSO3H oxidized sites are significantly increased and decreased, respectively (Figure 2.2.D, lower panel).

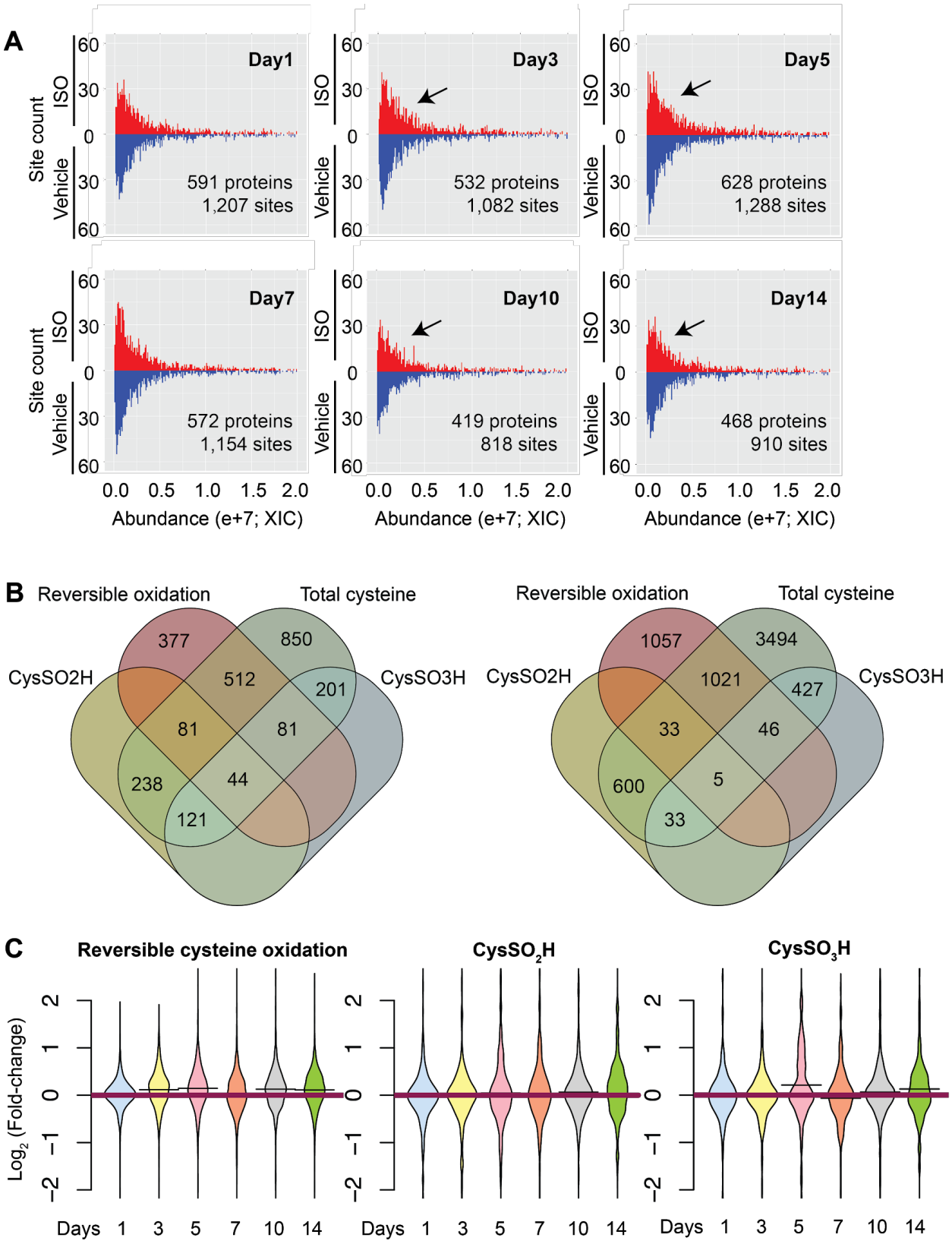
Pathway Name	# Entities Found	# Entities Total	Entities FDR	Submitted Entities Found	Signature
Laminin interactions	4	35	0.001	Q60675;P10493;Q61292;P02469	Significantly Decreased
MET activates PTK2 signaling	3	36	0.011	Q60675;Q61292;P02469	Significantly Decreased
MET promotes cell motility	3	59	0.025	Q60675;Q61292;P02469	Significantly Decreased
Ion homeostasis	3	61	0.025	O55143;P14094;P97370	Significantly Decreased
The citric acid (TCA) cycle and respiratory electron transport	4	156	0.026	Q91VD9;Q9Z2I9;Q91WD5;Q8K2B3	Significantly Decreased
Ion transport by P-type ATPases	3	76	0.026	O55143;P14094;P97370	Significantly Decreased
Citric acid cycle (TCA cycle)	2	26	0.038	Q9Z2I9;Q8K2B3	Significantly Decreased
Basigin interactions	2	32	0.052	P14094;P97370	Significantly Decreased
Signaling by MET	3	110	0.052	Q60675;Q61292;P02469	Significantly Decreased
Complex I biogenesis	2	59	0.083	Q91VD9;Q91WD5	Significantly Decreased
Post-translational protein phosphorylation	3	165	0.083	P32261;Q61292;P02469	Significantly Decreased
Pyruvate metabolism and Citric Acid (TCA) cycle	2	61	0.083	Q9Z2I9;Q8K2B3	Significantly Decreased
Cardiac conduction	3	169	0.083	O55143;P14094;P97370	Significantly Decreased
Respiratory electron transport	2	71	0.088	Q91VD9;Q91WD5	Significantly Decreased
Peroxisomal protein import	2	79	0.105	O35459;Q99J39	Significantly Decreased
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	2	95	0.111	Q91VD9;Q91WD5	Significantly Decreased
Degradation of the extracellular matrix	2	172	0.111	P10493;P18242	Significantly Decreased
Gluconeogenesis	5	40	0.002	P14152;P21550;P08249;P17751;P05201	Significantly Increased
Formation of the ternary complex, and subsequently	5	61	0.004	P62242;P97351;P62908;P25444;P62754	Significantly Increased

Pathway Name	# Entities Found	# Entities Total	Entities FDR	Submitted Entities Found	Signature
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	6	104	0.004	P62242;P97351;P62908;P62717;P25444;P62754	Significantly Increased
Nonsense-Mediated Decay (NMD)	6	104	0.004	P62242;P97351;P62908;P62717;P25444;P62754	Significantly Increased
Ribosomal scanning and start codon recognition	5	66	0.004	P62242;P97351;P62908;P25444;P62754	Significantly Increased
Cap-dependent Translation Initiation	5	73	0.004	P62242;P97351;P62908;P25444;P62754	Significantly Increased
Eukaryotic Translation Initiation	5	73	0.004	P62242;P97351;P62908;P25444;P62754	Significantly Increased
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	5	95	0.01	Q03265;Q99LC5;P52503;Q91YT0;Q9CQJ8	Significantly Increased
Protein methylation	3	21	0.01	Q01853;P10126;P25444	Significantly Increased
The citric acid (TCA) cycle and respiratory electron transport	6	156	0.012	Q03265;Q99LC5;P08249;P52503;Q91YT0;Q9CQJ8	Significantly Increased
Post-translational protein phosphorylation	6	165	0.015	Q61147;Q61554;P07724;P01027;Q921I1;P11276	Significantly Increased
Glucose metabolism	5	119	0.022	P14152;P21550;P08249;P17751;P05201	Significantly Increased
Respiratory electron transport	4	71	0.023	Q99LC5;P52503;Q91YT0;Q9CQJ8	Significantly Increased
HSF1 activation	2	8	0.026	Q01853;P10126	Significantly Increased
Molecules associated with elastic fibres	3	45	0.058	Q61554;Q8K4G1;P11276	Significantly Increased
Elastic fibre formation	3	52	0.079	Q61554;Q8K4G1;P11276	Significantly Increased
Creatine metabolism	2	16	0.082	Q04447;Q6P8J7	Significantly Increased
Translation	5	185	0.095	P62242;P97351;P62908;P25444;P62754	Significantly Increased
Complex I biogenesis	3	59	0.095	P52503;Q91YT0;Q9CQJ8	Significantly Increased
Mitochondrial biogenesis	2	27	0.189	P26443;Q03265	Significantly Increased
Plasma lipoprotein remodeling	2	30	0.21	P07724;P20918	Significantly Increased
Amino acid synthesis and interconversion (transamination)	2	33	0.22	P26443;P05201	Significantly Increased
Metabolism of polyamines	3	91	0.22	Q04447;Q6P8J7;Q60692	Significantly Increased
Platelet degranulation	4	166	0.22	P07724;P20918;Q921I1;P11276	Significantly Increased
Degradation of the extracellular matrix	4	172	0.226	P02463;Q61554;P20918;P11276	Significantly Increased
Response to elevated platelet cytosolic Ca2+	4	172	0.226	P07724;P20918;Q921I1;P11276	Significantly Increased
Non-integrin membrane-ECM interactions	2	43	0.268	P02463;P11276	Significantly Increased
Integrin cell surface interactions	3	110	0.268	P02463;Q61554;P11276	Significantly Increased

Pathway Name	# Entities Found	# Entities Total	Entities FDR	Submitted Entities Found	Signature
Detoxification of Reactive Oxygen Species	2	47	0.285	P24270;P46412	Significantly Increased
Cross-presentation of soluble exogenous antigens (endosomes)	2	50	0.29	Q61830;Q60692	Significantly Increased
MET promotes cell motility	2	59	0.3	P63001;P11276	Significantly Increased
Semaphorin interactions	2	63	0.3	P63001;O08553	Significantly Increased
Hedgehog ligand biogenesis	2	68	0.3	Q01853;Q60692	Significantly Increased
Plasma lipoprotein assembly, remodeling, and clearance	2	71	0.3	P07724;P20918	Significantly Increased
Iron uptake and transport	2	73	0.3	Q61147;Q921I1	Significantly Increased
MAPK6/MAPK4 signaling	2	80	0.3	P63001;Q60692	Significantly Increased
Signaling by PDGF	2	81	0.3	P02463;P20918	Significantly Increased
rRNA processing in the nucleus and cytosol	2	88	0.3	P25444;P62754	Significantly Increased
rRNA processing	2	88	0.3	P25444;P62754	Significantly Increased
Major pathway of rRNA processing in the nucleolus and cytosol	2	88	0.3	P25444;P62754	Significantly Increased
L1CAM interactions	2	88	0.3	P63001;O08553	Significantly Increased
Scavenging of heme from plasma	2	89	0.3	P07724;Q07456	Significantly Increased
PCP/CE pathway	2	94	0.3	P63001;Q60692	Significantly Increased
Glycolysis	2	98	0.3	P21550;P17751	Significantly Increased
Binding and Uptake of Ligands by Scavenger Receptors	2	108	0.3	P07724;Q07456	Significantly Increased
Cellular response to heat stress	2	110	0.3	Q01853;P10126	Significantly Increased
Signaling by MET	2	110	0.3	P63001;P11276	Significantly Increased
ABC-family proteins mediated transport	2	112	0.3	Q01853;Q60692	Significantly Increased
Antigen processing-Cross presentation	2	113	0.3	Q61830;Q60692	Significantly Increased
Regulation of Complement cascade	2	129	0.3	P01027;P06909	Significantly Increased
Cargo recognition for clathrin-mediated endocytosis	2	134	0.3	Q921I1;Q07113	Significantly Increased
Beta-catenin independent WNT signaling	2	136	0.3	P63001;Q60692	Significantly Increased
Complement cascade	2	140	0.3	P01027;P06909	Significantly Increased
Signaling by Hedgehog	2	161	0.3	Q01853;Q60692	Significantly Increased
Clathrin-mediated endocytosis	2	187	0.312	Q921I1;Q07113	Significantly Increased

Table 2.3. Enriched biological pathways of proteins bearing cysteine sites with significantly increased/decreased modification abundance.

Reactome identified 17 and 55 pathways that are enriched by proteins with significantly increased and decreased reversible cysteine sites, respectively (Figure 2.2.D, top panel). Specifically, proteins with significantly decreased reversibly oxidized sites are significantly enriched in 3 extracellular matrix organization and signaling pathways, 2 ion homeostasis and transportation pathways, and 2 TCA cycle and respiratory electron transport (TCA) pathway. Proteins with significantly increased reversibly oxidized cysteine sites are significantly enriched in 7 transcriptional and translational regulation pathways, 2 protein post-translational modification (PTM) pathways, 2 glucose metabolism pathways, and 3 TCA cycle and respiratory electron transport pathways. Notably, the "TCA cycle and respiratory electron transport" pathway is enriched in both significantly increased and decreased sites with associated proteins labeled by their UniProt ID. This observation reveals a significant and diverse reversible O-PTM regulation of mitochondrial metabolism during ISO-induced cardiac hypertrophy. In addition, significantly increased cysteine modifications of proteins susceptible to methylation and phosphorylation indicates further hierarchical interactions among PTMs. Zinc finger factor, ZEP1 is significantly decreased in CysSO3H (associated proteins are labeled by their UniProt ID), demonstrating redox regulation of transcriptional regulation during hypertrophy. Detailed information on featured sites, proteins, and their enriched pathways are listed in Tables 2.2 and 2.3, respectively.


A
Day1
591 proteins
1,207 sites
Day3
532 proteins
1,082 sites
Day5
628 proteins
1,288 sites
Day7
572 proteins
1,154 sites
Day10
419 proteins
818 sites
Day14
468 proteins
910 sites
Site count
ISO
Vehicle
Abundance (e+7; XIC)
B
Reversible oxidation
Total cysteine
CysSO2H
CysSO3H
377
850
512
201
81
81
238
44
121
1057
3494
1021
427
33
46
600
5
33
C
Reversible cysteine oxidation
CysSO2H
CysSO3H
Log2 (Fold-change)
Days 1 3 5 7 10 14

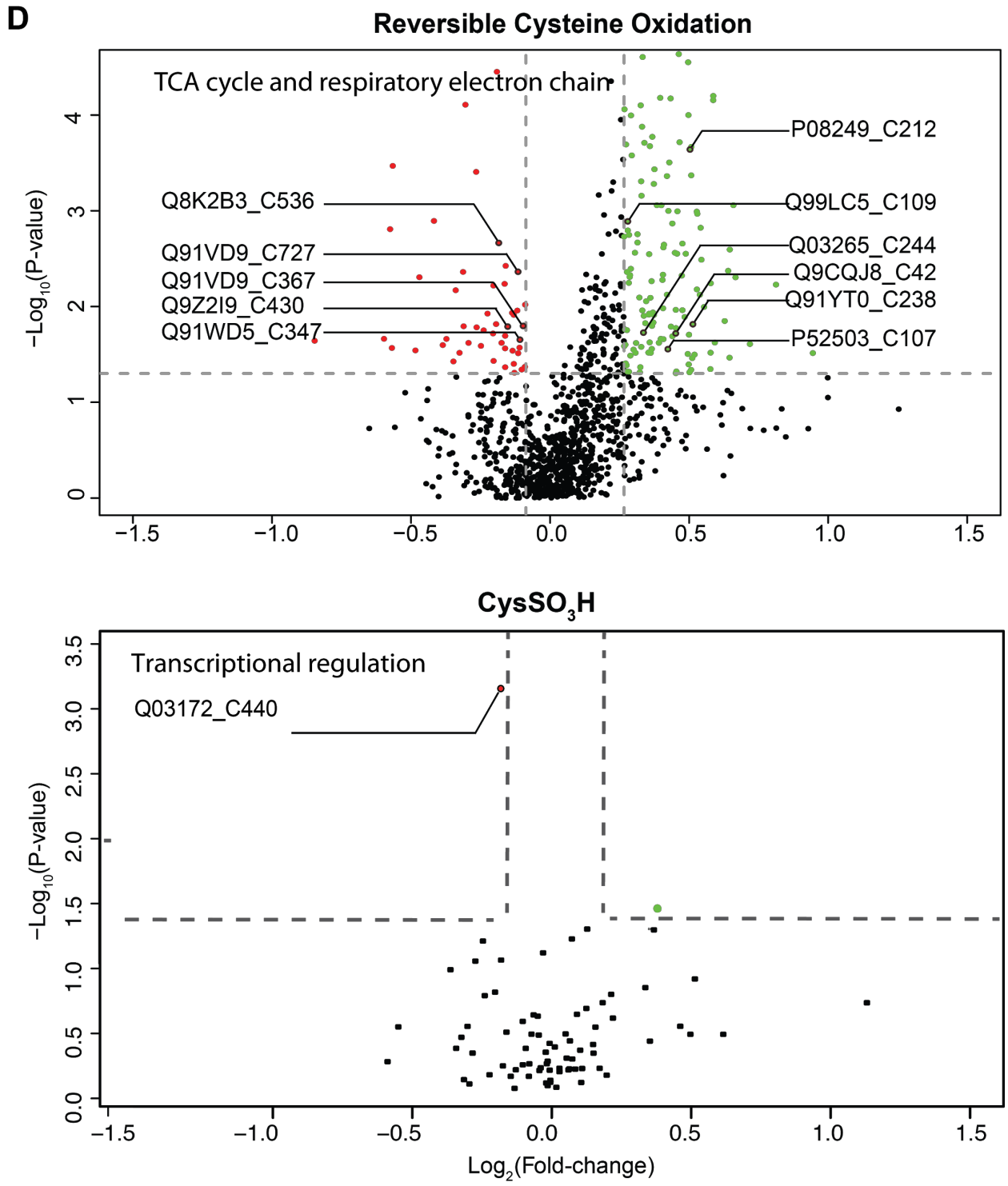


Figure 2.2. Impact of ISO-induced cardiac hypertrophy on the reversible cysteine O-PTM profile. (A) Abundance distribution of reversible cysteine O-PTM during ISO treatment. Six histograms exhibit the distribution of abundance of reversible cysteine O-PTM in each of the 6 time-points under both ISO (red) and Vehicle (blue) conditions. The absolute value of the vertical axis represents the site counts whereas the horizontal axis represents average abundance value (XIC) across all replicates. The black

arrow denotes a significant change in the abundance distribution of ISO occurring on Day5, where a sizeable upshift is observed in the ISO distribution of reversible cysteine O-PTM.

(B) Site and protein counts of reversible cysteine O-PTM, $CysSO_2H$, and $CysSO_3H$ during ISO-induced hypertrophy. A Venn diagram illustrates the number of reversibly oxidized proteins (left panel) and cysteine sites (right panel) for each of the cysteine O-PTMs that are identified from the integrated discovery platform. Data show that there are considerable and complex patterns of overlap among proteins and sites that are susceptible to modification.

(C) Distribution of cysteine O-PTM abundance during ISO treatment. Three violin plots exhibit the log_2 transformed abundance ratio of oxidation (y-axis) in ISO vs. Vehicle over 6 time points across 14 days (x-axis). The red horizontal line depicts the ratio value of zero, demonstrating no change. The black bar on the violin at each time point represents the population mean. The black arrow denotes a significant change in the abundance distribution of ISO occurring on Day5, where a sizeable upshift is observed in the ISO distribution of reversible cysteine O-PTM.

(D) Statistical validation of changes in oxidative molecular signatures. A volcano plot portrays significantly altered molecular signatures of reversible cysteine O-PTM (top panel) and $CysSO_3H$ (bottom panel) during ISO treatment. Each dot represents a cysteine O-PTM site with respect to its log2-transformed fold-change of abundance ratio in ISO/Vehicle (x-axis) and $-log_{10}$-transformed p-value from t-test (y-axis). Any sites with p-value <0.05 (horizontal line) and fold-change within the top or bottom 20th percentile (vertical line) are identified as significantly altered molecular signatures. Significantly increased and decreased modification sites are labeled in green and red, respectively. Pathway enrichment by Reactome identified "TCA cycle and respiratory electron chain" as the significantly enriched pathway for signatures with reversible O-PTM. "Transcriptional and translational regulation" is enriched in $CysSO_3H$. The UniProt IDs of proteins housing decreased and increased sites of cysteine O-PTM in these two enriched pathways are labeled. Detailed information regarding proteins of each cluster and their enriched pathways are listed in the Tables 2.2 and 2.3, respectively. Abbreviations: oxidative post-translational modification (O-PTM); isoproterenol (ISO); cystein sulfinylation ($CysSO_2H$); cystein sulfonylation ($CysSO_3H$); extracted ion chromatogram (XIC); TCA cycle and respiratory electron transport (TCA).

2.IV.C. Integrated temporal profiles of cysteine O-PTM in the hypertrophic heart proteome.

To characterize the dynamic patterns of reversible cysteine O-PTM during the progression of cardiac hypertrophy, we calculated the abundance ratios of each reversible cysteine O-PTM site in ISO over Vehicle for 1, 3, 5, 7, 10, and 14 days and employed a cubic spline-based smoothing approach for data de-noising. The cubic spline-fitted data were subsequently classified by K-means clustering to produce a total of 8 unique dynamic patterns (Figure 2.3.A). The largest cluster (Cluster 6, n = 198) remains unchanged during ISO treatment. These likely represent structural disulfide cysteines of housekeeping proteins. Consistent with this notion, Reactome analysis identified extracellular matrix (ECM) organization and signaling pathways to be enriched in this cluster, and ECM proteins are known to be rich in structural disulfide bonds.

The dynamic patterns of 8 clusters were represented in the stripcharts with smoothing lines, demonstrating the temporal feature of each cluster (Figure 2.3.B). Different temporal patterns of reversible O-PTM clusters may reveal the hidden association between dynamic regulation of biological processes and the progression of cardiac hypertrophy. We applied determination coefficient analysis to extract key sites as temporal signatures that are highly correlated (R^2 >0.8) with the temporal feature of their corresponding clusters. Cluster 6, with its lack of significant temporal changes to cysteine sites, is excluded from temporal signatures analysis. Among a total of 415 sites within 250 proteins that match the criteria as temporal signatures, the majority (404 sites within 243 proteins) are modified by reversible cysteine O-PTM, whereas 5 sites within 4 proteins by $CysSO_2H$ and 6 sites within 5 proteins by $CysSO_3H$. Notably, majority of $CysSO_2H$ sites (3 out of 5) are distributed in Cluster 1, exhibiting a continual decrease whereas the majority of $CysSO_3H$ sites (5 out of 6) are enriched in Cluster 7, exhibiting a continual increase. Notably, $CysSO_2H$ can be enzymatically reduced to free cysteine and can also be further oxidized into $CysSO_3H$, which is non-reducible (Lowther and Haynes 2011; Woo et al. 2005). These observations revealed distinct regulation pattern of different irreversible modification types during ISO-induced oxidative stress. Specifically, $CysSO_2H$ on Cys164 and $CysSO_3H$ on Cys167 in ATP7b (Q64446) shared the temporal pattern of continual increase (Cluster 7), suggesting an inactivation of ATP7b during cardiac hypertrophy. As a malfunction in ATP7b leads to cellular copper accumulation and cardiomyopathy in Wilson disease (Patil et al. 2013), the mapping of ATP7b's irreversible O-PTM temporal pattern offers mechanistic insights into cardiac complications of Wilson disease.

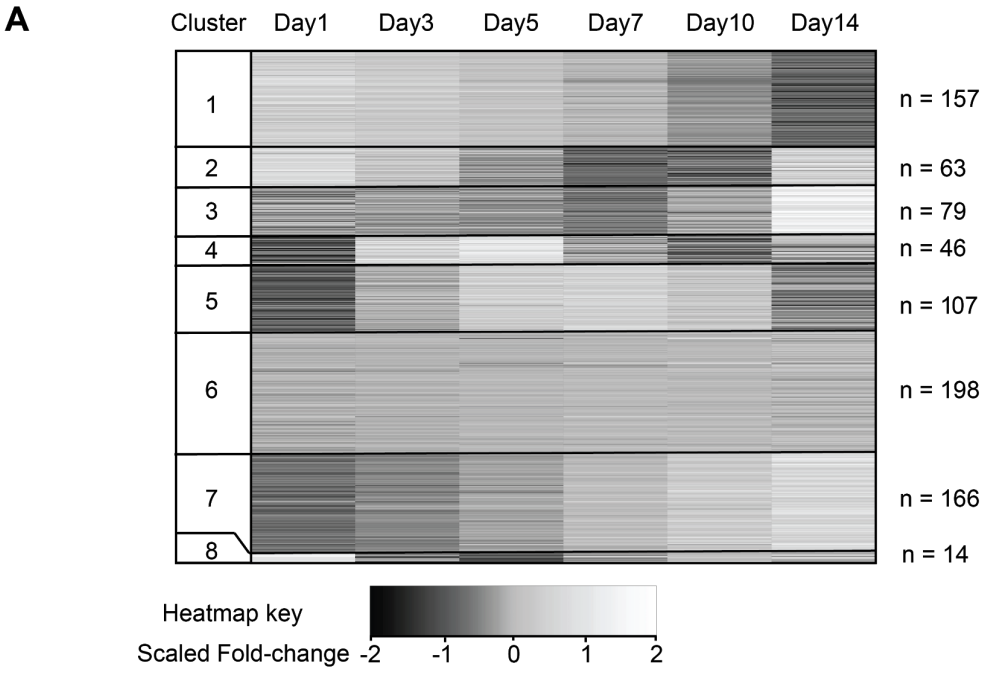
In a total of 243 reversibly oxidized proteins, 62 out of 90 bore at least two sites that fall into more than one temporal cluster. The distinct temporal regulation of cysteine sites within the same protein is possibly due to the unique microenvironment of the particular cysteine sites (e.g. cysteine pKa, enzymatic interaction). We utilized Reactome to perform pathway

enrichment with temporal signature proteins that are unique to one cluster. Cluster 1 is characterized by a continual decrease and significantly enriched in branched-chain amino acid (BCAA) catabolism, fatty acid beta-oxidation (FABO), and TCA pathways; Cluster 2 by an initial decrease that dips on Day 7 and significantly enriched in BCAA and TCA pathways; Cluster 3 by a sharp, increase after Day 7 and significantly enriched in BCAA, TCA, and glucose metabolism; Cluster 4 by a sinusoidal curve and significantly enriched in protein phosphorylation and complement cascade; Cluster 5 by an arched curve that peaks on Day 7 and significantly enriched in protein phosphorylation and creatine metabolism; Cluster 6 by no change and significantly enriched in extracellular matrix organization and signaling; and Cluster 7 by a continual increase and significantly enriched in BCAA, FABO, TCA, glucose metabolism, and creatine metabolism; and Cluster 8 by an initial sharp decrease before returning to baseline levels. Reversible cysteine O-PTM sites/proteins with significant alteration in abundance ratio across 14 days after ISO were widely distributed among 8 clusters, suggesting that dynamic temporal patterns is a state-of-the-art measurement and prediction that is independent of the absolute value of end-point measurements. Specifically, three 40S ribosomal proteins — S3, S6, and S11 — with significantly increased oxidation abundance are enriched in Cluster 1 (continual decrease), suggesting an initial wave of sharply increased oxidation in response to ISO treatment that diminishes as the heart hypertrophies. Notably, the majority of glucose metabolism proteins (4 out of 5) with significantly increased abundance belong to Cluster 7, demonstrating a strong and significant increase of reversible cysteine O-PTM in glucose metabolism during cardiac hypertrophy. In addition, the conserved Cys17 in the ATPase domain of the cytosolic chaperone protein, Hspa8, was continually oxidized (Cluster 7). As the perpetual oxidation of the conserved cysteine in the ER chaperone, Hspa5, has been shown to enhance chaperone function and ER homeostasis, the perpetual oxidation of Hspa8 indicates a potential function in cytosolic protein homeostasis during ISO-induced cardiac hypertrophy (J. Wang and Sevier 2016; J. Wang et al. 2014; O'Donnell et al. 2018).

The stoichiometry is an important feature of PTM and can be calculated by comparing the abundance of a modified form with all forms of a site of interest. Integrating the temporal patterns of total cysteine abundance and reversible cysteine O-PTM further reveals the comprehensive regulation of cysteine O-PTM stoichiometry. We applied cubic spline based clustering on total cysteine abundance and subsequently co-clustered temporal patterns of reversible cysteine O-PTM with total cysteine abundance (Figure 2.4.). Briefly, cubic spline-based K-mean clustering was implemented to determine the temporal clusters of the total cysteine proteome. Subsequently, temporal signature sites of total cysteines from each cluster were extracted by determination coefficient analysis with R^2 >0.8 as described and combined with the temporal pattern of oxidized cysteine sites. Clusters with more than 3 sites were considered as conserved patterns and included in the subsequent analysis.

Figure 2.4. shows a total of 10 co-clusters with more than 3 sites identified. These co-clusters consist of 3 temporal patterns from reversible and irreversible cysteine modification identified previously and 4 newly identified temporal patterns from total cysteine abundance. Cluster 1A is characterized by continual decrease in oxidation with continual increase in total abundance, suggesting a decrease in oxidation occupancy; Cluster 1B by continual decreased oxidation with no change in total abundance and enriched in fatty acid metabolism; Cluster 1C by continual decrease in both oxidation and total abundance and enriched in branched chain amino acid catabolism; Cluster 1D by continual decreased oxidation with initial decreased abundance that dips on Day 7 and enriched in respiratory electron transport; Cluster 7A by both continual increase in oxidation and total abundance and enriched in glucose metabolism, FABO, and lipoprotein remodeling; Cluster 7B by continual increased oxidation with no change in total abundance and enriched in glucose metabolism and ATP synthesis; Cluster 7C by continual increased oxidation with decreased total abundance and enriched in retinoic acid signaling, suggesting a continual increase in cysteine O-PTM occupancy in regulating retinoic acid, which

contributes to cardiac development; Cluster 7E by continual increase oxidation with arched curve that peaks on Day 7; Cluster 3A by initial decreased oxidation that dips on Day7 with continual increased total abundance and enriched in pyruvate metabolism and TCA cycle; Cluster 5B by an arched curve that peaks on Day 7 with no change in total abundance. In addition to sites modified by reversible O-PTM, two $CysSO_3H$ sites bore the temporal pattern of Cluster 7A. Notably, three major enriched pathways from Cluster1, BCAA, FABO, and TCA, were divided into Cluster 1B, 1C, and 1D, according to the temporal patterns of their total cysteine abundance, demonstrating the clustering of their temporal pattern is dependent on their biological function. Among clusters with altered stoichiometry during cardiac hypertrophy, Cluster 1B and Cluster 7B are enriched in FABO and glucose metabolism, respectively. As cardiac hypertrophy is featured by an increased reliance on glucose with an overall reduction in fatty acid metabolism (Kolwicz and Tian 2011), these observations provide a detailed molecular map that reveals an additional layer of regulation in mitochondrial energy metabolism via reversible cysteine O-PTM.



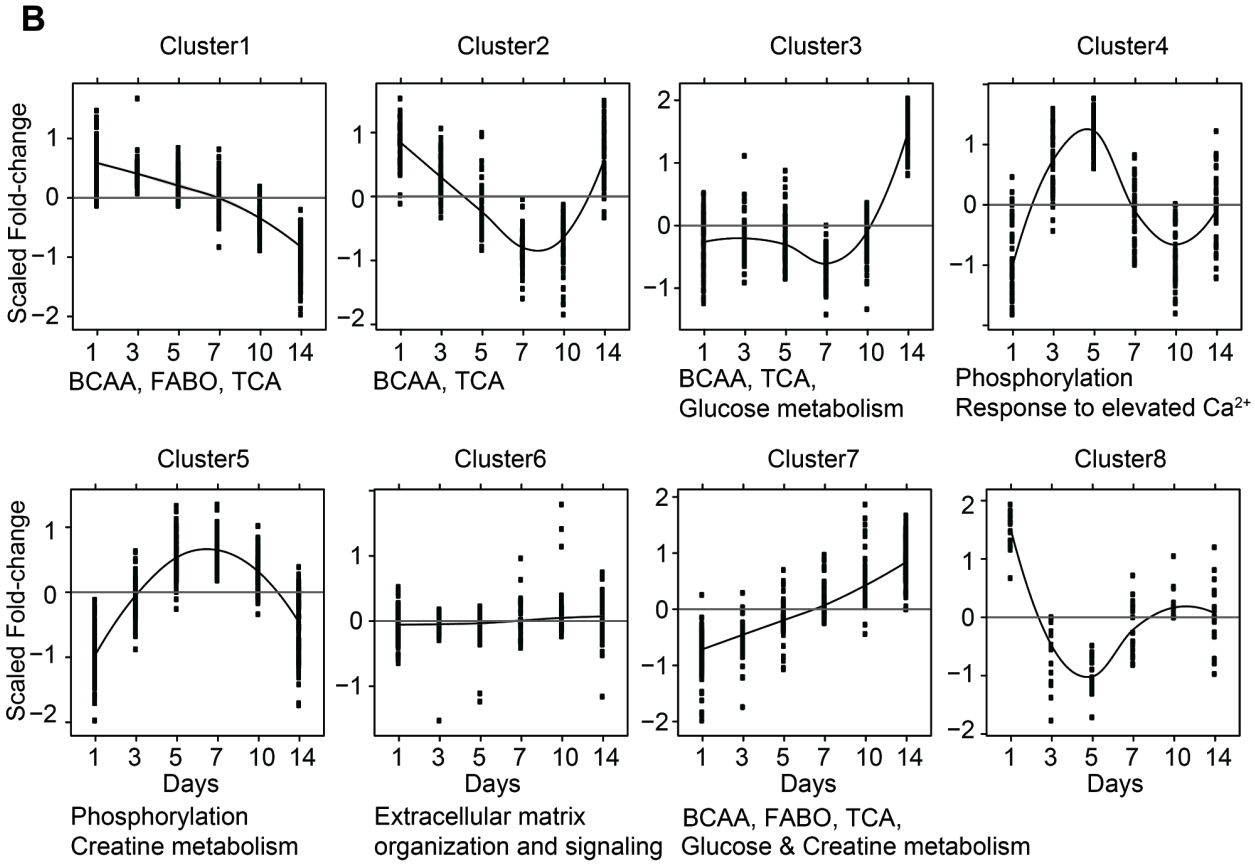


Figure 2.3. Temporal profiling of reversible cysteine O-PTM proteomes using cubic spline-based clustering. (A) The temporal mapping of the reversible cysteine O-PTM proteome. Cubic spline followed by K-mean clustering was applied to abundance ratio of reversible cysteine O-PTM sites in ISO vs. Vehicle across 6 time points (x-axis), as represented by a scale from 2-fold decrease (black) to 2-fold increase (white), in a heat map. Our analyses yielded 8 unique temporal patterns of change in the reversible cysteine O-PTM abundance ratio. The numbers of sites of each cluster are as listed to the right of the heatmap. **(B) Cluster-specific temporal patterns of reversible cysteine O-PTM.** Eight stripcharts exhibit the trends in the normalized abundance ratio of reversible cysteine O-PTM sites (ISO vs. Vehicle; y-axis) over 6 time points (x-axis). Cluster 1 of reversible cysteine O-PTM sites is characterized by continual decrease; Cluster 2 by an initial decrease that dips on Day 7; Cluster 3 by a sharp increase after Day 7; Cluster4 by a sinusoidal curve with an initial increase that shifts to decrease on Day 5 and back to increase on Day 10; Cluster 5 by an arched curve that peaks on Day 7; Cluster 6 by no change; Cluster 7 by continual increase; and Cluster 8 by an initial sharp decrease that dips on Day 5. Key cysteine sites and pathways are listed on the stripcharts of their corresponding clusters. Abbreviations: oxidative post-translational modification (O-PTM); isoproterenol (ISO); TCA cycle and respiratory electron transport (TCA); branched-chain amino acid (BCAA) catabolism; fatty acid beta-oxidation (FABO).

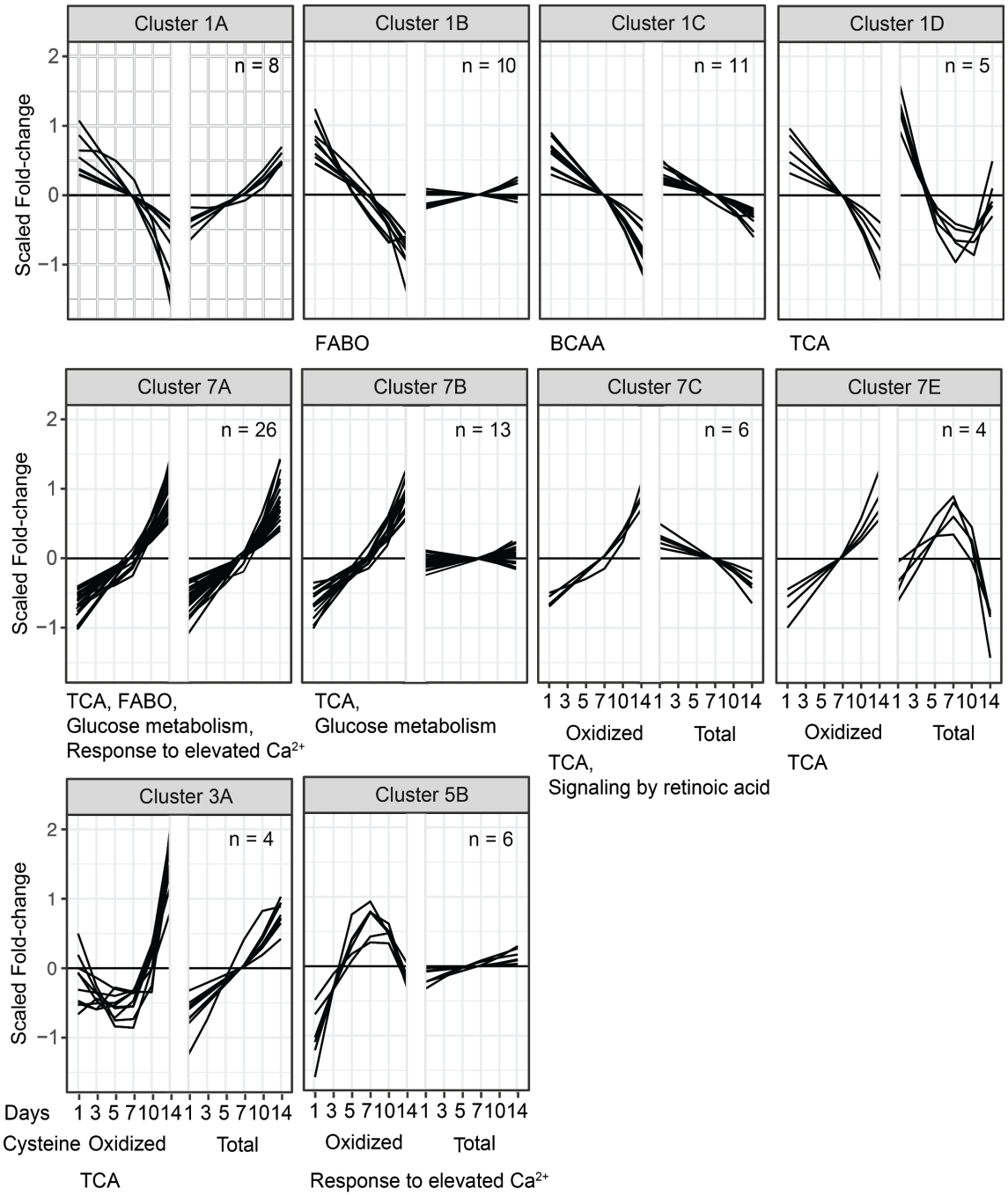


Figure 2.4. Temporal profiling combining reversible cysteine O-PTM and total cysteine proteomes using cubic spline-based co-clustering. Cubic spline followed by K-mean clustering was applied to abundance ratio of total cysteine sites (ISO vs. Vehicle) in the proteome. The profiles of reversible cysteine O-PTM temporal signatures (Figure 2.3.: R^2 >0.8) were co-plotted with the well-fitted temporal profiles of their total cysteine abundance (labeled as A-E; R^2 >0.8). Ten co-clusters with more than three cysteine sites were identified and demonstrated in the stripcharts, exhibiting the trends in the normalized abundance ratio of reversible cysteine O-PTM and total cysteine sites (y-axis) over 6 time points (x-axis). As previously described, Cluster 1 of reversible cysteine O-PTM sites is characterized by continual decrease; Cluster 3 by a sharp increase after Day7; Cluster 5 by an arched curve that peaks on Day 7; Cluster 6 by no change; and Cluster 7 by continual increase. Cluster A of total cysteine sites is characterized by continual increase; Cluster B by no change; Cluster C by continual decrease; Cluster D

by initial sharp decrease that dips between Day 7-10; Cluster E by an arched curve that peaks on Day7. Key pathways are listed on the stripcharts of their corresponding clusters. Abbreviations: oxidative post-translational modification (O-PTM); isoproterenol (ISO); TCA cycle and respiratory electron transport (TCA); branched-chain amino acid (BCAA) catabolism; fatty acid beta-oxidation (FABO).

2.IV.D. Correlation between hypertrophy phenotype and cysteine O-PTM abundance.

Temporal cluster analysis revealed the intrinsic temporal regulation of a cysteine O-PTM site, which may contribute to or respond to the dynamic pattern of the cardiac hypertrophy phenotype. The heart weight-body weight (HW/BW) ratios of the mouse strain were documented along with the left ventriclar sample collections to provide temporal phenotypic profiles. During ISO treatment, all mouse strains exhibited cardiac hypertrophy development starting from Day 1 post ISO-treatment (Figure 2.5.A). The averaged ratio of HW/BW across 6 time points increased 20% in ISO over Vehicle. Interestingly, we observed an arched curve of hypertrophy development that peaked on Day 5 (33% higher than control) and stabilized on Day 10 (~14% higher than control). This is consistent with previous observations of the ISO mouse model in the C57BL/6J40 mouse strain (Rau et al. 2015).

Protein	Cysteine	Correlation	P value	Oxidation Type	Signature	Entry Name	Protein Names	Gene Names
O55143	O55143_C875	-0.99	0.0001	Reversible cysteine O-PTM	Negative Correlation	AT2A2_MOUSE	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (SERCA2) (SR Ca(2+)-ATPase 2) (EC 3.6.3.8) (Calcium pump 2)	Atp2a2
O55143	O55143_C377	-0.96	0.0021	Reversible cysteine O-PTM	Negative Correlation	AT2A2_MOUSE	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (SERCA2) (SR Ca(2+)-ATPase 2) (EC 3.6.3.8) (Calcium pump 2) (Calcium-transporting ATPase sarcoplasmic reticulum type, slow twitch skeletal muscle isoform) (Endoplasmic reticulum class 1/2 Ca(2+) ATPase)	Atp2a2
P00397	P00397_C498	-0.96	0.0020	Reversible cysteine O-PTM	Negative Correlation	COX1_MOUSE	Cytochrome c oxidase subunit 1 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I)	Mtco1 COI mt-Co1
Q07417	Q07417_C246	-0.97	0.0010	Reversible cysteine O-PTM	Negative Correlation	ACADS_MOUSE	Short-chain specific acyl-CoA dehydrogenase, mitochondrial (SCAD) (EC 1.3.8.1) (Butyryl-CoA dehydrogenase)	Acads
Q3ULD5	Q3ULD5_C431	-0.99	0.0001	Reversible cysteine O-PTM	Negative Correlation	MCCB_MOUSE	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial (MCCase subunit beta) (EC 6.4.1.4) (3-methylcrotonyl-CoA carboxylase 2)	Mccc2

Protein	Cysteine	Correlation	P value	Oxidation Type	Signature	Entry Name	Protein Names	Gene Names
Q8BFR5	Q8BFR5_C290	-0.93	0.0065	Reversible cysteine O-PTM	Negative Correlation	EFTU_MOUSE	Elongation factor Tu, mitochondrial	Tufm
Q8BGH2	Q8BGH2_C65	-0.93	0.0079	Reversible cysteine O-PTM	Negative Correlation	SAM50_MOUSE	Sorting and assembly machinery component 50 homolog	Samm50
Q8K182	Q8K182_C140	-0.92	0.0085	Reversible cysteine O-PTM	Negative Correlation	CO8A_MOUSE	Complement component C8 alpha chain (Complement component 8 subunit alpha)	C8a
Q8K2B3	Q8K2B3_C238	-0.99	0.0002	Reversible cysteine O-PTM	Negative Correlation	SDHA_MOUSE	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (EC 1.3.5.1) (Flavoprotein subunit of complex II) (Fp)	Sdha
Q9D8B4	Q9D8B4_C115	-0.98	0.0006	Reversible cysteine O-PTM	Negative Correlation	NDUAB_MOUSE	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 11 (Complex I-B14.7) (CI-B14.7) (NADH-ubiquinone oxidoreductase subunit B14.7)	Ndufa11
Q5SSE9	Q5SSE9_C4865	-0.93	0.0070	SO2H	Negative Correlation	ABCAD_MOUSE	ATP-binding cassette sub-family A member 13	Abca13
A2ASQ1	A2ASQ1_C441	0.97	0.0012	Reversible cysteine O-PTM	Positive Correlation	AGRIN_MOUSE	Agrin [Cleaved into: Agrin N-terminal 110 kDa subunit; Agrin C-terminal 110 kDa subunit; Agrin C-terminal 90 kDa fragment (C90); Agrin C-terminal 22 kDa fragment (C22)]	Agrn Agrin
P01027	P01027_C727	0.95	0.0032	Reversible cysteine O-PTM	Positive Correlation	CO3_MOUSE	Complement C3 (HSE-MSF) [Cleaved into: Complement C3 beta chain; C3-beta-c (C3bc); Complement C3 alpha chain; C3a anaphylatoxin; Acylation stimulating protein (ASP) (C3adesArg); Complement C3b alpha' chain; Complement C3c alpha' chain fragment 1; Complement C3dg fragment; Complement C3g fragment; Complement C3d fragment; Complement C3f fragment; Complement C3c alpha' chain fragment 2]	C3
P07724	P07724_C302	0.94	0.0051	Reversible cysteine O-PTM	Positive Correlation	ALBU_MOUSE	Serum albumin	Alb Alb-1 Alb1
P07724	P07724_C485	0.99	0.0001	Reversible cysteine O-PTM	Positive Correlation	ALBU_MOUSE	Serum albumin	Alb Alb-1 Alb1
P14211	P14211_C137	0.93	0.0066	Reversible cysteine O-PTM	Positive Correlation	CALR_MOUSE	Calreticulin (CRP55) (Calregulin) (Endoplasmic reticulum resident protein 60) (ERp60) (HACBP)	Calr
P19221	P19221_C264	1.00	0.0000	Reversible cysteine O-PTM	Positive Correlation	THRB_MOUSE	Prothrombin (EC 3.4.21.5) (Coagulation factor II) [Cleaved into: Activation peptide fragment 1; Activation peptide fragment 2; Thrombin light chain; Thrombin heavy chain]	F2 Cf2

Protein	Cysteine	Correlation	P value	Oxidation Type	Signature	Entry Name	Protein Names	Gene Names
P62908	P62908_C119	1.00	0.0000	Reversible cysteine O-PTM	Positive Correlation	RS3_MOUSE	40S ribosomal protein S3 (EC 4.2.99.18)	Rps3
P68040	P68040_C249	0.97	0.0010	Reversible cysteine O-PTM	Positive Correlation	RACK1_MOUSE	Receptor of activated protein C kinase 1 (12-3) (Guanine nucleotide-binding protein subunit beta-2-like 1) (Receptor for activated C kinase) (Receptor of activated protein kinase C 1) (p205) [Cleaved into: Receptor of activated protein C kinase 1, N-terminally processed (Guanine nucleotide-binding protein subunit beta-2-like 1, N-terminally processed)]	Rack1 Gnb2-rs1 Gnb2l1
Q01853	Q01853_C572	0.93	0.0080	Reversible cysteine O-PTM	Positive Correlation	TERA_MOUSE	Transitional endoplasmic reticulum ATPase (TER ATPase) (EC 3.6.4.6)	Vcp
Q07113	Q07113_C886	1.00	0.0000	Reversible cysteine O-PTM	Positive Correlation	MPRI_MOUSE	Cation-independent mannose-6-phosphate receptor (CI Man-6-P receptor) (CI-MPR) (M6PR) (300 kDa mannose 6-phosphate receptor) (MPR 300) (Insulin-like growth factor 2 receptor)	Igf2r
Q3ULD5	Q3ULD5_C392	0.94	0.0047	Reversible cysteine O-PTM	Positive Correlation	MCCB_MOUSE	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial (MCCase subunit beta) (EC 6.4.1.4) (3-methylcrotonyl-CoA carboxylase 2)	Mccc2
Q60936	Q60936_C265	1.00	0.0000	Reversible cysteine O-PTM	Positive Correlation	COQ8A_MOUSE	Atypical kinase COQ8A, mitochondrial (EC 2.7.-.-) (Chaperone activity of bc1 complex-like) (Chaperone-ABC1-like) (Coenzyme Q protein 8A) (aarF domain-containing protein kinase 3)	Coq8a Adck3 Cabc1
Q61543	Q61543_C320	0.95	0.0040	Reversible cysteine O-PTM	Positive Correlation	GSLG1_MOUSE	Golgi apparatus protein 1 (E-selectin ligand 1) (ESL-1) (Selel) (Golgi sialoglycoprotein MG-160)	Glg1 Esl1 Mg160 Selel
Q61554	Q61554_C2413	0.92	0.0091	Reversible cysteine O-PTM	Positive Correlation	FBN1_MOUSE	Fibrillin-1 [Cleaved into: Asprosin]	Fbn1 Fbn-1
Q8VDN2	Q8VDN2_C249	1.00	0.0000	Reversible cysteine O-PTM	Positive Correlation	AT1A1_MOUSE	Sodium/potassium-transporting ATPase subunit alpha-1 (Na(+)/K(+) ATPase alpha-1 subunit) (EC 3.6.3.9) (Sodium pump subunit alpha-1)	Atp1a1
Q8VE96	Q8VE96_C254	0.99	0.0003	Reversible cysteine O-PTM	Positive Correlation	S35F6_MOUSE	Solute carrier family 35 member F6 (ANT2-binding protein) (ANT2BP) (Transport and Golgi organization 9 homolog)	Slc35f6
Q9DBH5	Q9DBH5_C204	0.95	0.0034	Reversible cysteine O-PTM	Positive Correlation	LMAN2_MOUSE	Vesicular integral-membrane protein VIP36 (Lectin mannose-binding 2) (Vesicular integral-membrane protein 36) (VIP36)	Lman2
Q9ESW4	Q9ESW4_C72	1.00	0.0000	Reversible cysteine O-PTM	Positive Correlation	AGK_MOUSE	Acylglycerol kinase, mitochondrial (EC 2.7.1.107) (EC 2.7.1.94) (Multiple substrate lipid kinase) (MuLK) (Multi-substrate lipid kinase)	Agk Mulk

Table 2.4. List of cysteine sites significantly correlated with the hypertrophy phenotype.

To directly map the association between temporal changes in cysteine O-PTM and that of the hypertrophy phenotype, we performed cubic spline-based Pearson correlation. We selected 29 cysteine O-PTM sites with a p-value <0.01 as key oxidation sites that significantly associated with cardiac hypertrophy. Among them, 18 cysteine sites within 17 proteins and 11 cysteine sites within 10 proteins were positively and negatively correlated with HW/BW changes, respectively (Figure 2.5.B). Pathway enrichment of proteins with a positive correlation to phenotype unveiled 4 secretory proteins in the extracellular space that regulate Insulin-like Growth Factor (IGF) transport, a key regulatory mechanism of cardiac hypertrophy (Hua, Zhang, and Ren 2012; Troncoso et al. 2014). In addition, a BCAA pathway protein, MCCC2, was identified in both correlation groups, demonstrating a diverse regulation of MCCC2 cysteine O-PTM. Notably, reversible O-PTM of an additional two key BCAA proteins, BCAT2 and DBT were observed in Cluster 1C, featured by continually decreased oxidation and total cysteine abundance. As a BCAA catabolic defect is a metabolic hallmark of the failing heart (Sun et al. 2016), diverse oxidative regulation of BCAA proteins can potentially contribute to hypertrophy by metabolic reprogramming. Detailed information on featured sites, proteins, and their enriched pathways are listed in Tables 2.4 and 2.5, respectively.

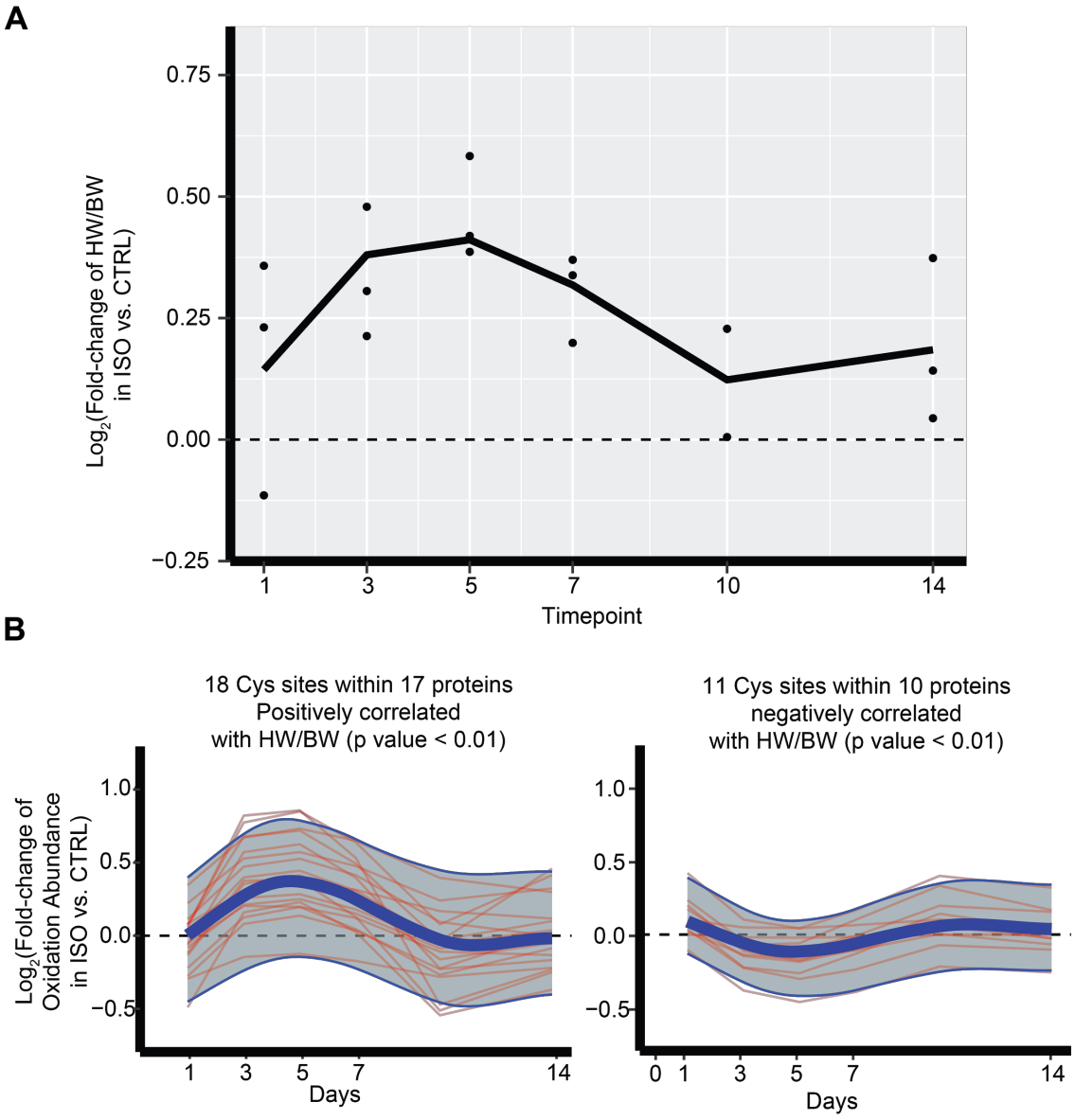


Figure 2.5. Phenotypic alteration and corresponding cysteine site fingerprints during ISO-induced cardiac hypertrophy. (A) The HW/BWs were temporally measured three times per time point during ISO-induced cardiac remodeling (time points Day 1, 3, 5, 7, 10, and 14), and subsequently normalized to HW/BWs of vehicle. The majority of $\log_2$ transformed measurements showed positive values, indicating HW/BWs of ISO mice increase relative to that of Vehicle mice, as expected. The difference of HW/BWs between ISO and Vehicle mice was continuously broadened and peaked at Day 5, and then decreased until Day 10, stabilizing thereafter. **(B)** Twenty-nine cysteine sites which are highly correlated with HW/BW changes were identified through Pearson correlation coefficient (p-value <0.01); their abundance ratios were displayed over time. The left panel shows eighteen cysteine sites within seventeen proteins that are positively correlated with HW/BW, while the right panel shows eleven cysteine sites within ten proteins that are negatively correlated with HWBW. The red line indicates each cysteine site and the thick blue line shows the average trend of the group. The two thin blue lines indicate upper bound and lower bound that cover 95% of the data. Detailed information regarding proteins of each cluster and their enriched pathways are listed in the **Tables 2.4 and 2.5**, respectively. Abbreviations: isoproterenol (ISO); heart weight-body weight ratio (HW/BW); control (CTRL).

Pathway Name	# Entities Found	# Entities Total	Entities FDR	Submitted Entities Found	Association
N-glycan trimming in the ER and Calnexin/Calreticulin cycle	2	40	0.063407806	Q01853;P14211	Positive Correlation
Post-translational protein phosphorylation	3	165	0.063407806	Q61554;P07724;P01027	Positive Correlation
Binding and Uptake of Ligands by Scavenger Receptors	2	108	0.090262572	P14211;P07724	Positive Correlation
Regulation of Complement cascade	2	129	0.090262572	P19221;P01027	Positive Correlation
Complement cascade	2	140	0.090262572	P19221;P01027	Positive Correlation
Peptide ligand-binding receptors	2	196	0.090262572	P19221;P01027	Positive Correlation

Table 2.5. Enriched biological pathways of proteins associated with cysteine sites significantly correlated with the hypertrophy phenotype.

2.IV.E. Signature pathways and proteins that contribute to cardiac hypertrophy.

Combining molecular signatures from differential analysis, cubic spline-based temporal clustering, and phenotypic correlation, we highlighted 7 biological processes that are reproducibly identified as signature pathways during cardiac hypertrophy, including TCA, FABO, BCAA, glucose metabolism, protein phosphorylation, creatine metabolism, and response to elevated Ca^{2+}. Specifically, among pathways enriched in proteins with significantly increased reversible modification, protein phosphorylation, creatine metabolism, and response to elevated Ca^{2+} pathways exhibited an increase in cysteine modification in early stages of hypertrophy, whereas glucose metabolism pathways were modified in the later stages (Figure 2.6.A). Comparatively, the cysteine oxidative profiles of TCA, FABO, and BCAA were more complicated, as metabolic reprogramming is a highly integrative process with complex compensatory mechanisms. We mapped all identified BCAA pathway proteins and their corresponding temporal dynamics. As expected, temporal trends of cysteine modification differ across pathways and even among cysteine sites within one protein. Interestingly, cysteine temporal profiles of valine catabolism sub-pathways exhibit a trend towards increase, whereas

that of leucine and isoleucine catabolism uniformly illustrates a decreased trend at early stage of hypertrophy. These observations demonstrate that cysteine O-PTM profiles are highly correlated with key biological processes during cardiac hypertrophy. Furthermore, these integrated molecular signatures of hypertrophy are novel and can be used to define phenotypes of metabolic reprogramming in new and informative ways.

A

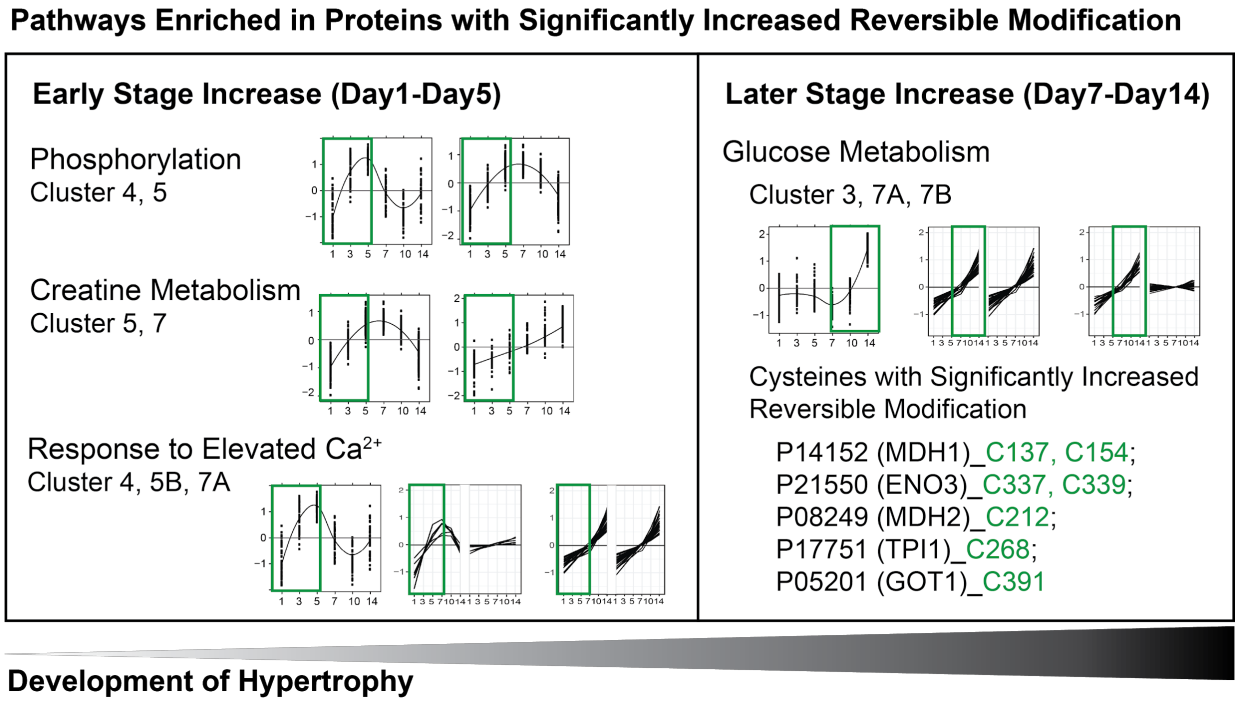
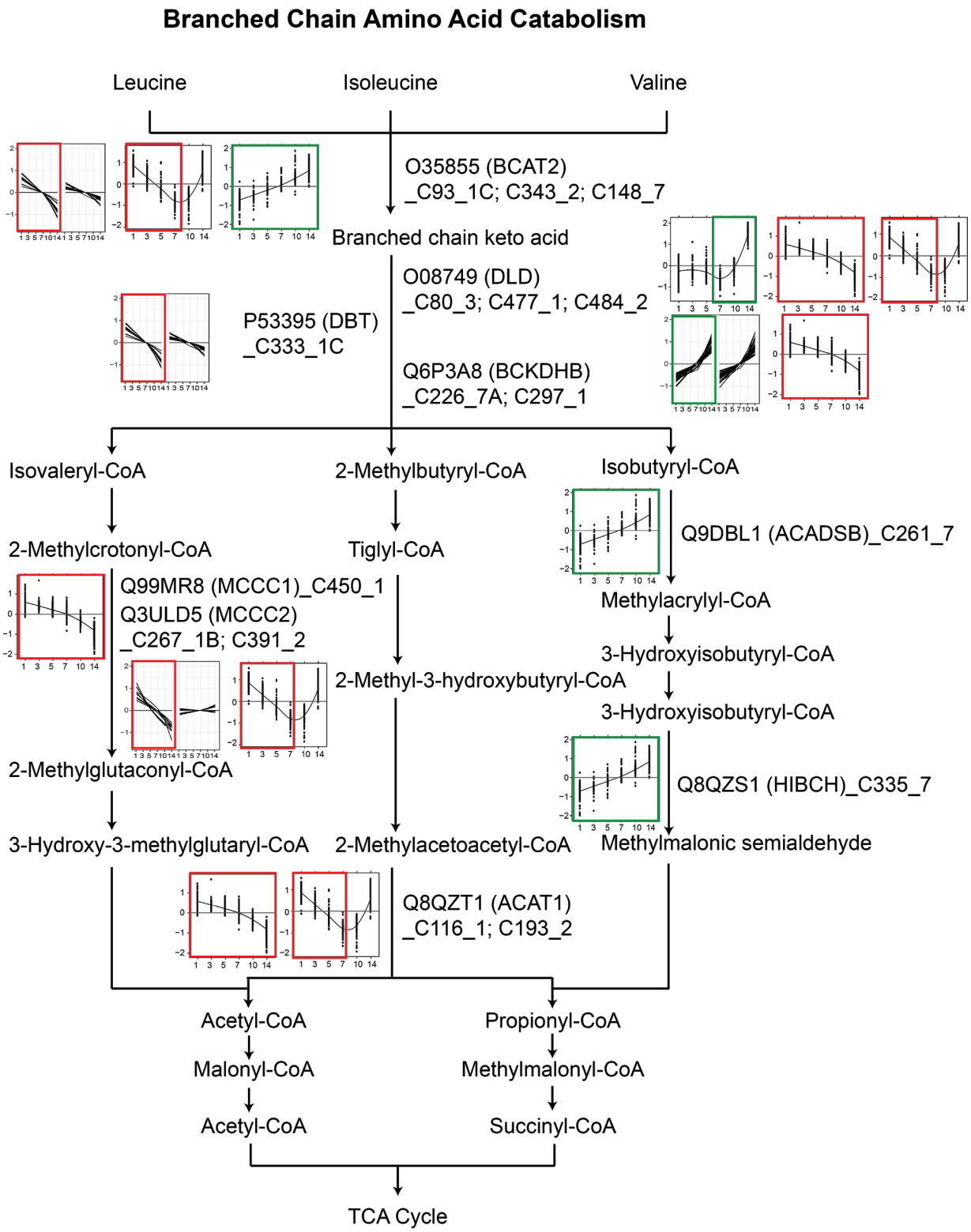


Figure 2.6. Signature pathways and proteins that contribute to cardiac hypertrophy. (A) The temporal changes of modified cysteine sites on pathways enriched in proteins with significantly increased modification abundance. Pathways and the temporal changes of their corresponding enriched temporal clusters are illustrated. Temporal clusters displaying increased trends of modification changes at indicated durations are boxed and color-coded as green. Protein phosphorylation, creatine metabolism, and response to elevated Ca2+ pathways are increased at early stages of hypertrophy whereas glucose metabolism has a trend of increased modification during later stages of hypertrophy. Cysteine sites with significantly increased modification in glucose metabolism pathways are labeled with accession number, protein name, and modified cysteine site. These cysteine sites with significantly increased modification are color-coded as green. **(B)** The temporal changes of modified cysteine sites on BCAA pathway proteins are labeled and illustrated. Each molecular player is labeled with accession number, protein name, modified cysteine site, and temporal cluster. Temporal clusters with increased or decreased trends at indicated durations are boxed and color-coded as green or red, respectively. Most sites in the shared pathways among BCAA pathways exhibit diverse temporal trends. Notably, the valine catabolism pathway exhibits an increased trend of cysteine modification whereas the temporal trend of leucine and isoleucine is decreased during early stages of hypertrophy. These observations suggest that oxidative cysteine regulation of BCAA pathways is highly compartmentalized, based on biological functions. Abbreviations: branched-chain amino acid catabolism (BCAA).

B


Branched Chain Amino Acid Catabolism
Leucine
Isoleucine
Valine
O35855 (BCAT2)
_C93_1C; C343_2; C148_7
Branched chain keto acid
O08749 (DLD)
_C80_3; C477_1; C484_2
P53395 (DBT)
_C333_1C
Q6P3A8 (BCKDHB)
_C226_7A; C297_1
Isovaleryl-CoA
2-Methylbutyryl-CoA
Isobutyryl-CoA
Q9DBL1 (ACADSB)_C261_7
2-Methylcrotonyl-CoA
Tiglyl-CoA
Methylacrylyl-CoA
Q99MR8 (MCCC1)_C450_1
Q3ULD5 (MCCC2)
_C267_1B; C391_2
3-Hydroxyisobutyryl-CoA
2-Methyl-3-hydroxybutyryl-CoA
3-Hydroxyisobutyryl-CoA
2-Methylglutaconyl-CoA
Q8QZS1 (HIBCH)_C335_7
3-Hydroxy-3-methylglutaryl-CoA
2-Methylacetoacetyl-CoA
Methylmalonic semialdehyde
Q8QZT1 (ACAT1)
_C116_1; C193_2
Acetyl-CoA
Propionyl-CoA
Malonyl-CoA
Methylmalonyl-CoA
Acetyl-CoA
Succinyl-CoA
TCA Cycle

2.IV.F. Role of cysteine O-PTM in cardiovascular biology.

ISO induced beta-adrenoceptor stimulation has been shown to evoke cardiac oxidative stress (G.-X. Zhang et al. 2005, 2007). Specifically, this process is mediated through increased mitochondrial ROS production, as mitochondrial targeted antioxidants diminish oxidative stress and its downstream biological events (Andersson et al. 2011). Meanwhile, the capacity of nitric oxide synthesis is also increased in response ISO, likely from enhanced expression of endothelial and inducible nitric oxide synthases (Krenek et al. 2009). During heart failure development, elevated RO/NS production in turn leads to modification of reactive cysteine thiol groups. The highly reactive thiol group of cysteine residues has made it difficult to isolate native versus non-biological oxidized peptides in the cardiovascular system, and subsequently, to elucidate their biological role. Pioneering studies on both technological and biological fronts have overcome this barrier and broken new ground in cardiovascular medicine. Recent work from Jennifer van Eyk's laboratory identified distinct subpopulations of nitrosylated cysteines through a dual-labeling of nitrosylation that reduces labeling bias (Chung et al. 2015). This advanced methodology was then utilized to explore glycogen synthase kinase 3β (GSK3β) regulation by nitrosylation and it was discovered that nitrosylation reduces GSK3β kinase activity and promotes its nuclear translocation (S.-B. Wang et al. 2018).

The biological and physiological roles of cysteine O-PTM in both vascular systems and cardiac muscle have been pioneered and carried into clinical translation by the work of Jonathan Stamler and Elizabeth Murphy, respectively. A seminal paper in 2004 by the Stamler laboratory was the first demonstration of a role of nitrosylation in innate immunity and vascular function (L. Liu et al. 2004). Mice devoid of S-nitrosoglutathione reductase showed marked increased in nitrosylation, vascular damage, and mortality following endotoxic challenge. More recent work unveiled a role for the hemoglobin β Cys93 residue in nitrosylation-based vasoactivity, demonstrating that S-nitrosohemoglobin plays a role in cardioprotection (R. Zhang et al. 2016).

This and other notable contributions (Stomberski, Hess, and Stamler 2019; Seth and Stamler 2015; Gonzalez et al. 2009) have revolutionized our understanding of cysteine O-PTM in vascular signaling. Studies from Elizabeth Murphy's lab have advanced knowledge of cysteine O-PTM in cardiac muscle biology. This group pioneered novel labeling approaches for measuring nitrosylation occupancy, the fraction of a given protein that is nitrosylation modified in the myocardium, and demonstrated that nitrosylation occupancy levels following ischemic preconditioning protect against cysteine O-PTM (Kohr et al. 2012). More recent, formative work by this lab demonstrated increased nitrosylation abundance at Cys144 of the cardioprotective protein, TRIM72, is a molecular switch preventing TRIM72 degradation following an oxidative insult, which increases cardiomyocyte survival (Kohr et al. 2014). Other work unveiled complex profiles of nitroso-redox signaling and nitrosylation of cardiac proteins in failing versus non-failing human cardiac tissue. Sex-specific differences in S-glutathionylation of endothelial nitric oxide synthase were discovered, adding to the overall complexity of these pathways in cardiac muscle (Menazza et al. 2015).

Our measurements represent the sum of all types of cysteine O-PTM events, including disulfide bonds, nitrosylation, glutathionylation, and irreversible cysteine O-PTMs. Among these, different modifications result from diverse regulatory mechanisms (Murray and Van Eyk 2012) and can be interconvertible. For example, glutathionylation of Cys63 on the ER stress protein BiP is mediated by sulfenylation (J. Wang and Sevier 2016). Meanwhile, both sulfenylation and glutathionylation of BiP have been shown to enhance BiP's activity in preventing aggregation. Meanwhile, different modifications can also lead to distinct biological consequences. For example, two types of oxidative posttranslational modifications have been shown to occur on Cys674 of SERCA2: reversible S-glutathiolation increases SERCA activity, whereas irreversible oxidative CysSO3H is associated with decreased activity (Qin et al. 2013). Conventionally, irreversible modifications were considered to be markers of cellular damage. Yet recent

advances suggest that these modifications have regulatory capacities. For example, the unique active cysteine sites of nitrile hydratase and thiocyanate hydrolase are responsible for metal coordination and can be modified by CysSO2H. The fully reduced forms of these two enzymes appear inactive, suggesting that CysSO2H is critical in maintaining their catalytic activity (Murakami et al. 2000; Arakawa et al. 2009). Our study provides key signatures modified by reversible and irreversible cysteine O-PTMs, facilitating the target prioritization for studying hierarchical regulation among different types of cysteine O-PTMs.

2.V. Discussion.

2.V.A. Technical Considerations.

Capitalizing on several advancements and variations of the biotin switch method (García-Santamarina et al. 2014; R. Li, Huang, and Kast 2015), we developed a quantitative approach for in-depth characterization of the cysteine oxidized cardiac proteome. Firstly, NEM alkylates free cysteines via a faster, more specific Michael addition reaction than the nucleophilic substitution reaction with IAM (Kramer et al. 2015; Reisz et al. 2013). Efficient labeling with less pH dependence makes NEM a great choice to preserve the labile reversible cysteine O-PTM and minimize artificial chemical alteration and non-specific labeling during sample processing over IAM. Secondly, biotin maleimide labeling offers the efficiency and specificity accompanying the Michael addition reaction. In addition, as the avidin-biotin interaction is strong and rapid (Haugland and You 2008), enrichment with high-capacity NeutrAvidin agarose enables a large-scale pull-down with high efficiency. Thirdly, stable-isotope dimethyl labeling using reductive amination is a reliable widely-used approach in MS-based quantitative proteomics (Boersema et al. 2009; Hsu et al. 2003). Specifically, dimethylation enhances fragmentation efficiency of collision-induced dissociation (CID) by increasing the number of positive charges on a peptide

(Fu and Li 2005). In addition, high cost-effectiveness makes dimethylation applicable for studies with any sample size.

Nevertheless, there are a few limitations to consider. The multi-stage process, including biotin maleimide labeling, dimethyl labeling, and avidin enrichment requires milligrams of starting material. The development of biotin maleimide-based isotope reagents would be a great tool to combine this two-stage labeling. While DTT is a potent and efficient reducing reagent for all types of reversible cysteine O-PTM, proteins with a large number of stable disulfide bonds will be prevalently enriched in this assay. On the other hand, this method can facilitate discovery of functional and catalytic disulfide bonds and less reactive cysteine O-PTMs that require a more potent reductant. Despite the limitations mentioned above, our biotin switch-based reversible cysteine discovery platform provides a reliable workflow allowing robust detection of reversible cysteine sites at multiple time points during cardiac hypertrophy development.

2.V.B. Future Directions.

To comprehensively understand the impact of ISO-induced hypertrophy on cysteine O-PTM sites and proteins identified from this robust biotin switch based quantitative proteomic approach, we used a bioinformatics approach that dissects the temporal pattern of the cysteine O-PTM proteome using a cubic spline-based K-mean clustering (Bhasi, Forrest, and Ramanathan 2005; Straube et al. 2015). The cubic spline method generates a fitted curve for all the cysteine O-PTM abundance values across time points for each oxidation site. This de-noising method extracts the most fitted temporal pattern from a group of highly variable values with cross-validation. The smoothed temporal curve facilitates the subsequent unsupervised clustering to accurately identify temporal clusters of interests. However, the assumption of this method is that alteration across time points is gradual and smooth; this will potentially have a tradeoff with a sudden alteration between time points that is biologically significant. Thus, more

time points are required or multiple temporal analysis methods widely used in the genomics field, including advanced pattern extraction (Tchagang et al. 2009), can be adapted and applied to reveal the true biological signal from the technical bias.

2.VI. Conclusions.

Utilizing a novel cysteine O-PTM discovery platform comprised of customized redox proteomics and advanced computational analysis, we present the first proteome-wide study of multi-type cysteine O-PTM on the well-known mouse model of ISO-induced cardiac hypertrophy. The novelty and strength of our study lie in our ability to cluster the temporal behavioral profiles of cysteine residues and visualize distinct patterns of change, as well as how they correlate to the hypertrophic phenotype. The advanced computational and statistical platform enabled us to create a dynamic and integrated picture of the entire cysteine oxidative proteome, a molecular signature that can be used for in-depth profiling and defining of various health and diseased states.

CHAPTER 3: COMPUTATIONAL APPROACHES TO IDENTIFY METABOLOME FINGERPRINTS OF PATHOLOGICAL STAGES FOLLOWING HEART FAILURE TREATMENT.

3.I. Abstract.

Pathologic cardiac remodeling from a healthy to diseased state is characterized by an intricate and coordinated multitude of alterations among various biomolecules over time. Conversely, mechanical circulatory support devices (MCSD) are a promising strategy for unloading the heart and reversing this process. Understanding the molecular drivers of remodeling and reverse remodeling would inform new mechanistic insights and more nuanced diagnostics, prognostics, and therapeutics for hypertrophy and/or heart failure (HF). Plasma metabolomics holds great promise for the temporal profiling and characterization of healthy and diseased phenotypes, but large-scale technological platforms with machine learning (ML) pipelines for clinical temporal metabolomics profiling are still lacking. In this study, we established a Multiple Reaction Monitoring (MRM)-based MS quantitative platform and a corresponding ML-based computational workflow to unravel the metabolomics fingerprints of healthy humans, HF patients, and a mouse model. Accordingly, we quantified the plasma levels of 610 metabolites, and identified those exhibiting high correlation to phenotypic changes. In summary, a total of 30 metabolites were significantly altered during the cardiac remodeling: 12 metabolites were continuously upregulated during ISO-treatment, belonging to the biological categories of amino acids and sphingomyelin (SM). In parallel, 18 metabolites were downregulated, consisting of 10 free fatty acids (FFA), 6 ceramides, and 2 phosphatidylcholine (PC). Among these FFAs, we identified 3 saturated FFAs, 6 monounsaturated FFAs (with 1 double bond) and 5 polyunsaturated fatty acids (PUFA). Importantly, we have identified 4 driver metabolites that are significantly associated with changes in HW/BW ratio, all of which are FFAs: C14:0 (Myristic acid), cis-C16:1w7 (Palmitoleic acid), cis-C18:1w9 (Oleic acid), and cis-C18:3w3 (α-Linolenic acid, ALA). Our study highlights a novel ML workflow that, in conjunction with benchmark datasets, will facilitate future biomarker discovery.

3.II. Introduction.

Maladaptive cardiac remodeling and heart failure (HF) are common stages of many heart diseases and pose major public health problems in the United States (McMurray 2010). To date, advanced HF patients are clinically treated as a homogenous group with similar standard therapies (WRITING COMMITTEE MEMBERS et al. 2013). However, HF is a multifactorial disease and results not only from cardiac overload or injury, but also from a complex interplay among genetic, neurohormonal, inflammatory, and biochemical alterations, requiring refined diagnostic means to stratify patient groups for divergent therapies (S. J. Shah et al. 2015; Taylor et al. 2004). Many factors including cultural, environmental, and comorbidities (e.g., hypertension, obesity, kidney disease, and diabetes) can predict the severity and outcome of HF (East et al. 2004; Mensah et al. 2005; Husaini et al. 2011). However, rigorous, quantifiable methods that reliably characterize the complexity of changes associated with pathological remodeling and hold clinical predictive value are lacking.

Cardiac metabolism is a likely target for pathologic cardiac remodeling (Neubauer 2007), in that the mechanical nature of cardiomyocytes makes them substantial energy consumers and even more so during remodeling when workloads are heightened. Under healthy conditions, the catabolism of fatty acids through β-oxidation provides approximately 70-90% of the total ATP used by the heart (Doenst, Nguyen, and Abel 2013). In addition to fatty acids, other metabolites are supplemented into the oxidative phosphorylation pathway, including carbohydrates, amino acids, and ketone bodies (Kolwicz, Purohit, and Tian 2013). Complex metabolic pathways including lipolysis, glycolysis, fatty acid oxidation, tricarboxylic acid cycle, oxidative phosphorylation, and regulatory signaling cascades work in concert to maintain functional stability of the heart. Both basic research and clinical studies have reported a shift toward glucose metabolism from fatty acid β-oxidation in hypertrophied and failing hearts. Consequently, ATP synthesis in failing hearts is compromised, leading to cardiac energy deficits

(Kolwicz, Purohit, and Tian 2013). Furthermore, inadequate cardiac output leads to insufficient hemodynamic circulation and end-organ perfusion, which subsequently alters global metabolism. Hence, maladaptive remodeling and advanced HF are associated with disrupted levels of multiple metabolites both in the heart and the body (Neubauer 2007). Comprehensive quantitation of metabolite abundance in blood plasma, also known as plasma metabolomics profiling, can provide mechanistic insights into the molecular alterations underlying maladaptive remodeling and HF (Cheng et al. 2015). The large-scale quantitation of circulating metabolites may also identify metabolic changes for clinical classification, which may predict the outcomes of HF patients before and after optimal medical therapy and surgical interventions (Lloyd-Jones 2010). These methods can be established as important avenues for biomarker and drug-target development, as well as the validation of treatments.

With the advances in high-throughput quantification by MS and Nuclear Magnetic Resonance, components of metabolic pathways have been identified as potential biomarkers in complex diseases (Emwas 2015; Soininen et al. 2015; A. Zhang, Sun, Wang, et al. 2012; A. Zhang et al. 2016). Biomarkers are quantifiable small molecules that can be utilized as internal indicators (Chen et al. 2011; Graham et al. 2013; J. Kang et al. 2015) or therapeutic targets of disease (Arakaki, Skolnick, and McDonald 2008; Ganti et al. 2012; Tolstikov et al. 2014). Metabolic survey may offer novel mechanistic insights on disease origin and progression beyond traditional approaches (Roux et al. 2011; A. Zhang, Sun, Wu, et al. 2012). A few studies have identified metabolites as cardiometabolic risk factors (Sansbury et al. 2014; Soininen et al. 2015; Würtz et al. 2015). In these investigations, several biological classes of metabolites or metabolic pathways were targeted based on previous knowledge of cardiac pathology (Armenian et al. 2014). Thus far, a platform for unbiased, discovery-based temporal metabolomics profiling empowered by advanced machine learning-based approaches is still

missing, which has largely hindered our capabilities for identifying metabolites relevant to maladaptive remodeling and HF.

Biological Class	Metabolites	Biological Class	Metabolites
Ceramides	136	Biogenic Amines	18
Phosphatidylcholines	85	Phosphatidylglycerols	18
Free Fatty Acids	63	Eicosanoids & Other PUFAs	17
Phosphatidylethanolamines	54	Lysophosphatidylcholines	15
Acylcarnitines	42	Energy Metabolism Intermediates	9
Sphingomyelins	40	Lysophosphatidylethanolamines	9
Phosphatidylserines	27	Neurotransmitter	8
Bile Acids	22	Oxidative Status	6
Amino Acids	20	Hexoses	1
Steroids	19	Lysophosphatidylglycerols	1

Table 3.1. Biological Classes of 610 Targeted Metabolites in Mouse and Human Plasma. An MRM-based mass spectrometry approach detected and quantified the absolute abundances of a total 610 metabolites belonging to 20 biological classes in mouse and human plasma samples. The number of metabolites in each biological class is also shown.

In this study, we utilized an MRM-based MS quantitative platform to conduct large-scale plasma metabolomics profiling with the intent to capture the temporal dynamics of metabolites in plasma. We first applied this to chronic isoproterenol (ISO)-treated mouse strains with diverse susceptibilities towards β-adrenergic overstimulation, temporally characterizing 610 plasma metabolites that encompass 20 categories of biomolecules based on their chemical structure and biological function (Table 3.1). The experimental mouse dataset was used to develop a computational workflow, which elucidates the connection between metabolic dynamics and phenotypic alterations. We then applied our technical platform and corresponding computational

workflow to human datasets. Blood samples and clinical assessments were acquired from 26 HF patients receiving Mechanical Circulatory Support Device (MCSD) therapy. Temporal metabolomics profiling characterized the metabolic fingerprints following MCSD implantation and identified metabolites relevant to phenotypic alterations. Our study illustrates that a computational workflow, in conjunction with a benchmark dataset of healthy and diseased human plasma metabolomes, may pave the groundwork for future metabolomics studies with clinical translation.

3.III. Methods and Materials.

3.III.A. Experimental Procedures

Study Approvals.

Mouse experiments were performed in accordance with the National Research Council's guidelines for the care and use of laboratory animals and were approved by the Animal Research Committee (ARC) of UCLA under ARC Protocol #2002-172. Human procedures were performed according to protocols approved by the Institutional Review Board (IRB) of UCLA under IRB protocols #12-000351 and #12-000899. All participating human subjects received a detailed explanation of the study and provided written informed consent.

Mouse Model of Maladaptive Cardiac Remodeling and Sample Collection.

Cardiac maladaptive remodeling was induced in mice of six genetic strains (Figure 3.1.) with chronic isoproterenol infusion, using a previously reported protocol (Drews et al. 2010).

Human Cohort and Sample Collection.

A pilot human cohort was assembled with 26 advanced HF Patients undergoing MCSD implantation and 12 Healthy Human Subjects. The venous blood samples were obtained and clinical assessments were concomitantly documented one day before the surgery (T0), and subsequently within time windows from one day post-surgery up to 14 weeks depending on the availability of patients (Figure 3.4.).

Metabolic Profiling of Blood Plasma.

Utilizing a Multiple-Reaction-Monitoring (MRM) based MS quantitative approach, we surveyed 9 quantification panels of plasma metabolites (P180, Energy Metabolism, Free Fatty Acids, Eicosanoids and other oxidation products of PUFAs, Lipids, Steroids, Neurotransmitters, Bile acids, Oxidative status assays) using service offered by Biocrates Life Science AG (Innsbruck, Austria). We targeted a total of 610 plasma metabolites (Table 3.1); the absolute abundance levels of plasma metabolites were quantified for subsequent computational analyses.

Healthy Human Subjects									
No.	Age (y/o)	Gender	Ethnicity	BMI	No.	Age (y/o)	Gender	Ethnicity	BMI
H1	61	F	Caucasian	25.6	H7	38	M	Asian	25.0
H2	43	F	Caucasian	25.0	H8	40	M	Asian	25.8
H3	58	M	Asian[a]	23.3	H9	51	F	Asian	22.7
H4	39	M	Hispanic	25.6	H10	36	M	Asian	30.5
H5	45	F	Asian	22.5	H11	34	M	Asian	24.9
H6	46	M	Asian	25.3	H12	27	M	Asian	24.7

Heart Failure Patients									
No.	Age (y/o)	Gender	Ethnicity	BMI	Etiology	SOFA1	SOFA2	SOFA3	INTER MACS
P1	59	F	Asian	23.3	Non-ischemic	6	3	2	2
P2	65	M	Caucasian	27.6	Non-ischemic	8	15	18	2
P3	55	M	Caucasian	21.0	Non-ischemic	5	5	2	Medium[c]
P4	38	F	Hispanic	29.7	Non-ischemic	5	4	2	3

No.	Age (y/o)	Gender	Ethnicity	BMI	Etiology	SOFA1	SOFA2	SOFA3	INTER MACS
P5	24	F	Hispanic	32.4	Non-ischemic	4	1	N/A	3
P6	43	M	Hispanic	28.3	Non-ischemic	12	2	N/A	3
P7	57	M	Caucasian	24.1	Non-ischemic	7	3	1	3
P8	74	M	Caucasian	28.0	Ischemic	14	4	3	2
P9	62	M	Asian	28.0	Ischemic	16	8	N/A	1
P10	64	M	Caucasian	27.3	Non-ischemic	8	6	3	2
P11	43	M	African American	24.1	Non-ischemic	3	4	1	Medium
P12	30	F	Caucasian	20.1	Non-ischemic	4	1	N/A	3
P13	65	M	Hispanic	26.3	Ischemic	7	13	4	2
P14	68	M	Hispanic	25.6	Non-ischemic	10	9	2	2
P15	61	M	Hispanic	24.7	Non-ischemic	6	8	3	2
P16	71	M	Caucasian	19.7	Ischemic	5	4	2	4
P17	66	M	Hispanic	22.0	Non-ischemic	6	4	1	3
P18	62	M	Caucasian	23.4	Ischemic	10	15	N/A	2
P19	37	F	Caucasian	29.2	Non-ischemic	5	6	10	2
P20	61	M	Caucasian	21.0	Non-ischemic	13	7	4	2
P21	63	M	Other[b]	20.3	Ischemic	12	17	N/A	1
P22	36	M	African American	27.0	Non-ischemic	8	N/A	1	2
P23	46	M	African American	22.4	Non-ischemic	13	3	4	2
P24	80	M	Caucasian	16.3	Ischemic	8	7	21	3
P25	59	F	Hispanic	47.9	Non-ischemic	15	19	N/A	2
P26	65	M	Hispanic	22.5	Non-ischemic	4	5	3	2

Table 3.2 Demographics of the Human Cohorts. A sample cohort was assembled with 12 healthy human subjects and 26 advanced heart failure patients undergoing Mechanical Circulatory Support Device (MCSD) implantation. BMI represents Body Mass Index; SOFA1 represents SOFA score of patient on 1 day pre-surgery; SOFA2 represents SOFA score of patient 5-6 days post-surgery; and SOFA 3 represents SOFA score of patient 1 month (or closest date to 1 month) post-surgery. a. Asian represents Asian and Pacific Islander; b. Other indicates no available ethnicity record; c. Medium presents all profiles above profile 4.

3.III.B. Computational Analyses and Annotation Workflow.

Data Pre-processing.

To construct a reliable temporal dataset of metabolomics profiling for subsequent computational analyses, the original dataset was pre-processed to remove redundancy and to replace missing values.

Differential Expression Analysis with t-Tests.

To characterize the temporal dynamics of plasma metabolites in response to maladaptive cardiac remodeling in mice, a Differential Expression Analysis was performed. We applied a Multiple Testing Correction to the p-values calculated by the t-tests, using the Benjamini-Hochberg method to reduce the false positives (Benjamini and Hochberg 1995). In parallel, t-tests were applied to 12 Healthy Human Subjects and 26 advanced HF Patients undergoing MCSD implantation. For both mice and humans, a threshold of adjusted p-value < 0.001 and ±1.5 fold change was applied to detect metabolites with significant changes. Volcano plots, Heatmaps, and PCAs were generated to depict the results of Differential Expression Analyses.

Identification of Housekeeping Metabolites.

A certain set of metabolites are regularly produced and maintained by the basic processes of metabolism. This set of metabolites can serve as internal controls to adjust for experimental batch effects, and individual differences prior to an experiment. Leveraging the concept of housekeeping proteins, we defined the housekeeping metabolites as constitutively expressed components to maintain the basic cellular function. Accordingly, they sustain a constant plasma concentration over time. As oppose to metabolites that were significantly altered over time, housekeeping metabolites are selected with an adjusted p-value > 0.1 in the comparison between Untreated mice and ISO-treated mice for the mouse population. For the human datasets, housekeeping metabolites were selected using the same criteria.

Adjustment for Sample Relatedness Using a Linear Mixed Model.

In the effort of identifying plasma metabolites that are highly associated with the phenotypic characteristics of ISO treatment or MCSD implantation (i.e., HW/BW or SOFA Score), we applied the following Linear Mixed Model to account for any population relatedness represented

by the metabolite expression profiles (Furlotte and Eskin 2015). In this model, y represents the measurable phenotypic characteristics, μ is the phenotypic mean of a population, β represents the strength of metabolite effects, x is the metabolite expression profiles, u represents the effects of population stratification with Var(u) = σ2gK, and ε represents the measurement errors which follows a normal distribution with mean of zero and variance of σ2e. In a genome-wide association study, the population structure is mainly defined by the genetic relatedness among strains or individuals, and thus the relatedness matrix (K) is constructed using all genotypes. However, the concentrations of plasma metabolites can be influenced by multidimensional factors including genetic variances, behavior patterns, and environmental impacts. Accordingly, it is more adequate to use the basal expression levels of the "housekeeping" metabolites to construct the relatedness matrix. The variances σ2g and σ2e are estimated using a restricted maximum likelihood (REML) approach, and the phenotype association is performed through computing the F-statistic to test whether β = 0. The relatedness matrix construction, restricted maximum likelihood estimation, and the F-statistic computation are carried out using "Pylmm". Since each metabolite is tested independently, the Benjamini-Hochberg procedure is applied to correct for multiple testing. A metabolite is identified as highly associated with the phenotypic characteristics if its associated weight (β) is significantly greater or smaller than zero with an adjusted p-value < 0.01.

3.IV. Results.

3.IV.A. Characterization of Plasma Metabolome Dynamics in Mice following ISO-Treatment.

Temporal Dynamics of Mouse Plasma Metabolome During Maladaptive Remodeling.

We conducted ISO treatment of six genetic mouse strains that previously displayed variable susceptibilities towards hypertrophy caused by chronic β-adrenergic stimulation (Figure 3.1.)

(Rau et al. 2015). Using plasma metabolomics profiling, we obtained a temporal dataset containing 6 mouse strains, 6 time points, and 610 metabolites (Table 3.1) for computational analyses (see details in Experimental Procedures). Accordingly, an integrated computational workflow was conceptualized to facilitate data interpretation. To identify metabolites sensitive to the progression of ISO-induced cardiac remodeling, the differential expression patterns of all 610 metabolites were investigated over time using paired Wilcoxon Signed-Rank tests followed by the multiple testing corrections (FDR < 0.05 as the threshold).

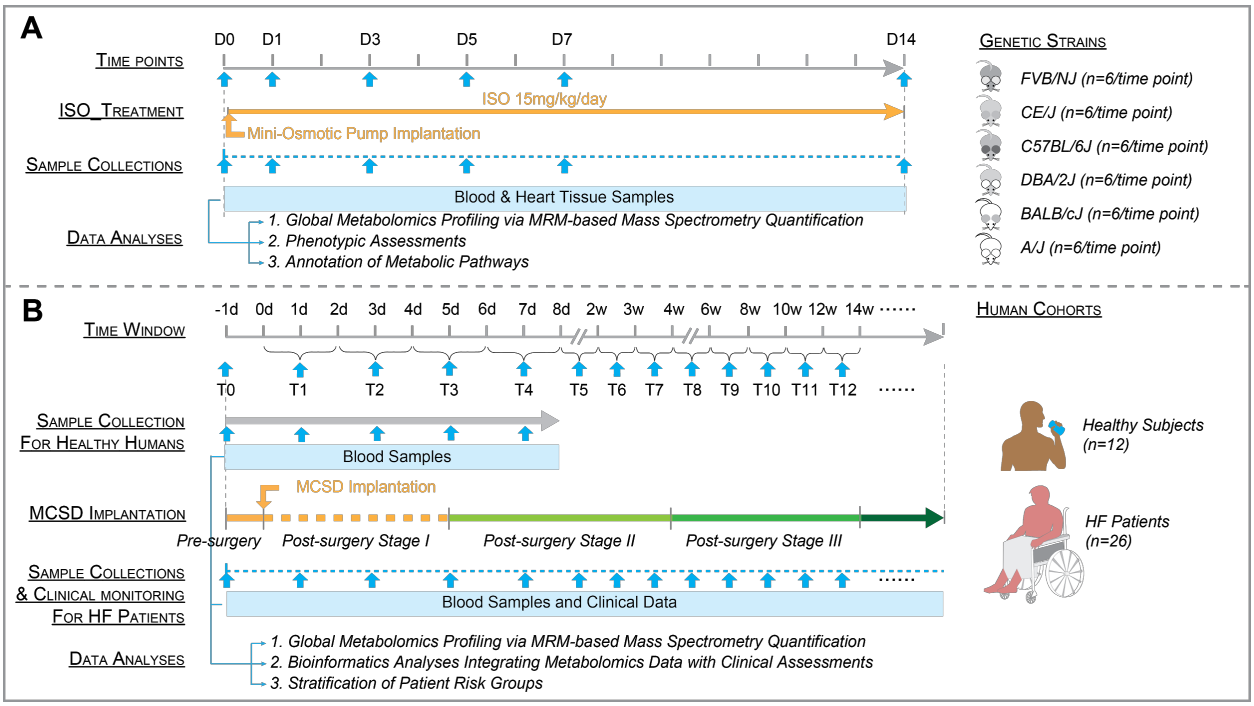
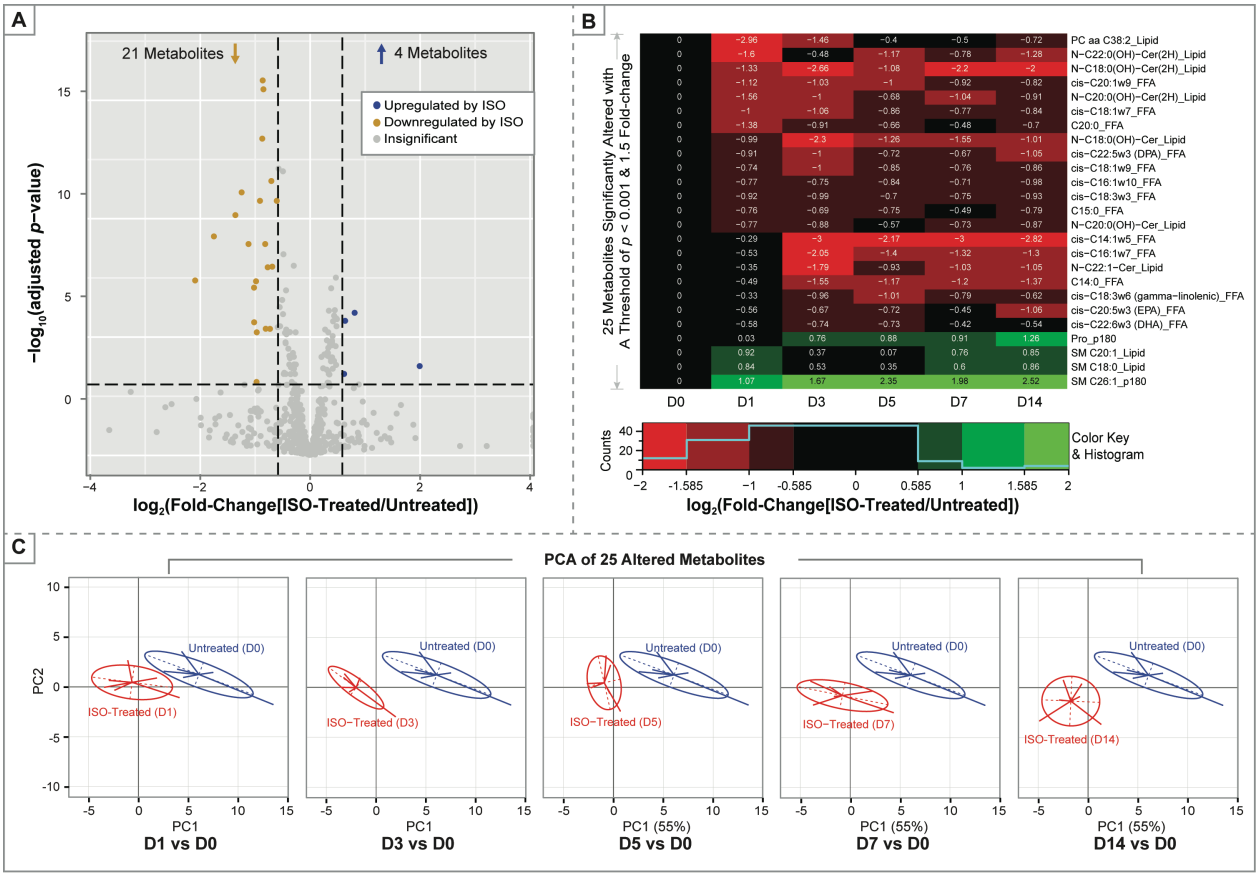


Figure 3.1. Schematic Overview of Experimental Protocol for Global Plasma Metabolomics Profiling in ISO-stimulated Mice, Healthy Humans, and HF Patients. (A) Six genetic mouse strains with varying susceptibilities towards maladaptive cardiac remodeling by isoproterenol (ISO) stimulation were continuously treated with 15mg/kg/day ISO for 14 days using surgically implanted mini-osmotic pumps. At baseline (time point D0) and during ISO treatment (time points D1, D3, D5, D7, and D14), blood samples and whole hearts were collected for plasma metabolomics profiling (3 technical replicates), cardiac phenotypic assessments, and annotation analyses of metabolic pathways. (B) A total of 26 advanced heart failure (HF) patients undergoing mechanical circulatory support device (MCSD) implantation were analyzed for their plasma metabolomics profiles while their clinical data were obtained. Venous blood samples were collected one day prior to surgery (-1d, labeled as T0) and subsequently at 12 time periods up to 14 weeks post-surgery (14 weeks, T12). The post-surgery period was subdivided into Stage I (surgical recovery, T0-T3), Stage II (short-term post-MCSD, T4 up to T7), and Stage III (long-term post-MCSD, T8-T12). A cohort of 12 healthy human subjects was also analyzed for their plasma metabolomics profiles at 5 time periods as control (T0-T4).

A total of 30 metabolites were significantly altered across all time points (Figure 3.2.C). Among these altered metabolites, 12 metabolites (Proline, Sphingomyelins C18:0, C20:1, C26:1, etc.) belonging to the biological categories of amino acids and sphingomyelin (SM) were continuously upregulated during ISO treatment. In parallel, 18 metabolites consisting of 10 free fatty acids (FFA), 6 ceramides, and 2 phosphatidylcholine (PC). Among these FFAs, we identified 3 saturated FFAs, 6 monounsaturated FFAs (with 1 double bond) and 5 polyunsaturated fatty acids (PUFA) were downregulated (Figure 3.2.A). A temporal trend in fold change among the 25 differentially expressed metabolites was visualized using a Heatmap (Figure 3.2.B). Each individual metabolite displays a unique pattern across the 14 days of treatment, with some metabolites (e.g., Proline and Sphingomyelin) continually increasing, and others (e.g. FFAs) reaching a sustained decrease. Using the 25 significantly altered metabolites, Principal Component Analysis (PCA) was able to stratify the untreated (D0) mice from the ISO-treated mice (D1, D3, D5, D7, D14), indicating that the screening process is sufficient to recognize metabolic distinctions between pathological stages (Figure 3.2.B and 3.2.C). At each time point, Unsupervised Hierarchal Clustering was applied to stratify the 6 genetic strains (Figure 3.2.D). Notably, each heatmap displays unique relationships between expression patterns, suggesting that the metabolic alterations are sensitive to both genetic predispositions and temporal development.



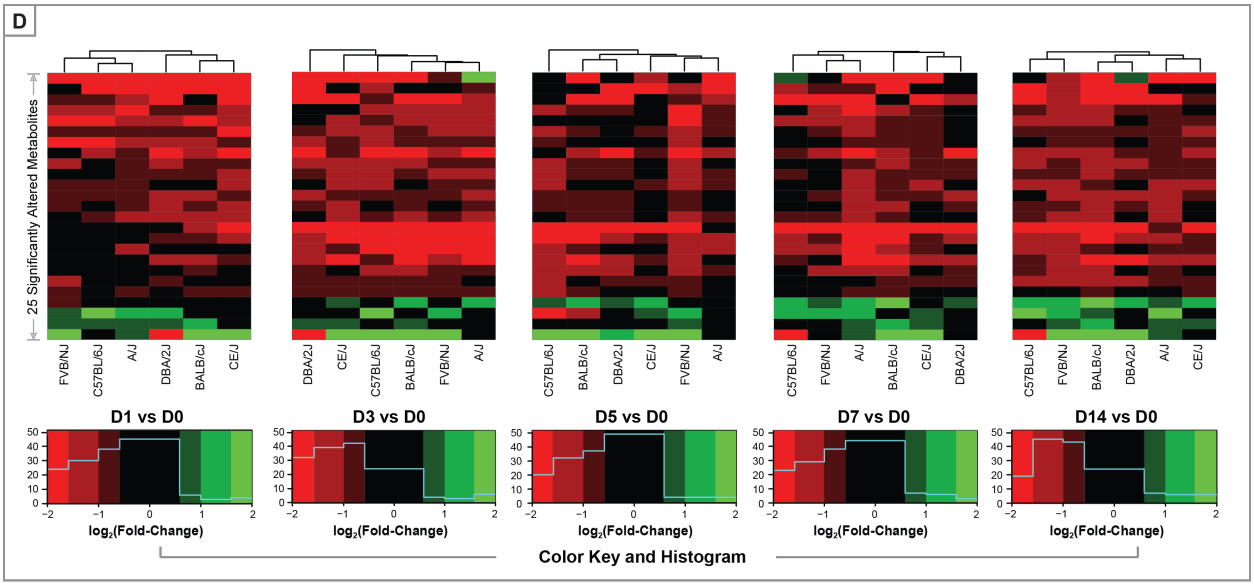


Figure 3.2. Metabolomics Dynamics in Mouse Plasma During ISO-induced Cardiac Remodeling. The absolute abundances of 610 plasma metabolites were quantified in six genetic mouse strains during ISO stimulation and subjected to statistical analyses to identify differentially expressed metabolites. Over time, these metabolites underlie the metabolic responses associated with the progression of ISO-induced

maladaptive remodeling of the heart. The results indicate that the global metabolomics profiling is a sensitive approach to identify metabolic markers with pathophysiological relevance. **(A)** A paired t-test was performed to identify metabolites significantly altered following ISO treatment (D0 vs. D1, 3, 5, 7 and 14). One volcano plot depicted as Log2 (Fold-Change expressed as ISO-treated/Untreated) (X-axis) against –Log10 (adjusted p-value) (Y-axis) was generated accordingly. A threshold ($p < 0.001$ and ±1.5 Fold-Change, dotted lines) was applied to identify metabolites with significant changes. During ISO-induced maladaptive cardiac remodeling, 21 metabolites were downregulated (Yellow dots) and 4 were upregulated (Blue dots). The other metabolites did not display significant changes (Grey dots). **(B)** A temporal Fold-Change profile was generated for the 25 differentially expressed metabolites identified in panel 2A. The heatmap shows the metabolite levels expressed as Fold-Change between ISO-treated mice on consecutive time points (D1, D3, D5, D7 and D14) and Untreated mice (D0). For all metabolites, the Fold-Change value at each time point is indicated in each tile. The color keys (green, upregulation; red, downregulation) and histograms (blue line, distribution of metabolites in each color range) are displayed at the bottom of heatmap. Both Proline (Pro) and Sphingomyelin (SM) showed an increasing Fold-Change over time while multiple species of Ceramides (Cer) and Free Fat Acids (FFA) showed consistent downregulation. **(C)** Principal Component Analyses (PCAs) were performed on the 25 significantly altered metabolites (Panel 2A) to compare ISO-treated mice during multiple time points (D1, D3, D5, D7, and D14) to Untreated mice (D0). The X-axis is depicted as the first principal component (PC1) representing the space with the largest variance in data. And the Y-axis is depicted as the second principal component (PC2) representing the space with the second largest variance. In each scatterplot, ISO-treated mouse strains are colored in red and untreated mouse strains are colored in blue. Ovals are 95% inertia ellipses. Of the 25 metabolites, a clear separation between the ISO-treated mice and untreated mice can be observed starting on D1. The more profound separation on D3, D5, D7 and D14 reflects prolonged and more severe stress by ISO treatment. In contrast, a PCA of all 610 metabolites did not show a separation during all time points, which demonstrates the identified metabolites in panel 6A-C reflecting pathophysiological changes over time. **(D)** A Hierarchical Clustering was generated using Spearman correlation coefficients of the metabolites Fold-Change profiles. Accordingly, the heatmaps display the metabolite levels expressed as a Fold-Changes between Untreated mice (D0) and ISO-treated mice at multiple time points (D1, D3, D5, D7, and D14) of the 25 differentially expressed metabolites (Panel 2A). The color keys (green, upregulation; red, downregulation) and histograms (blue line, distribution of metabolites in each color range) are displayed at the bottom of heatmap. The dendrograms on the top of the heat maps represent the distance among the mouse strains.

Metabolomics Fingerprints of Six Unique Genetic Mouse Strains with Differing Susceptibilities to ISO Stimulation.

To identify metabolomics fingerprints that closely associate with phenotypic characteristics, the heart weight-body weight (HW/BW) ratios were documented concomitantly with the plasma sample collection to provide temporal phenotypic profiles in parallel with metabolic alterations. During ISO treatment, mouse strains C57BL/6J, CE/J and FVB/NJ exhibited relatively lower severity of cardiac hypertrophy; the HW/BW ratios ended at ~120% of baseline after 14 days. The result agrees with previous observations that these three mouse strains are more resilient to cardiac stress following ISO treatment (Rau et al. 2015). In contrast, strains DBA/2J, A/J and BALB/cJ showed more substantial hypertrophy, with HW/BW ratios ~140% after 14 days, demonstrating a greater degree of susceptibility to ISO-induced cardiac stress (Figure 3.3.A).

In our computational workflow, a Linear Mixed Model (LMM) was employed to correlate temporal metabolome dynamics with HW/BW ratios. To account for individual relatedness and thereby improve statistical accuracy, we incorporated a Relatedness Matrix Correction into the LMM. To generate the matrix, we selected the metabolites with an adjusted p-value > 0.1 in the aforementioned paired t-test and a total of 309 metabolites, termed “housekeeping metabolites”, were identified. These “housekeeping” metabolites maintain consistent levels in plasma over time, whereas their similarities among mouse strains indicate individual relatedness. We also performed a Multiple Testing Correction to control the false discovery rate. Applying a threshold of < 0.01 on adjusted p-values, 4 metabolites were identified as significantly associated with changes in HW/BW ratio, all of which are FFAs: C14:0 (Myristic acid), cis-C16:1w7 (Palmitoleic acid), cis-C18:1w9 (Oleic acid), and cis-C18:3w3 (α-Linolenic acid, ALA). Their temporal changes in plasma levels are displayed in parallel with HW/BW ratios (Fig 3B). In mouse strains sharing similar susceptibility to ISO treatment, the 4 metabolites displayed similar temporal behaviors during maladaptive remodeling.

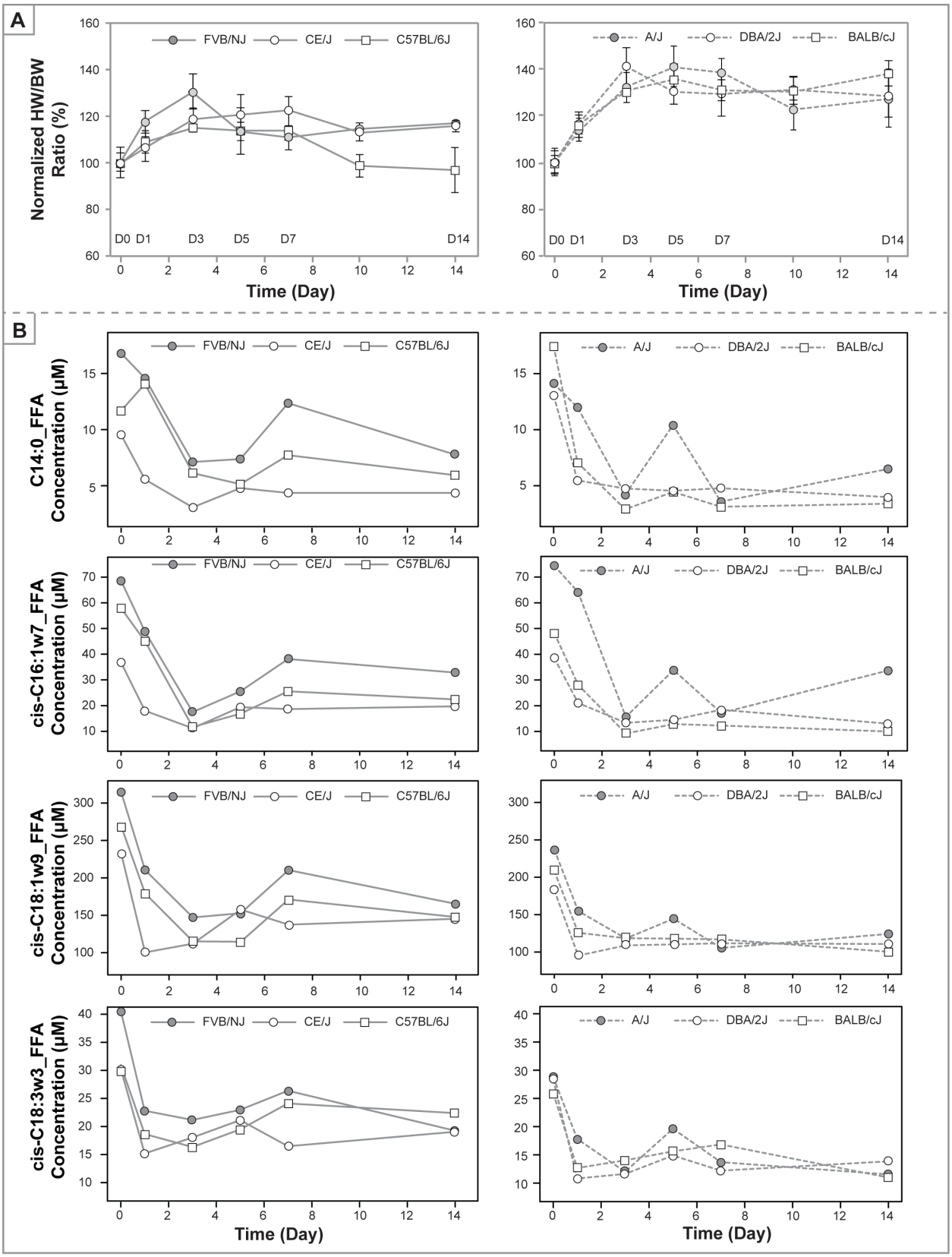

A
Normalized HW/BW Ratio (%)
FVB/NJ
CE/J
C57BL/6J
A/J
DBA/2J
BALB/cJ
D0 D1 D3 D5 D7 D14
Time (Day)
B
C14:0_FFA Concentration (μM)
cis-C16:1w7_FFA Concentration (μM)
cis-C18:1w9_FFA Concentration (μM)
cis-C18:3w3_FFA Concentration (μM)
Time (Day)

Figure 3.3. Phenotypic Alteration and Corresponding Metabolic Fingerprints During ISO-induced Cardiac Remodeling in Mice.
(A) The heart weight–body weight ratios (HW/BWs) were temporally measured for the six mouse strains during ISO-induced cardiac remodeling (time points D1, D2, D3, D5, and D14), and subsequently normalized to HW/BW of untreated mice (D0). The BALB/cJ, A/J, and DBA/2J mouse strains showed an elevation up to 40% at an earlier stage, indicating a more severe response to ISO treatment; the C57BL/6J, CE/J, and FVB/NJ mouse strains displayed a stabilized increase of 20% at D14, suggesting a more resilient response. **(B)** Four metabolites (C14:0, cis-C16:1w7, cis-C18:1w9, cis-C18:3w3) were identified through paired t-test and LMM analyses; their absolute abundances were displayed over time, in line with the mice's susceptibilities towards ISO treatment. Within each mouse strain, the four metabolites presented consistent temporal dynamics, indicating shared or synchronized metabolic pathways. In mouse strains sharing similar susceptibility towards ISO treatment, the four metabolites displayed comparable behavior during the progression of maladaptive remodeling. For example, Balb/c, A/J, and DBA/2J mouse strains (ISO-susceptible) showed relatively less basal abundance of α-linolenic acid (cis-C18:3w3) than the C57, CE/J, and FVB mouse strains (ISO-resilient). During 14days' ISO-treatment, ISO-resilient mouse strains displayed better capability to immediately restore the plasma level of α-linolenic acid than ISO-susceptible strains

3.IV.B. Characterization of Plasma Metabolomics Dynamics in HF Patients Receiving MCSD Implants.

Enrollment and Demographics of Human Cohort.

To translate our computational workflow from mouse dataset analyses to human data interpretation, we first assembled a test cohort over a 2-year period from Aug 2012 to Nov 2014, including 12 healthy human volunteers (H1, H2, … H12) and 26 advanced HF patients (NYHA III-IV) who had received MCSD implants (P1, P2, … P26) (Table 3.2). The healthy human group contains 9 Asians, 2 Caucasians, and 1 Hispanic with an average age of 43 ± 10 (range: 27-61) years. The healthy humans include 4 females and 8 males with an average Body Mass Index (BMI) of 25.1 ± 2.0 (range: 20.1-32.4), presenting a normal distribution of body type. The HF patients include 11 Caucasians, 9 Hispanics, 3 African-Americans, 2 Asians, and 1 without an ethnic record. This group includes 6 females and 23 males with an average age of 56 ± 14 (range: 24-80) years. The average of BMI was 25.5 ± 5.9 (range: 16.3-47.9). The etiological categories of HF patients were classified as ischemic cardiomyopathy (ICM) or non-ischemic cardiomyopathy (NICM). The Sequential Organ Failure Assessment (SOFA) score of each individual patient was assessed 1 day before MCSD implantation and continuously monitored

post-surgery. Within the 5-day period after MCSD implantation, all patients showed a significant increase in SOFA score that is most likely induced by post-surgical inflammatory response and recovery. We exhibited this period as Stage I (T1-T3, up to 6 days post-surgery, Fig 5). A short-term clinical assessment Stage II (T4-T7, 7 days to 28 days post-surgery) was chosen to evaluate the direct results of MCSD implants and their metabolic imprint due to improved systemic hemodynamics and vital signs. A long-term phase was indicated as Stage III (T8-T12, 29 days up to 120 days post-surgery) to determine the metabolic imprints of the patients with respect to their recovery and outcome. Before surgery, the INTERMACS score for each patient was assessed to classify the HF patients before MCSD implantation (7.79% in profile 1; 53.85% in profile 2; 26.92% in profile 3; 3.85% in profile 4, and 7.69% in profile 5-7).

Plasma Metabolomics Stability in Healthy Human Subjects.

To validate the practicality and reliability of our technical platform and corresponding computational workflow, blood samples were collected from 12 Healthy Human Subjects at 5 consecutive time points (T0, T1, …T4). A total of 610 metabolites were absolutely quantified for the time series of samples. A paired t-test showed that only one metabolite, Cortisol, was significantly altered over time (Figure 3.5.A). Cortisol is a human steroid hormone released in response to stress and also sensitive to the Circadian rhythm. In addition, cortisol is only slightly above the threshold to be considered significant, suggesting general metabolome stability over time in Health Human Subjects. In addition, almost all metabolites were identified without statistical significance from these biological replicates, in which no pathological difference should exist. The results suggest that our technical platform and computational workflow only address metabolites in response to pathological alterations and clinical interventions.

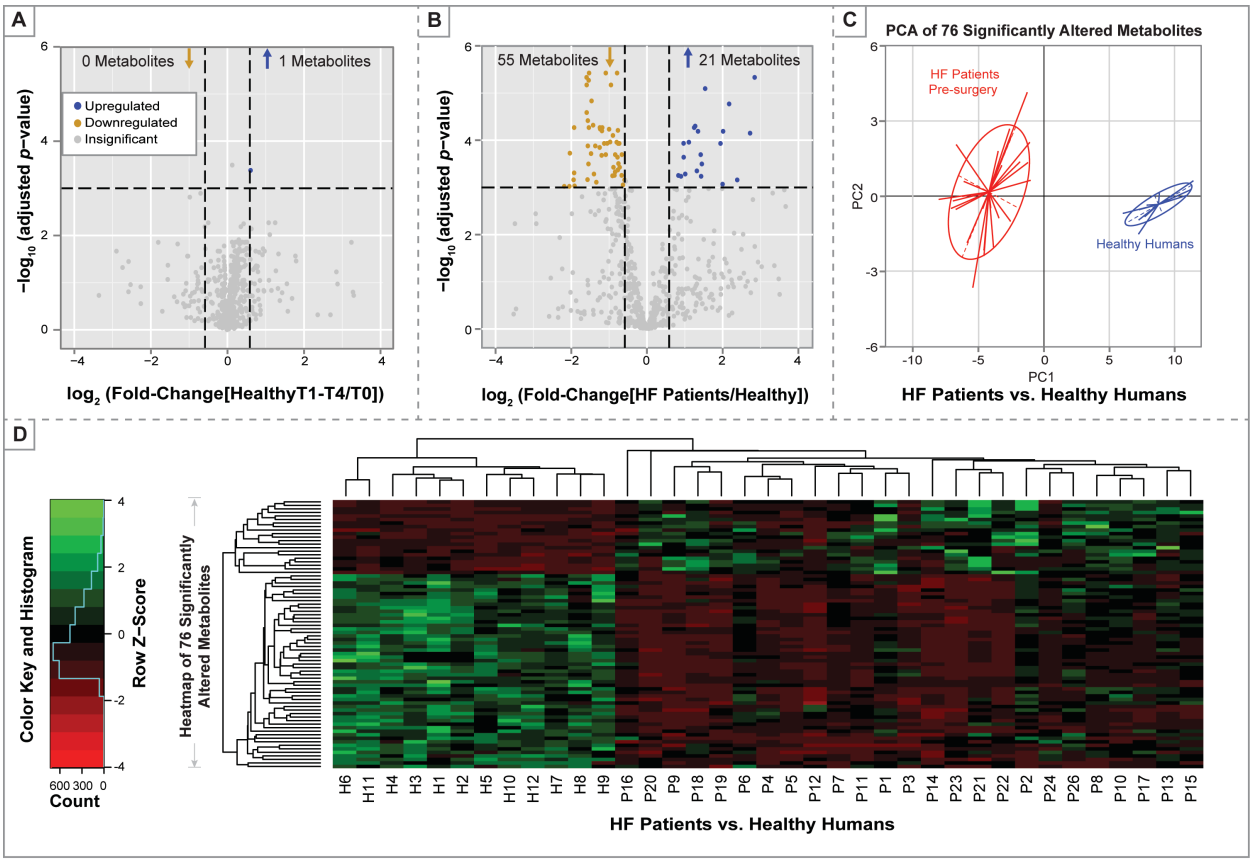
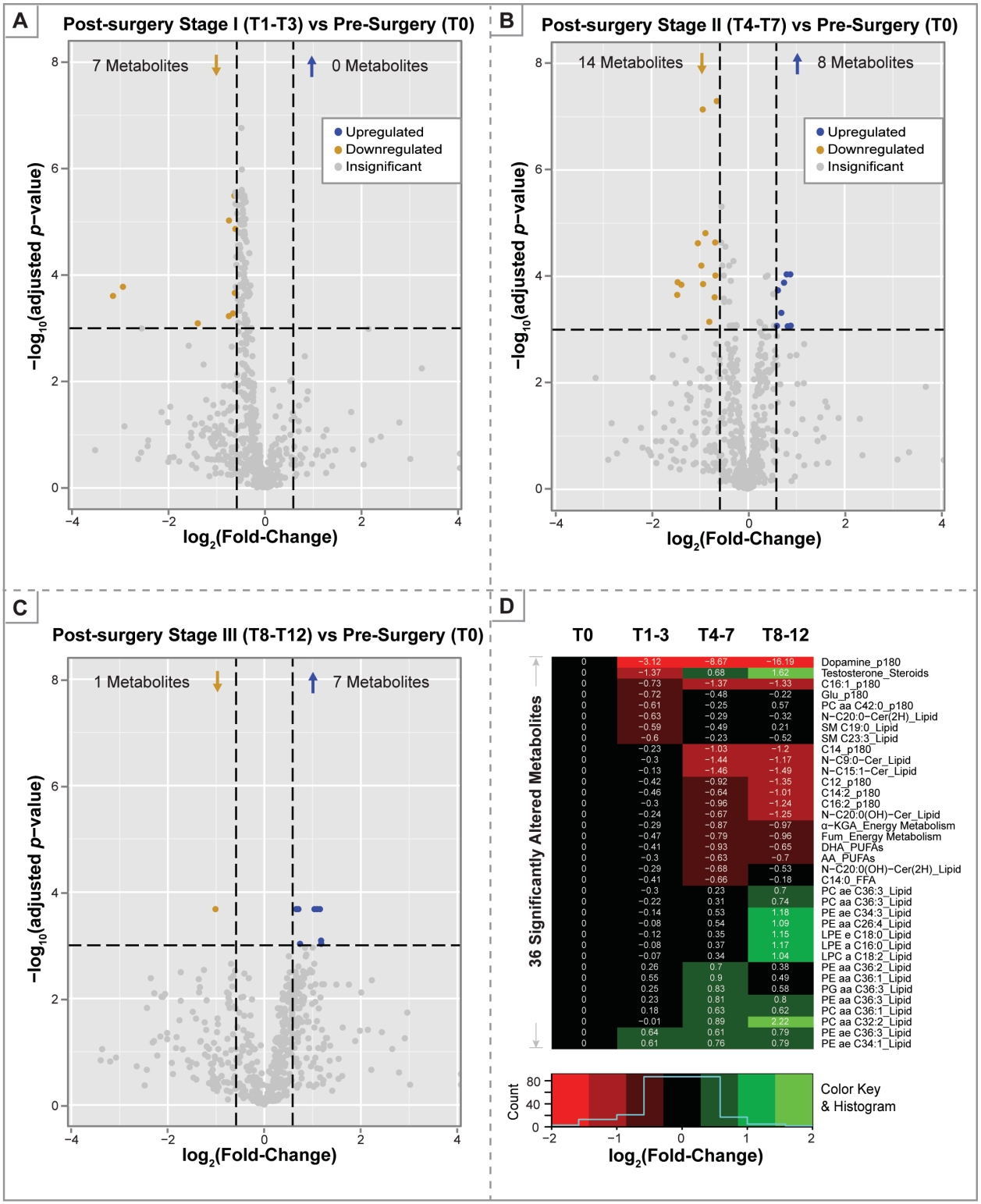


Figure 3.4. Healthy Human Subjects Plasma Metabolomics Stability and Their Distinction from HF Patients. The absolute abundances of plasma metabolites were quantified in 12 Healthy Human Subjects at 5 consecutive time points (T0-T4) and in 26 HF Patients before receiving MCSD implantation (T0). Resulting datasets were subjected to statistical analyses to identify differentially expressed metabolites. **(A)** A paired *t*-test was performed to generate a volcano plot depicted as Log_2 (Fold-Change expressed as T1-T4/T0) (X-axis) against $-Log_{10}$ (adjusted *p*-value) (Y-axis) to highlight the differential expression patterns of 610 plasma metabolites quantified in 12 Healthy Human Subjects over 5 consecutive time points (T0-T4). A threshold ($p < 0.001$ and ±1.5 Fold-Change, dotted lines) was applied to identify metabolites with significant changes. Over T0-T4 only one metabolite was upregulated (Blue dot) and no metabolites were downregulated (Yellow dots) in healthy humans. The other metabolites did not display significant changes (Grey dots), indicating a notable stability of the plasma metabolome under basal conditions. **(B)** With the same threshold as in Panel **(A)**, a *t*-test was performed to compare the differential expression patterns of 610 metabolites in Healthy Human Subjects and HF Patients before MCSD implantation. The volcano plot illustrates 76 metabolites that were differentially expressed, of which 21 metabolites were downregulated (Yellow dots) and 55 were upregulated (Blue dots) in HF patients, respectively. **(C)** Using the 76 metabolites identified in Panel **(B)**, the Principal Component Analysis (PCA) reveals that Healthy Human Subjects and HF Patients are evidently separated, indicating distinct metabolomics fingerprints that differentiate healthy and compromised plasma metabolomes. **(D)** The 76 metabolites identified in Panel **(B)** were further subjected to unsupervised hierarchal clustering; this method validates that the HF Patients are well-clustered and separate from the Healthy Human Subjects, corroborating that the two cohorts have distinct metabolomic fingerprints.


A
Post-surgery Stage I (T1-T3) vs Pre-Surgery (T0)
7 Metabolites
0 Metabolites
Upregulated
Downregulated
Insignificant
−log10(adjusted p−value)
log2(Fold-Change)
B
Post-surgery Stage II (T4-T7) vs Pre-Surgery (T0)
14 Metabolites
8 Metabolites
Upregulated
Downregulated
Insignificant
−log10(adjusted p−value)
log2(Fold-Change)
C
Post-surgery Stage III (T8-T12) vs Pre-Surgery (T0)
1 Metabolites
7 Metabolites
−log10(adjusted p−value)
log2(Fold-Change)
D
T0
T1-3
T4-7
T8-12
36 Significantly Altered Metabolites
Dopamine_p180
Testosterone_Steroids
C16:1_p180
Glu_p180
PC aa C42:0_p180
N−C20:0−Cer(2H)_Lipid
SM C19:0_Lipid
SM C23:3_Lipid
C14_p180
N−C9:0−Cer_Lipid
N−C15:1−Cer_Lipid
C12_p180
C14:2_p180
C16:2_p180
N−C20:0(OH)−Cer_Lipid
α−KGA_Energy Metabolism
Fum_Energy Metabolism
DHA_PUFAs
AA_PUFAs
N−C20:0(OH)−Cer(2H)_Lipid
C14:0_FFA
PC ae C36:3_Lipid
PC aa C36:3_Lipid
PE ae C34:3_Lipid
PE aa C26:4_Lipid
LPE e C18:0_Lipid
LPE a C16:0_Lipid
LPC a C18:2_Lipid
PE aa C36:2_Lipid
PE aa C36:1_Lipid
PG aa C36:3_Lipid
PE aa C36:3_Lipid
PC aa C36:1_Lipid
PC aa C32:2_Lipid
PE ae C36:3_Lipid
PE ae C34:1_Lipid
Count
Color Key
& Histogram
log2(Fold-Change)

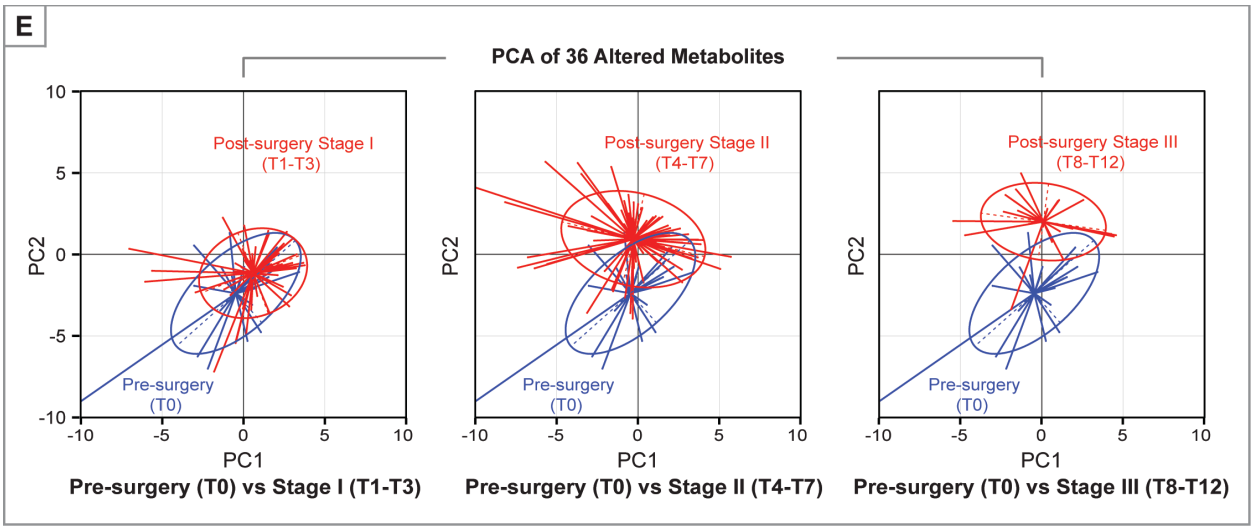


Figure 3.5. Metabolomics Dynamics in Plasma from HF Patients Before and After MCSD Implantation. The absolute abundances of plasma metabolites were quantified in HF patients receiving MCSD implantation and subjected to statistical analyses to identify differentially expressed metabolites. The significantly altered metabolites over time underlie the metabolic responses associated with mechanical unloading of the heart during advanced HF.
A paired *t*-test was performed to generate volcano plots depicted as Log_2 (Fold-Change expressed as post-surgery/pre-surgery) (X-axis) against $-Log_{10}$ adjusted *p*-value (Y-axis) to highlight the differential expression patterns of 610 plasma metabolites before (T0) and after the MCSD implantation at **Stage I** (surgical recovery, T1 - T3), **Stage II** (Short term post-MCDS, T3 - T7), and **Stage III** (Long term post-MCDS, T7 - T12). A threshold ($p < 0.001$ and ±1.5-fold change, dotted lines) was applied to identify metabolites with significant changes In Stage I **(A)**, a total of 7 metabolites were significantly altered, of which 7 were upregulated (Blue dots) while none were downregulated (Yellow dots) compared with T0. In Stage II **(B)**, a total of 22 metabolites were significantly altered, of which 8 were downregulated (Yellow dots) while 14 were upregulated (Blue dots) compared with T0. In Stage III **(C)**, a total of 8 metabolites were significantly altered, of which 7 were downregulated (Yellow dots) while 1 was upregulated (Blue dots) compared with T0. The other metabolites did not display significant changes (Grey dots). **(D)** A temporal Fold-Change profile was generated for the sum of 36 differentially expressed metabolites found in Stage I-III. The color keys (green, upregulation; red, downregulation) and histograms (blue line, distribution of metabolites in each color range) are displayed at the bottom of each heatmap. Dopamine_P180 showed an decreasing Fold-Change over time, while PC aa C32:2_Lipd showed an increasing fold-change over time; The other metabolites were consistently upregulated, or downregulated over time. **(E)** Three PCAs were performed for post-surgery Stages I, II and III against the metabolic conditions before MCSD implantation (T0). In each scatterplot, post-surgical status of MCSD recipients are colored in red and their pre-surgical status (T0) are colored in blue. Ovals are 95% inertia ellipses. Of the 36 metabolites, no separation can be observed between Stage I and T0. However, Stage II exhibited a moderate separation and Stage III displayed an increased separation, reflecting the metabolic impacts of prolonged MCSD unloading for advanced HF patients.

Differences of Human Plasma Metabolomes between Healthy and HF Conditions.

To assess the scale of metabolic distinction between Healthy Human Subjects and HF Patients, a t-test was applied to data points of Healthy Human Subjects versus HF Patients before their MCSD implantation (Figure 3.5.B). A total of 21 metabolites were upregulated in HF Patients

while 55 metabolites were downregulated. Performing PCA with these 76 significantly differentiated metabolites, the first two principal components clearly separated healthy humans from HF Patients before MCSD implants (Figure 3.5.C). In parallel, an Unsupervised Hierarchical Clustering analysis was applied and the resulting heatmap and dendrograms revealed that patients are well clustered away from healthy humans. Furthermore, the upregulated metabolites and downregulated metabolites are clustered separately with unique relationships (Figure 3.5.D). Through the Differential Expression Analysis, we identified metabolites that are potentially related to disease-induced metabolic abnormality. However, end-stage HF patients displayed notable pathological conditions dissimilar to those of healthy humans; thus, it is not surprising to observe that these severe physical dysfunctions underlie drastic metabolic differences.

Temporal Metabolic Alterations Induced by MCSD Implantation in HF Patients.

Beyond the characterization of readily-discernable metabolomics distinctions between healthy humans and HF patients, we aimed to pinpoint the metabolic alterations before and after MCSD implantation. Paired t-tests were performed to characterize the differential expression patterns of 610 plasma metabolites at the following time points: post-surgery Stage I (surgical recovery, up to 6 day post-surgery, see details in Experimental Procedures), Stage II (short-term post-MCSD, up to 4 weeks post-surgery), and Stage III (long-term post-MCSD, up to 14 weeks post-surgery). In Stage I, only 7 metabolites were significantly downregulated (Figure 3.6.A). Subsequent analysis revealed 14 downregulated and 8 upregulated metabolites in Stage II. Notably, all 8 upregulated metabolites are Glycerophospholipids, while the majority of downregulated metabolites belong to biological classes of Acylcarnitines and Ceramides (Figure 3.6.B). Finally, we found that 1 Acylcarnitine was downregulated and 7 Glycerophospholipids were significantly upregulated within Stage III (Figure 3.6.C). Accordingly, a temporal trend of fold change among the 36 differentially expressed metabolites was visualized using a heatmap

(Figure 3.6.D). Plasma metabolites such as Dopamine, Testosterone and several phospholipids exhibited a steady fold change after MCSD implantation reflecting the temporal effects of mechanical unloading of the heart. Three PCAs were performed for post-surgery Stages I, II and III against the metabolic conditions before MCSD implantation (T0). Of the 36 metabolites that were cumulatively found to be differentially expressed in either Stages I, II or III, no separation could be observed between Stage I and T0. However, Stage II exhibited a moderate separation and Stage III displayed an increased separation, reflecting the metabolic impacts of prolonged MCSD unloading for advanced HF patients.

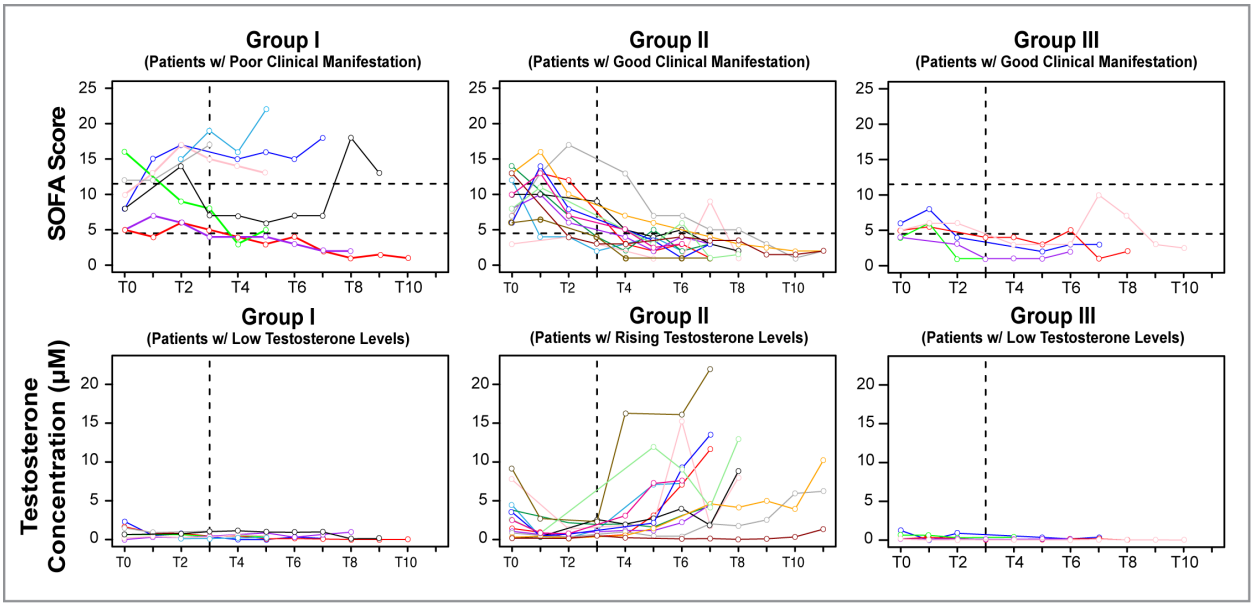


Figure 3.6. Temporal Changes of Clinical Manifestation and Corresponding Metabolomic Dynamics in HF Patients Following MCSD Implantation. In conjunction with Differential Expression Analysis, a Linear Mixed Model (LMM) was applied to the temporal metabolomics dataset of 26 HF Patients and identified 11 metabolites that were both significantly altered over time and highly associated to the clinical manifestations after MCSD implantation. Among these metabolites, 4 acylcarnitines, 4 phosphotidylethanolamines, 1 phosphotidylcholine, 1 sphingomyelin, and 1 hormone were identified, which we consider as potential novel biomarkers with respect to clinical diagnosis, prognosis and gradation of disease severity. Particularly, the temporal patterns of testosterone were utilized to stratify the patients into 3 risk groups. In Group I, 8 deceased patients (P2, P3, P9, P16, P18, P21, P24, and P25) were mostly males (n=7) who constantly maintained a very low Testosterone level. In Group II, 13 patients (P6, P7, P8, P10, P11, P13, P14, P15, P17, P20, P22, P23, P26) were all males with good clinical manifestation. Notably, these patients stayed at low Testosterone levels shortly after implantation but exhibited a steady increase of Testosterone after Stage II. In Group III, 5 female patients (P1, P4, P5, P12, P19) shared similar low Testosterone levels while showing a promising clinical recovery.

Potential Biomarker Candidates Identified by Temporal Metabolomics Profiling and Computational Workflows.

When we applied our integrated computational workflow, including both Differential Expression Analysis and LMM, to the MCSD recipient cohort, a total of 11 metabolites were found significantly altered over time as well as highly associated with clinical manifestations (SOFA Scores). When applied to human datasets, LMM with Relatedness Matrix Correction provides notable different results compared to Simple Linear Regression Model. Therefore, QQ-plots were used to evaluate the performance of these two computational models. The LMM analyses using matrix correction generated more precise p-values than Simple Linear Regression Model without matrix correction. Among these metabolites, 4 Acylcarnitine, 4 Phosphotidylethanolamine, 1 Phosphotidylcholine, 1 Sphingomyelin, and 1 hormone were identified. These metabolites may offer significant predictive value to the patient's clinical manifestation and outcome. In particular, Testosterone level could classify patients into three groups (Figure 3.7). In Group I, 8 deceased patients (P2, P3, P9, P16, P18, P21, P24, and P25) were mostly males (n=7) who constantly maintained a very low Testosterone level throughout the entire study period post-MCSD. In Group II, 13 patients (P6, P7, P8, P10, P11, P13, P14, P15, P17, P20, P22, P23, P26) were all males who presented with good clinical manifestation according to their SOFA scores. Notably, these patients stayed at low Testosterone levels shortly after implantation but exhibited a steady increase of Testosterone after Stage I, while further increasing in Stage II & III. In Group III, 5 patients (P1, P4, P5, P12, P19) shared similar low Testosterone levels as patients in Group I while their SOFA scores showed a promising clinical recovery. Interestingly, these were all female patients. It appears that Testosterone plasma levels may have more accurate predication of clinical outcomes in male patients, potentially as a biomarker candidate in clinical classification.

3.V. Discussion.

Advanced HF is a multifactorial disease in which a patient's hemodynamics are severely compromised. MCSD has been proven to promptly restore blood circulation and end-organ perfusion, thus enhancing global metabolism. To fully comprehend the failing heart and cardiovascular systemic responses towards MCSD implantation, it is essential to characterize the dynamics of metabolic networks. Several successful clinical investigations identified plasma biomarkers for diagnosis of cancers and neurodegenerative Diseases (Duarte, Rocha, and Gil 2013; Jové et al. 2014; G. Wang et al. 2014; Xu et al. 2013). In the cardiovascular research field, the metabolic networks in the heart and their perturbations during cardiac disease progression have become an emerging research focus (Cheng et al. 2015; Hoefer et al. 2015; Stegemann et al. 2014). However, the lack of an effective workflow to systematically screen disease-relevant metabolites has limited the clinical application of metabolomics profiling. Within this pilot study, we first sought to develop a technical platform with computational tools that can adequately handle the volume and analysis of the mouse plasma metabolome in a temporal manner. Our established workflow was subsequently refined to characterize complex clinical datasets of human HF. Finally, we translated our findings to clinical diagnosis, therapeutic validation, and risk stratification.

3.V.A. Methodological Considerations of Animal Models and Human Studies.

Advanced HF is characterized by prolonged stress and inflammation caused by elevated levels of neurohormonal stimuli, including cortisol and catecholamines. Previous studies demonstrated that implantation of MCSDs may improve patient quality of life by decreasing clinical symptoms and reducing neurohormonal stimulation in patients, which is recognized as reverse remodeling in HF (Liem et al. 2014). Technically, it is challenging to simulate a reverse remodeling model in mice. Thus, we decided to utilize an ISO-induced hypertrophy model, in which cardiac function is compromised by constant neurohormonal overstimulation (opposite to reverse remodeling). In

addition, the pathological phenotypes of ISO treatment (e.g., HW/BW ratio) have been well-documented and can be used as guidance for experimental quality control. Six genetically distinct mouse strains were selected for their variable susceptibilities towards ISO treatment, which offers a spectrum of molecular responses to ISO-induced stress that mimics genetic variances among humans.

We explored the scope of our large-scale metabolic survey to determine the dynamic range and sensitivity for data acquisition over time. Theoretically, any plasma metabolite could be accurately quantified if we apply extra isolation/enrichment procedures. However, temporal metabolomics profiling needs to be balanced between the quantity of identifiable metabolites and the technical practicality. Accordingly, we applied 9 quantitative panels of plasma metabolites and eventually obtained the absolute abundances of 610 metabolites. They encompass 20 categories of biomolecules based on their similarities in chemical structure and biological functions. We considered this set of metabolites as the most accessible pool for human and mouse plasma profiling without additional enrichment procedures.

In contrast to animal models, in which sample collections follow a well-planned schedule, clinical human studies often face challenges that may prevent blood sampling at consistent intervals. This necessitates a certain degree of flexibility in the construction of datasets. Accordingly, we defined intervals to which a time window is assigned. In brief, smaller intervals of sample collection were conducted right after surgery, and larger intervals were used as the patient stabilized and continuously recovered. This enabled data-series from multiple patients to be compared in a temporal manner, while maintaining a high level of accuracy to reflect clinical manifestations.

Before committing to extensive temporal profiling studies, it was important to gain confidence in our technical design by confirming the plasma metabolomics stability under basal conditions. Thus, 12 Healthy Human Subjects were recruited and blood samples were collected individually at 5 consecutive time points. A differential expression analysis validated that our computational workflow is tolerant of the insignificant changes among 610 metabolites that may occur due to varying factors such as sample collection, diet, physical activity, etc (Figure 3.5.). This result provided the confidence for conducting the two-year enrollment of MCSD recipients for blood sample collections up to 14 weeks. Furthermore, this dataset of healthy humans could be used as a benchmark reference for plasma metabolomic homeostasis.

3.V.B. Conceptualization of Computational Workflow Modules.

Unlike animal models, human studies present major challenges due to human biochemical individuality. Variability in genetic assembly and metabolic homeostasis may prevent deduction of mechanistic insights, especially in the presence of comorbidities. Large, longitudinal clinical studies take years to conduct, and the scope of metabolite profiling is usually limited by financial costs. Studies with smaller cohorts provide enhanced possibility for “data-driven” biomarker discoveries. However, small numbers of human subjects restrains the statistic power for in-depth characterization using standard computational methods. We demonstrate in our study that temporal metabolomics profiling datasets with smaller cohorts, in conjunction with customized computational approaches, may offer the benefit of sensitivity in response to pathological stress or therapeutic regimens, enabling in-depth analysis of biomarker candidates. To date, the most widely-used computational analyses in metabolomics studies are Univariate Analysis (e.g., t-test or ANOVA), Multivariate Analysis (e.g., PCA or PLS-DA) and Cluster Analysis (e.g., hierarchical clustering or partitional clustering) (Xia et al. 2012, 2009, 2015), offering snapshot comparisons of static conditions (e.g., healthy vs. diseased). However, these

methods have not been integrated to characterize temporal datasets extracted from large-scale metabolomics profiling.

The plasma metabolome is a summation of global metabolism, containing metabolites whose plasma levels range from stable to vastly dynamic. In this study, paired t-tests were first utilized to process temporal data series and differentiate metabolites' responses towards stress or therapies over time. A pilot subset of 12 healthy humans was initiated to establish a preferable threshold. With an adjusted p-value less than 0.001, we only identified one metabolite, Cortisol (stress hormone), which was slightly above the significance threshold. The stringent threshold reduces false positive rates potentially caused by systematic error (e.g., lack of biological replicates). Using the same criteria, this approach enabled us to identify significantly altered metabolites from HF patients following MCSD implantation. Our benchmark dataset of Healthy Humans and HF Patients revealed not only the stability but also the dynamics of plasma metabolome under varying pathological conditions. Such references may serve as guidance for future clinical investigations.

Following the identification of significantly altered metabolites over time, we sought to recognize a strong association between metabolic dynamics and pathophysiological alterations. The Regression Analysis has been widely used to correlate metabolite concentration and clinical characteristics (Riesmeyer et al. 2012; Weinstock-Guttman et al. 2011; Adibi et al. 2009). However, few metabolomics studies have considered Random Effects in their statistical equation, in which each element of one given variable presents different (weighted) impact on the final result (Bonate 2013). In clinical studies, Random Effects need to be taken into account because factors such as age, gender, or comorbidities contribute unevenly to disease phenotypes. To address these Random Effects in our regression model, a Relatedness Matrix Correction was applied. Although common in genomics studies (H. M. Kang et al. 2010; Orozco et al. 2015; H. M. Kang et al. 2008), such matrix correction has not been employed in

metabolomics investigations. The immediate challenge is that SNPs are routinely used in genomics to generate the relatedness matrix, whereas such genetic background is not available for human subjects in our study. In addition, genetic relatedness may not completely translate into metabolic similarities. It is more logical to generate the matrix using original plasma metabolite concentrations, which reflects the intrinsic metabolic differences among individuals and should remain consistent throughout the entire temporal profiling study. Given that a fraction of metabolites are more dynamic, and a Relatedness Matrix will be universally applied to correct all data points for the entire time series, we decided to select metabolites with stable plasma levels to generate the matrix. Similar to the concept in proteomics, we term those metabolites as "housekeeping metabolites". Thus, we computed a pairwise Relatedness Matrix among patients according to their housekeeping metabolite abundances, and performed the Linear Mixed Model (LMM) with the Matrix Correction using a software package pyLMM (https://github.com/nickFurlotte/pylmm) (Orozco et al. 2015). In this study, a metabolite must fulfill the criteria for both the differential expression analyses and LMM to be considered as a biomarker candidate. However, those extremely stable metabolites also bear biological importance. The pool-size differences of metabolite plasma reservoirs may also indicate the individual's capability to handle stress and perturbation, offering new directions for further investigations.

3.V.C. Gradation of Cardiac Remodeling and HF Using Identified Metabolites as Biomarker Candidates.

In the ISO treatment mouse model, we identified four metabolites that are both differentially altered over time and highly associated with ISO-induced phenotypic alteration (i.e., HW/BW ratio). They belong to the biological class of Free Fatty Acids (FFAs), including C14:0 (Myristic acid), cis-C16:1w7 (Palmitoleic acid), cis-C18:1w9 (Oleic acid) and cis-C18:3w3 (α-Linolenic acid). Following the initiation of maladaptive remodeling, their plasma levels significantly

dropped to below 30% of baseline levels within 3 days, suggesting that elevated cardiac energy demand directly led to plasma FFA deficiency (Figure 3.4). The stratification of mouse strains with these 4 FFAs is in line with their susceptibility towards ISO. In particular, the plasma α-linolenic acid (ALA) levels in ISO-resilient strains were promptly restored after the first wave of ISO-induced stress, which maintained in a level as much ~2-fold as that of ISO-susceptible strains. In a recent study using a rat model with β-adrenergic overstimulation, ALA was found to prevent ISO-induced myocardial fibrosis and hypertrophy by preserving the Src-PI3K protective pathway also known as the β2AR pro-survival pathway (Folino et al. 2015). Similar studies confirmed that an ALA-enriched diet prevents myocardial damages in an independent animal model (Fiaccavento et al. 2006) and might be associates with cardioprotective effects in human (Djoussé et al. 2005; Singh et al. 1997). According to our results, the temporal profiles of plasma ALA can be used as a potential biomarker to differentiate ISO-resilient mouse strains from ISO-susceptible ones.

In this human HF study, our integrated computational workflow identified a total of 11 metabolites that were significantly altered over time and highly associated with the patient's clinical manifestation (i.e., SOFA Score). Among these metabolites, 4 acylcarnitines, 4 phosphotidylethanolamines, 1 phosphotidylcholine, 1 sphingomyelin, and 1 hormone were identified, which we consider as potential novel biomarkers with respect to clinical diagnosis, prognosis and gradation of disease severity. In addition to static differences in plasma levels, we have considered the temporal fold change (Figure 3.6.D) of potential metabolomics biomarkers as a predictive clinical parameter. For clinically established biomarkers such as B-type natriuretic peptide (BNP) and Troponin-I, their fold changes over time (in parallel to their absolute plasma levels) provide valuable clinical insights on the pathological state of the heart (Twerenbold et al. 2012). Hence, a 10 fold change of plasma Troponin-I over time indicates more severe cardiac ischemic damage compared to a 2-fold change. Similarly, our results

indicate that plasma metabolites such as Dopamine, Testosterone and several phospholipids exhibit a steady fold change after MCSD implantation, reflecting the temporal impacts of mechanical unloading of the heart.

Remarkably, the human plasma levels of the hormone Testosterone appear to have strong clinical predictive values. In 8 HF patients deceased after MCSD implantation, their plasma testosterone values remained consistently low over phase I, II and III (Group I in 3.7); concomitantly, their SOFA scores remained in high-risk ranges. In contrast, 13 HF patients who showed favorable clinical recoveries after MCSD implantation (illustrated by a consistently improving SOFA score over time) displayed a steady increase of Testosterone plasma levels starting in Phase II, and further increasing over Phase III (Group II in 3.7). Lastly, 5 HF patients who showed good clinical recoveries with improving SOFA scores post-MCSD demonstrated low plasma testosterone levels (Group III in Figure 3.7.). Interestingly, these 5 patients were all females, which explains the lower hormone values despite the good clinical manifestation. Accordingly, our findings on plasma testosterone levels suggest that monitoring plasma testosterone in male HF patients before and after therapeutic intervention may furnish accurate predictive value to the patient's clinical outcome. Low testosterone levels post-MCSD may predict complications and a poor clinical performance, whereas an increasing plasma value may indicate that the HF patient is responding well to MCSD. Notably, 3 patients (P3, P9, P16 in Group I) displayed good SOFA scores but still expired post-MCSD. However, these 3 patients exhibited consistently low Testosterone levels, suggesting that plasma Testosterone levels may remain predictive even when the SOFA score is ambiguous. Correspondingly, previous observations in patients who were critically ill with trauma, shock or sepsis in the intensive care unit indicated that hypotestosteronemia was correlated with a poorer clinical outcome (Almoosa et al. 2014; Nierman and Mechanick 1999). To our knowledge, by applying our technical platform and novel computational workflow, we are the first group to identify a correlation

between hypotestosteronemia in advanced HF patients before and after MCSD implantation and their clinical outcome, in parallel to other plasma metabolomics fingerprints that may be utilized as potential biomarkers.

3.V.D. Future Perspectives and Directions.

In the current pilot study, we developed a technical platform and computational workflow to characterize the temporal dynamics of the plasma metabolome. A mouse model of maladaptive remodeling was applied as well as a human model of advanced HF and reverse remodeling. Our results suggest that this novel workflow is adequate for identifying significant metabolic changes that are correlated to the cardiac phenotypes and clinical outcomes. To further validate and optimize our technological platform and computational workflow (i.e. defining it's sensitivity, specificity, and positive & negative predictive values), it is critical to expand the current study to large HF patient cohorts. Moreover, advanced HF is a complex heterogeneous disease with a multitude of underlying molecular mechanisms and different etiologies, which must be taken into account within our computational workflow when being applied to large patient cohorts. For instance, many comorbidities and clinical characteristics relevant to cardiovascular disease (e.g., lipid status, blood chemistry, organ (dys)function, medication panel), and etiological subgroups (e.g., ischemic- vs. non-ischemic heart disease) most likely impact the plasma metabolome fingerprints (Stegemann et al. 2014; S. H. Shah, Kraus, and Newgard 2012; T. J. Wang 2011). Thus, their interferences should be assessed in greater detail using large patient cohorts.

In contrast to the mouse model consisting of 6 homogeneous genetic strains, each patient in our study exhibits unique demographics and individual variability in lifestyle. Consequently, different ethnic backgrounds, environmental stimuli, personality types, dietary preferences and physical activities most likely impact their metabolomes in advanced HF pre- and post-therapy. To meet

the challenge of taking these individual variability's into account, it is a mandatory endeavor to extract information from multi-dimensional data sets; this includes combining conventional clinical data from electronic health records with "omics" data from high-throughput technologies and computational workflows (Collins and Varmus 2015). It is the combination of diverse datasets from large patient cohorts that will enable a better understanding of disease risks, complex disease mechanisms, and optimal therapies as well as suitable intervention points in advanced heart failure.

CHAPTER 4: CLOUD-BASED COMPUTATIONAL KNOWLEDGEBASE TO ANALYZE, ANNOTATE, AND INTEGRATE METABOLOMICS AND PROTEOMICS DATASETS.

4.I. Abstract.

The integrated cardiovascular proteomics and metabolomics knowledgebase project entails a computational pipeline to fully characterize the plasma metabolome amid the development of heart failure (HF), and to establish a cloud-based bioinformatics platform that enables analysis, annotation, and integration of proteomics and metabolomics datasets. This project aims to unravel the interplay of the cardiovascular proteome and metabolome, ultimately to deliver novel knowledge in cardiovascular medicine.

The progression of HF on the molecular level is determined by the complex interplay among a broad spectrum of biomolecules and cellular pathways, where both proteins and metabolites are key players. Identifying how various biomolecules contribute to the pathological progression of HF is a major goal in cardiovascular research for informing pathological mechanisms, indicating disease phenotypes, and identifying new therapeutic targets. However, one current bottleneck for realizing this goal is the lack of computational tools and algorithms that can effectively integrate large-scale multi-omics data, including clinical datasets.

Accordingly, in this Chapter, we will discuss the outcome of three specific projects:

- Project 1: Development of a simplified computational pipeline to characterize the plasma metabolome and application to two specific cardiovascular scenarios: first, metabolome datasets from a mouse model of HF (isoproterenol [ISO]-treated mice); and second, metabolome datasets from HF patients.
- Project 2: Development of bioinformatics strategies to integrate proteome datasets with metabolome datasets and application to datasets obtained from six genetic mouse strains.
- Project 3: Establishment of a cloud-based platform, MetProt, to streamline and empower the above bioinformatics strategies, as well as disseminate the pipeline to the cardiovascular community at-large.

4.II. Intorduction.

4.II.A. Omics Phenotyping of HF.

HF causes >300,000 deaths a year in the US, with 5.8 million Americans currently afflicted (Mancini and Colombo 2015; Braunwald 2015; Roger 2013; Goldberg et al. 2007). A major reason why this common disease is so intractable is its complex, multifactorial nature, where overall phenotypic traits and disease outcomes are the effect of multiple biological and environmental factors. Technologies and approaches used to characterize HF and develop therapeutics must be able to embrace the complexity of the data.

4.II.A.a. Systematic characterization of the cardiac proteome.

As molecules that perform most mechanical and biochemical functions in the heart, proteins provide critical information to systems biology studies that aim to uncover the mechanisms of disease susceptibility. Technologies now exist that can easily identify and quantify large numbers of proteins in a single experiment (Lotz et al. 2014; Mann 2006). These studies have advanced our understanding of the components constructing protein interaction networks (Orchard et al. 2014), the role of post-translational modifications (PTMs) in modulating cellular signals (Streng et al. 2013), the dynamics of individual protein half-life in response to perturbation (Lam et al. 2014), and the totality of proteins contributing to a given cellular process or organelle (Kim et al. 2014). One of the principal lessons from the proteomics revolution is that gene expression does not necessarily mirror protein expression; one example of this is following ISO challenge on the heart we see protein abundance change independently of mRNA (Lam et al. 2014).

4.II.A.b. Temporal profiling of the cardiac metabolome.

The continuous contraction of cardiomyocytes demands tremendous amounts of energy; thus, maintaining metabolic homeostasis is essential for heart function. Accumulating evidence

suggests that perturbation of cardiac metabolism plays an important role in the pathological progression of HF (Neubauer 2007). Comprehensive quantification of metabolite abundance in blood plasma, also known as plasma metabolomics profiling, provides mechanistic insights into the molecular alterations underlying HF (Cheng et al. 2015). The large-scale quantification of circulating metabolites is useful for clinical applications, which can help classify HF patients before and after optimal medical therapy and surgical interventions (Lloyd-Jones 2010). These metabolomic approaches can be established as important avenues for biomarker and drug-target development, as well as validation of therapeutic efficacy.

4.II.A.c. Multi-omics phenotyping of HF.

Along with advances in high-throughput technologies, current studies have successfully integrated gene expression profiling with proteomics, and have facilitated our understanding of the pathophysiology and the molecular mechanisms of HF (Hou et al. 2015; Dos Remedios et al. 2003). However, there are no systematic approaches for the integration of large proteomic datasets with metabolome datasets due to the lack of strategies to connect proteins to metabolites and elucidate their relationships.

4.II.B. Computational Approaches to Identify Driver Molecules.

The availability of metabolomics data analysis tools is currently limited; for example, the Aztec discovery index lists 778 proteomics tools, but only 147 metabolomics tools. To our knowledge, there is no application that comprehensively analyzes and integrates proteomic and metabolomic datasets, and a simplified computational pipeline to analyze and annotate large-scale proteomics datasets with metabolomics datasets currently does not exist. This limitation hinders our capability to identify biomolecules relevant to HF.

4.II.B.a. Differential Expression Analysis with Paired t-Test.

Differential Expression Analysis is a univariate method widely used in omics studies for targeting molecules that are altered in a statistically significant manner (i.e., up-regulated or down-regulated) after a treatment or physical exercise (Xia et al. 2015, 2012, 2009). We employ a paired sample t-test, which is a statistical technique that is used to compare two population means in 'before-after' studies. Following the paired t-test, multiple testing correction to the p-values is applied using the Benjamini-Hochberg method to reduce false positives.

4.II.B.b. Principal Component Analysis.

Principal Component Analysis (PCA) is a well-established method for dimensional reduction of large datasets with many variables. This facilitates visualization, clustering, pattern recognition, and identification of the key variables that vary most significantly across a population. PCA can aid investigators in understanding complex, higher-dimensional datasets by projecting them into a 2- or 3-dimensional space, where they can be more easily visualized without sacrificing fidelity (Xia et al. 2015, 2012, 2009).

4.II.B.c. Unsupervised Hierarchical Clustering Analysis.

Unsupervised Hierarchical Clustering is a widely used data analysis method to build a binary tree from data by merging similar groups of points based on the calculated distances across samples or molecules. Usually, the distances are displayed as dendrograms (Xia et al. 2015, 2012; Xia and Wishart 2011b, [a] 2011; Xia et al. 2009).

4.II.B.d. Linear Mixed Model with Sample Relatedness Correction.

Linear Mixed Model (LMM) with Sample Relatedness Correction analyzes time-series omics data generated from heterogeneous samples to identify driver molecules that are highly associated with phenotypic characteristics. LMM correlates two variables and incorporates the

samples' relatedness, represented by omics expression or demographic factors, to improve statistical accuracy. We have applied LMM with Sample Relatedness Correction in our preliminary metabolomics study to find fingerprints of pathological stages during HF progression (Kirby et al. 2010; H. M. Kang et al. 2008).

4.II.B.c. Pathway Enrichment Analysis.

Significantly altered metabolites may be searched against several metabolomics databases, including Chemical Entities of Biological Interest (ChEBI, UK) (Hastings et al. 2013), Human Metabolome Database (HMDB, Canada) (Wishart et al. 2013, 2009, 2007), and LIPID Metabolites And Pathways Strategy (LIPID MAPS) (Fahy et al. 2009, 2007) to cross-reference identifiers that are commonly used by pathway enrichment search engines and databases. The pathways enriched with identified metabolites may be further annotated with resources including MetaboLights (Haug et al. 2013) and Reactome (EMBL-EBI, UK) (Fabregat et al. 2018, 2016; Milacic et al. 2012).

4.II.C. Cloud-based Computational Knowledgebase.

To extract biological meaning from multi-omics datasets, researchers rely heavily on computational resources that analyze and annotate the molecules of interest with known information. Currently, access to tools/annotations is not straightforward because they reside in fragmented and incomplete repositories. To address this challenge, we are creating a novel distributed query system and cloud-based infrastructure, MetProt, that is capable of providing unified access to protein and metabolite datasets, allowing users to submit a single query to access multiple resources including Reactome (Fabregat et al. 2018, 2016; Milacic et al. 2012), UniProt (UniProt Consortium 2015), MetaboLights (Haug et al. 2013), BioGPS (Wu et al. 2016; Wu, Macleod, and Su 2013), Gene Wiki (Tsueng et al. 2016), and COPaKB (N. C. Zong et al.

2013; H. Li et al. 2013). We will also engineer customized application programming interfaces (APIs) to provide direct access for information in MetProt.

4.II.C.a. COPaKB.

COPaKB (N. Zong et al. 2014; N. C. Zong et al. 2013; H. Li et al. 2013) (http://heartproteome.org) is an omics analysis platform with two key components: (1) a peptide spectral search engine; and (2) spectral library modules for knowledge annotation.

4.II.C.b. Reactome.

The Reactome (Croft et al. 2014; Milacic et al. 2012) (http://reactome.org) platform is a suite of network analysis tools for performing topology analysis and over-representation analysis of gene/protein networks. It comprises a manually curated database of human pathways with views of protein and metabolite structures overlaid with expression data. We will interface with Reactome to identify key pathways of interest in our omics data.

4.II.C.c. MetaboLights.

MetaboLights (http://ebi.ac.uk/metabolights/) is a public database dedicated to the submission and sharing of metabolomics data, mass spectra, annotated biological roles, and other derived information (Haug et al. 2013). Based on spectral similarities and chemical structures, we will employ its search services to analyze and interpret metabolite data we collect from the studies.

4.II.C.d. BioGPS/Gene Wiki:

BioGPS (http://biogps.org) is an interface for omics research (Wu et al. 2016). It provides a user-customizable portal with aggregated information on protein annotations and target list analyses. Gene Wiki translates molecular information into structured knowledge; relevant pages representing molecular transducers will be aggregated and recruited into MetProt.

4.III. Results.

4.III.A. Multi-Omic Data Acquisition from Public Repositories.

We have at hand large human and mouse plasma metabolomics datasets that are clinically relevant to advanced HF. The datasets include multiple reaction monitoring (MRM) mass spectromertry (MS)-based absolute quantification measurements of 610 distinct plasma metabolites (including lipids, steroids, amino acids, and energy metabolism metabolites) from 26 advanced HF patients and 12 healthy controls recruited at UCLA. In parallel, we established a systems genetics model of cardiac hypertrophy in which the metabolomics of six genetic mouse strains were measured. In total, the datasets comprise 21,960 measurements of absolute concentrations in mouse plasma, and 147,620 measurements of absolute concentrations in human plasma. In addition, a large dataset of protein dynamics data, including the abundance of 8,064 cardiac proteins and turnover rates of 3,228 cardiac proteins in the normal and hypertrophic hearts of six mouse strains, was acquired from ProteomeXchange (PXD002870) and Sage Synapse (syn2289125). This dataset contains over 120,000 protein quantification data points and covers over 200 cellular pathways, and used to support development of a strategy to integrate protein and metabolite data as detailed in Project 2.

4.III.B. Simplified computational pipeline for triaging and prioritizing molecular markers.

We have established a computational pipeline to identify biomarker candidates that are associated with both phenotypic alterations and temporal dynamics following medical therapy and surgical intervention. In our pipeline, we explored the analytical power of two major components, a Differential Expression Analysis and a Linear Mixed Model. To account for the individual relatedness, we incorporated the Relatedness Matrix Correction into the Linear Mixed Model by exploiting the “housekeeping metabolites”, which remain at constant plasma levels after mechanical circulatory support device (MCSD) implantation. With these preliminary

analyses, we successfully identified a number of plasma metabolites that are altered in cardiac remodeling.

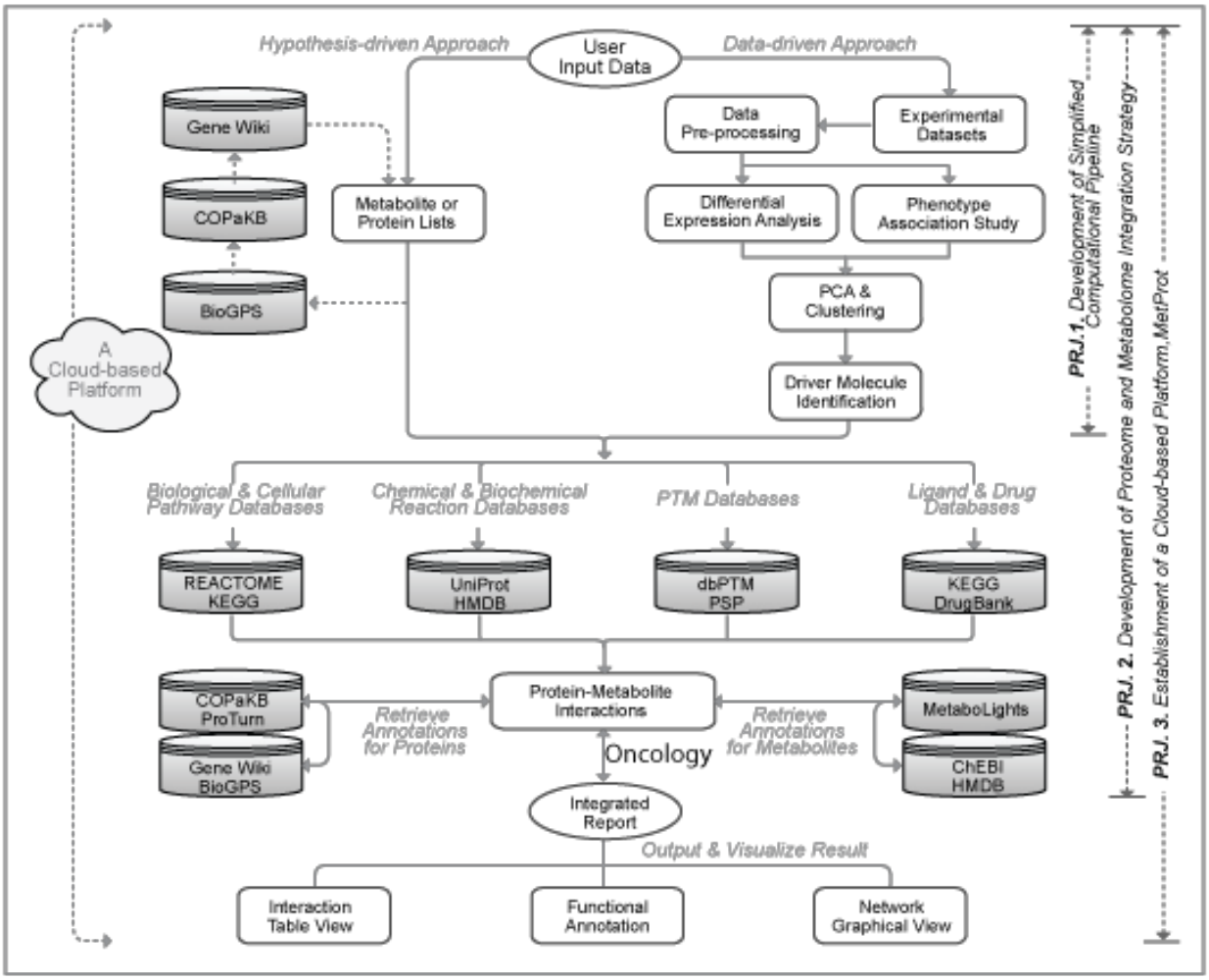


Figure 4.1. MetProt Integrated Workflow. This workflow diagram depicts the individual elements and logical connections in the proposed bioinformatics platform for prioritizing and analyzing data, finding connections between proteins and metabolites, and annotating molecular functions. User data will be submitted in two major categories: 1) lists of protein and/or metabolite identifiers from hypothesis-driven research with biological information enriched by Gene Wiki, COPaKB, and/or BioGPS, or 2) large-scale molecular data from data-driven investigations with their quantitative values and phenotypic information. For the latter category, MetProt will review and analyze the data to identify driver molecules that are significantly altered over time and highly associated with phenotype (Project 1). The molecules of interest are then annotated using information fetched from Reactome/KEGG (biological pathways), Uniprot/HMDB (chemical reactions), dbPTM/PhosphoSitePlus (PTM data), and KEGG LIGAND/DrugBank (ligand & drug information). The extracted information are now integrated to construct interaction graphs among proteins and metabolites, with additional annotations from COPaKB/Gene Wiki and MetaboLights/CheBI (Project 2). Collectively, we will engineer a cloud-based platform to enable these analyses on the cloud and to generate integrated reports, including table views of protein-metabolite interactions, functional annotations, and graphical views as output (Project 3). Detailed platform system architecture is shown in Figure 4.3.

4.III.B.a. Develop a simplified computational pipeline to characterize plasma metabolomics data. Plasma metabolomics profiling has great potential to identify biomarker candidates as internal indicators of diseases or as novel therapeutic targets. However, its clinical applications are challenged by the complex nature of patient cohorts, hospital protocols, and policy. Subject individuality, affected genetic variability, diverse metabolic homeostasis, and interference of comorbidities can further complicate accurate quantification of metabolites. To overcome these challenges, large-scale clinical cohorts have been assembled to provide sufficient statistical power, though the studies were limited to a “hypothesis-driven” metabolomics approach. Alternatively, temporal datasets using small clinical cohorts can be established as a profiling approach to pursue “data-driven” metabolomics studies. However, an integrated computational workflow simplified for clinical metabolomics datasets has never been fully explored. We first implement a data pre-processing component, which replaces the Limit of Detection (LOD) values with half of the minimum value in the original data and remove any metabolites missing more than 50% of the quantification values. Second, Differential Expression Analysis will be performed on the filtered dataset with a stringent threshold of adjusted p-value < 0.001 and ±1.5-fold change to detect metabolites with significant changes. To depict the results of Differential Expression Analysis, volcano plots and heatmaps will be automatically generated. Next, to identify metabolites that are highly associated with the phenotypic characteristics (e.g., HW/BW or SOFA Score), we first account for the individual relatedness by incorporating the relatedness matrix into a linear mixed model (LMM). The relatedness matrix will be generated using “housekeeping metabolites”, which are metabolites that maintain a consistent level of plasma concentration over time. Differences in plasma concentration exist among individuals, which can indicate individual relatedness. Housekeeping metabolites will be defined using a threshold of adjusted p-value > 0.1 and subsequently used to create the relatedness matrix. The LMM employs time series metabolite concentrations along with phenotypic values, and it adjusts the regression based on their relatedness matrix. We perform the multiple testing correction,

using the Benjamini-Hochberg method, and apply a threshold of FDR < 0.01 to identify metabolites highly correlated with phenotypic alterations. The metabolites that are detected by both Differential Expression Analysis and Linear Mixed Model with relatedness correction are considered driver molecules. We also use Principal Component Analysis (PCA) to verify that the driver molecule expression pattern has sufficient power to stratify patients with distinct outcomes (Figure 4.1.). Users upload metabolite identification and quantification datasets with phenotypic information and matching time points in .tsv or .csv file formats, using file transfer and management protocols we have previously established in our knowledgebase, COPaKB. Output contains: (i) a list of metabolites that are significantly altered over time; (ii) a list of metabolites that are highly associated with the phenotypic characteristics; (iii) a list of metabolites commonly detected in both (i) and (ii); and (iv) p-values, FDRs, and metabolite annotations for each detected metabolite (Figure 4.1.).

We utilize a mouse model of maladaptive cardiac remodeling via β-adrenergic overstimulation to validate the computational pipeline with stringent thresholds. To investigate whether the pipeline is applicable to different organisms, we subsequently apply our integrated computational pipeline to a clinical temporal dataset of HF patients undergoing MCSD implantation.

4.III.B.b. Develop bioinformatics strategies to integrate proteome datasets with metabolome datasets.

The proteome and the metabolome are both intermediate phenotypes that are tightly connected to the biochemical functional output of a system. Both proteomics and metabolomics methods are now being increasingly deployed to identify disease markers and mechanisms. There are intrinsic connections between proteins and metabolites: (i) proteins in major metabolic pathways function to synthesize and degrade metabolites, (ii) metabolites in turn can modulate protein activities via allosteric interactions, and (iii) metabolomics and proteomics experiments share mass spectrometry (MS) techniques. Unfortunately, current efforts to integrate these two types

of molecules are limited. Our goal is two-fold: (i) first, to identify the biological interface of proteins and metabolites; and (ii) second, to develop a bioinformatics strategy to integrate proteome and metabolome datasets. Therefore, we enable users with either hypothesis-driven or data-driven approaches to annotate and enrich their data, and to gain biomedical insights with metabolite-level information (Figure 4.3.). We start by enumerating various direct linkages via which metabolite information may be relevant to data interpretation and hypothesis generation from proteomics data. We then lay out a computational strategy and software tool to integrate multiple data types, drawing broadly from existing databases on protein and metabolite information. The utility, applicability, and scalability of our approach are rigorously validated using large-scale proteomics and metabolomics data on six distinct genetic mouse strains and perturbations that we have already acquired.

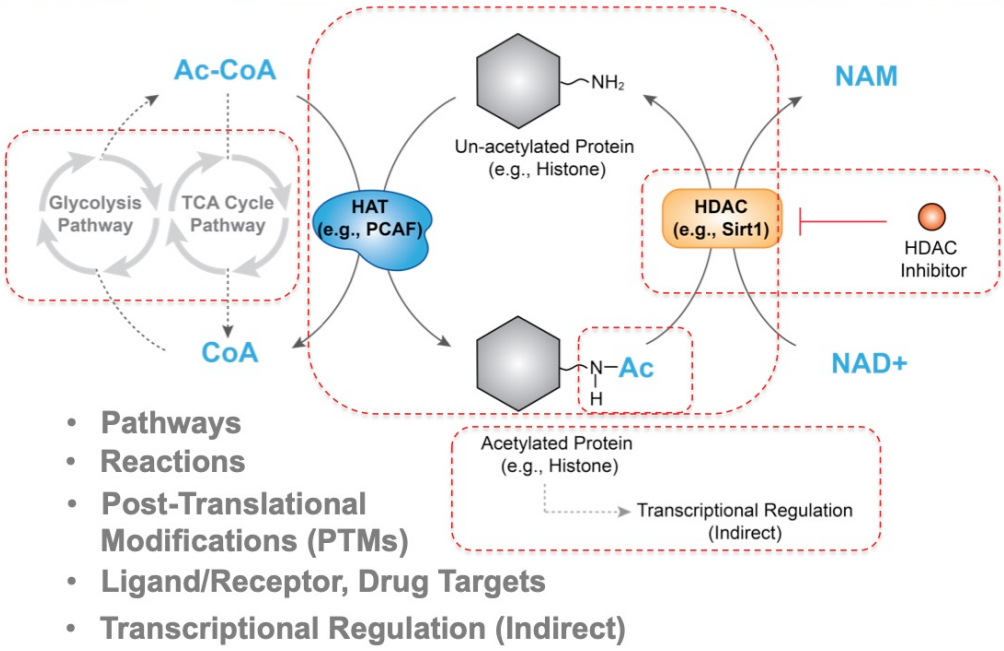


Figure 4.2. Intersections where metabolome meets proteome. meet proteins. Protein PTMs are critically modulated by the presence and concentration of particular metabolites. With Histone Acetyltransferase (HAT), e.g., enzyme PCAF (lysine) acetyltransferase, an Acetyl group from metabolite, Acetyl-CoA, is transferred to histone protein, forming the Acetylated protein. The Acetylation is considered as a PTM.

We identify five important relationships between proteins and metabolites essential to the annotation of a dataset in biomedical investigations: (i) given a list of quantified proteins, we find the enriched biological and cellular pathways from Reactome and KEGG, and we retrieve a list of metabolites that participate in these pathways; (ii) given a protein that catalyzes an enzymatic reaction, we retrieve its natural biochemical substrates and products from UniProt, HMDB, Reactome, and KEGG; (iii) given a protein known to participate in a particular type of PTM, we retrieve metabolites known to critically induce, suppress, or otherwise participate in these modifications (e.g., NAD+ to acetylation and ADP-ribosylation) from dbPTM; (iv) given a protein that is a receptor or is allosterically regulated by a known metabolite, we retrieve its natural agonist and/or allosteric regulator (e.g., from KEGG LIGAND database); and (v) given a protein with known transcriptional regulation, we retrieve metabolites that are known to induce or suppress its transcription (e.g., oxysterols and liver X receptor downstream targets) (Figure 4.2 and 4.3).

To enable these functionalities, we begin by building broad interfaces with large-scale databases of proteomics and metabolomics data. Currently, we are at pace to complete a dedicated web API to retrieve protein cardiovascular disease relevance information from COPaKB. Data from Reactome, MetaboLights, and UniProt will be retrieved through existing and to-be-developed high-performance APIs. We develop a scoring schema to weigh retrieved protein-metabolite connections from each category based on the strength of existing evidence. As the data retrieval and integration strategies mature, we demonstrate the strategy's utility in the analysis of a large protein and metabolite dataset of cardiac hypertrophy that we have already acquired (see Preliminary Data). We construct protein-metabolite graphs based on the retrieved connections, followed by visualization of the integrated protein-metabolite networks on the browser.

In addition, biological and chemical functional annotations for proteins and metabolites are retrieved from COPaKB/Gene Wiki and MetaboLights/ChEBI. MetProt will be hosted on Amazon Web Services (AWS), enabling users to perform these analyses on the cloud and generate integrated reports, including a table view of protein-metabolite interactions, and functional annotations as well as graphical views of networks and pathways.

Users input a list of proteins with Uniprot accession numbers and columns of numerical qualities in text formats. Output will be annotations for each Uniprot accession in each of the aforementioned classes of protein-metabolite linkage (Figure 4.3.). We use the proteomics and metabolomics data from the mouse hypertrophy model as a test case to determine whether the knowledge integration approach between protein and metabolite data in a common disease model can be used to discover new pathways and generate new hypotheses (Project 3).

4.III.C. Build a cloud-based platform, MetProt, to enable the computational/bioinformatics strategies.

Cloud-based platforms have become an increasingly popular mode of large-scale high-dimensional omics datasets (big data) analysis and dissemination of biomedical knowledgebases. The many operational advantages of cloud-based knowledgebases include minimal requirements of end-user investment in computational infrastructure, strong software compatibility, lack of installation hassle, and high visibility on the web, all working to ensure maximal use and access of the tools. Successful examples include UniProt and Reactome; both currently support broad user engagement in global biomedical research. Accordingly, we have chosen to establish a cloud-based platform, tentatively named MetProt, as an approach enabling the above computational and bioinformatics strategies (Projects 1 & 2). This cloud-based platform connects to collaborating partners and knowledgebases to promote interoperability, utilization, and dissemination.

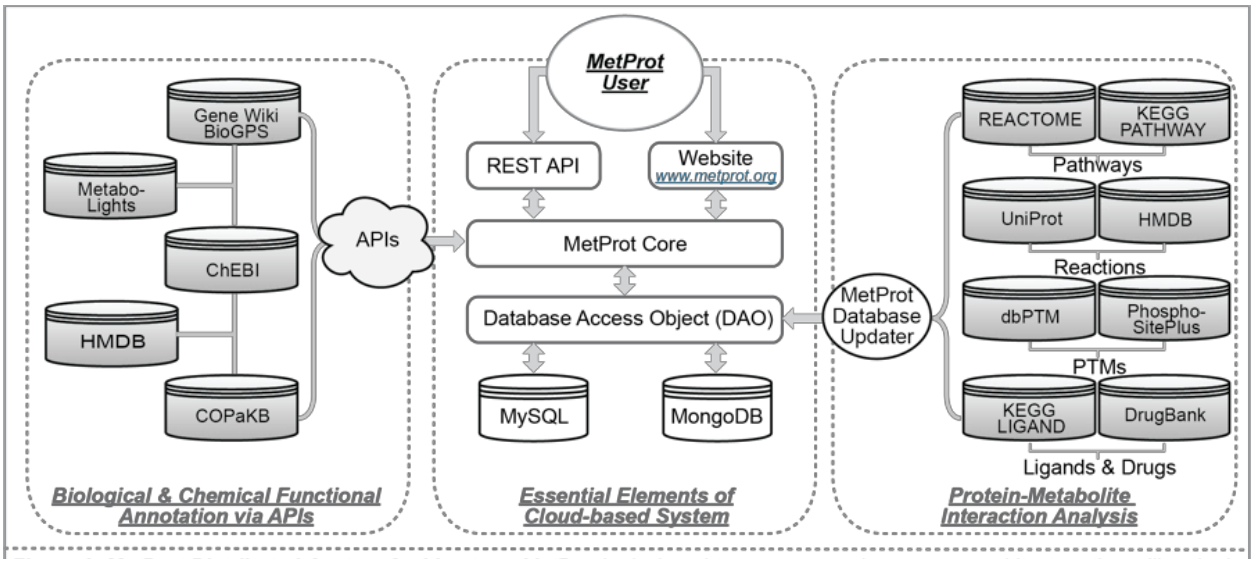


Figure 4.3. MetProt Distributed System Architecture. MetProt is designed to run on a modern system architecture that will make it robust, versatile, and extensible. An Amazon Web Service (AWS) EC2 server will host the MetProt Core. Users will programmatically access and retrieve search results via REST APIs or the web interface on www.metprot.org. MetProt Database Updater will integrate various types of biomolecular interaction information extracted from REACTOME, KEGG PATHWAY, UniProt, HMDB, dbPTM, PhosphoSitePlus, KEGG LIGAND, and DrugBank, and will store the comprehensive interaction information among proteins and metabolites in MetProt Databases. In addition, MetProt will provide biological and chemical functional annotation for biomolecules by utilizing APIs provided by Gene Wiki, BioGPS, MetaboLights, HMDB, ChEBI, and COPaKB.

Our team previously developed COPaKB, a cardiovascular knowledgebase for cloud-based proteomics analysis. COPaKB enables users to upload and analyze MS data files remotely on a highly scalable cloud server (N. C. Zong et al. 2013; H. Li et al. 2013) and has received over 177,604 page views from ~125 countries as of June 2016. The spectral library search engine on COPaKB is an emerging alternative to sequence database search, and allows fast identification of peptides with low error rates. In addition, API integration has enabled COPaKB to feature a pathway enrichment analysis tool, a Gene Wiki embedding widget, and a protein-protein interaction and visualization tool. These efforts demonstrate our credentials in developing and deploying widely-utilized, community-centric data science platforms.

MetProt system architecture includes: (i) a Wiki interface, which serves as evolving documentation and a canvas for users to document protein and metabolite annotations; (ii) a MetProt Core, which receive and process user queries and connect to Gene Wiki, BioGPS,

MetaboLights, COPaKB, HMDB, and ChEBI via APIs provided by these knowledgebases; (iii) a Database Access Object (DAO), which interfaces with MetProt Database Updater, retrieving and integrating information from collaborating knowledgebases (i.e., Reactome, KEGG, UniProt, etc.); and (iv) MetProt will build upon two complimentary database systems (MySQL and MongoDB) and ensure efficient search/query operations (Figure 4.3.).

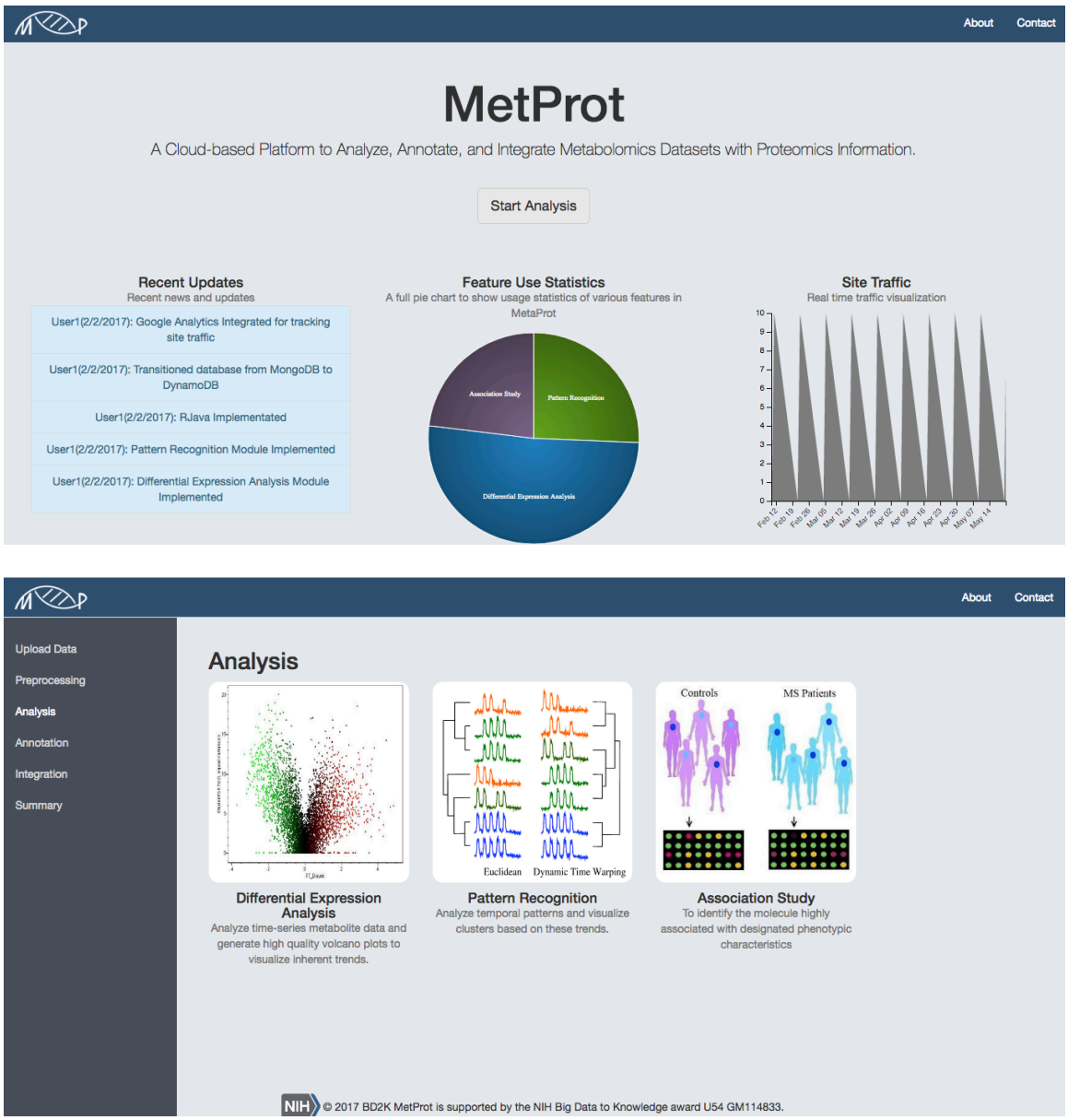


Figure 4.4. MetProt User Interface. Technologies used in MetProt user interface include HTML/CSS, JavaScript, and Bootstrap. Back-end/server is written in Java and R programming languages. The MetProt can be accessed at http://metaprot-env.us-west-2.elasticbeanstalk.com/.

We anticipate users upload protein identification and quantification datasets, using file transfer and management protocols we have established from our experience with COPaKB. Output contains the list of proteins that are linked to known metabolites and metabolite annotations, and the manner in which such conceptual or molecular connections are established. For example, users are able to identify the number of proteins from the uploaded list that participate in major metabolic pathways and visualize the proteins that are driving chemical and biochemical reactions, as well as proteins whose protein PTMs may be critically modulated by the presence and concentration of particular metabolites. These connections serve to identify the potential involvement of various metabolite species in the biological model being studied, and will be systematically organized and presented to facilitate data interpretation, knowledge integration, and hypothesis generation. Additional features will be developed (which will extend beyond my dissertation), where we will support metabolite datasets as input and allow reverse queries of protein-level regulation of metabolite levels.

MetProt is evaluated by the number of installed servers and the volume of depositions in the MetProt database at the end of the project. In addition, the user access statistics of MetProt, including number of unique users, pageviews, and pageviews per visit, are measured by Google Analytics (Figure 4.4.). We will also continuously monitor the usage of our MetProt API through Google Analytics.

4.IV. Discussion.

4.IV.A. Significance and Innovation of Completed MetPro Projects.

The proposed three projects have two major aspects of significance. First, our studies will establish a computational pipeline, applying it to two benchmark plasma metabolomic datasets of ISO-treated mouse strains and HF patients, to pave the groundwork for future metabolomics

studies with clinical translation. Protein and metabolite identities/quantities are tightly regulated parameters reflecting cardiac function.

- Conceptual innovation: We took advantage of a novel systems genetics model, in which natural mouse genetic strains span a wide spectrum of HF susceptibility, to discover disease mechanisms.
- Technological innovation: Proteomics and metabolomics investigations are rarely integrated, despite the obvious connections between proteins and metabolites that are critical for fully understanding biological mechanisms. Our studies innovate approaches that enable multi-omics investigations and empower the broad user base of cardiovascular investigators.
- Translational innovation: The proposed studies directly support experiments aimed at understanding in vivo protein and metabolite dynamics of HF patients.

4.IV.B. Feasibility of Integrating MetPro with Other Existing Resources.

4.IV.B.a. Aztec search results for other available resources that conduct multi-omics analysis and integration (http://aztec.bio/).

The availability of multi-omics data analysis tools is currently limited; for example, the Aztec discovery index lists 778 proteomics tools and 147 metabolomics tools. However, Aztec resulted in only one multi-omics visualization tool, and to the best of our knowledge, there is no application that comprehensively analyzes and integrates proteomic and metabolomic datasets. A simplified computational pipeline to analyze and annotate large-scale proteomics datasets with metabolomics datasets currently does not exist. This limitation hinders our capability to identify integrated biomolecular profiles relevant to HF.

4.V.B.b. Function, application and accuracy of the existing/alternative methods.

Previous studies have systematically examined the interplay between proteomes and metabolomes and reported that approximately 20% of proteins bind to at least one hydrophobic metabolite (X. Li et al. 2010; X. Li and Snyder 2011), suggesting that a large portion of the proteome physically interfaces with metabolites. However, the resources for the true integration of proteins and metabolites remain limited. One notable effort that successfully tackled this topic is the development of “Search Tool for Interacting Chemicals” (STITCH), which is a database integrating the disparate data sources into a singular resource (Szklarczyk et al. 2016; Kuhn et al. 2014, 2008).

STITCH facilitates access to the information of protein-metabolite interactions that can be extracted from the databases of metabolic pathways, crystal structures, binding experiments, and drug-target relationships. Further, this database provides a comprehensive list as well as visualization of protein-metabolite interactions. Specifically, STITCH enables users to query the interacting partners of the list of proteins or metabolites of interest. A full-text search is available for identifiers (e.g., STRING identifier) and common names of chemicals and proteins. The chemical structures and protein sequences can also be submitted for a database search. In addition, STITCH contains a visualization tool that displays the search results in a network view, providing an overview of the complex networks among proteins and metabolites and possibly biological insights. Further, STITCH allows users to filter out the molecules in the results that are believed not to be present in a specified tissue. This filtering capability reduces the false positives in the protein-metabolite network. STITCH aggregates information from multiple sources into a unified global network of protein–metabolite interactions. It is widely applicable to targeted small-scale analyses via web interface to large-scale analyses via their API access.

Due to its powerful capacity for bridging metabolic pathways, protein 3D structures, protein-

metabolite binding data, and drug-target relations, STITCH enables synergistic data sharing with MetProt. This resource offers granular details for biological and cellular pathways, describing chemical and biochemical reactions at each step. Moreover, it provides critical information for metabolites when they function either as ligands for protein-based receptors or cofactors for translational and post-translational regulations. This effort affords opportunities for comprehensive knowledge pertaining to interfaces between the proteome and metabolome; furthermore, it facilitates an in-depth understanding of molecular and cellular function underlying disease initiation and progression.

4.V.B.c. Potential limitations and alternative strategies.

The proposed platform will incorporate information from multiple resources. Some resources provide APIs allowing us to programmatically retrieve information; some provide only spreadsheets without structured data. Our approach is to retrieve programmatically where possible and manually where APIs are not available. Hence, a potential pitfall is the heterogeneity and incompatibility of vocabularies from these different data sources; for example, KEGG and Reactome may have different terminologies for some identical pathways. We will apply community standards to particular resources and either request the resources, adopt a consistent community standard, or alternatively, host the converted information on our local server, with periodical updates.

4.V.C. Future Direction of MetProt Development.

4.V.C.a. Differentiate physiological relevant protein-metabolite relations from “enzymatic promiscuity” (“Comprehensive Natural Products II” 2010).

For an enzyme, its “native function” is physiologically relevant to the organism hosting this enzyme. Many enzymes have more than one physiological function, or exhibit a broad

specificity towards a whole range of similar substrates, which still belong to the category of naturally evolved “native function”. However, purely accidental protein-metabolite interactions occur, which comprise the idea of enzymatic promiscuity. When a high-throughput, unbiased proteomics dataset has been established, we need to differentiate the physiologically relevant associations from non-specific ones. The traditional binding assay-based method is labor-intensive and not practical for validating hundreds of thousands of relations within a global proteome-metabolome network. Therefore, computational approaches, in particular, machine learning (ML)-based strategies may be incorporated into the current infrastructure of MetPro.

4.V.C.b. Analysis of protein-metabolite regulatory events that are relevant to a particular phenotype.

Large-scale characterization of metabolomes has gained traction over the last decade, and as with proteomics, advanced MS techniques can now be used to detect and quantify thousands of chemical compounds in a single run. Data analysis pipelines are not quite as well developed for metabolomics as for standard proteomics data, and since the structures of many metabolites cannot be readily distinguished above a certain mass tolerance, rigorous validation criteria are required for confident identification. While the main bottleneck in computational analysis remains the lack of annotation for many uncharacterized metabolites, serendipitously the global study of metabolite–protein interactions will likely provide an elegant solution to address this (Guo, Peng, and Emili 2017). To this end, a major goal of MetProt moving forward is to systematically address the metabolite annotation conundrum.

It is estimated that there are >1 million metabolites in a biological cell, and the number of metabolites is approximately 100-fold larger than proteins (Milo 2013; Bennett et al. 2009). However, the interaction between the proteome and metabolome has been difficult to

systematically define on a large-scale. While bioinformatics analysis of proteomes and metabolomes has traditionally been performed separately, integrative data analysis strategies are urgently needed to obtain a deeper mechanistic understanding of biological processes and complex protein-metabolite networks. From a technical standpoint, the allosteric, or noncovalent nature of the binding between proteins and metabolites has hampered the development of methods for systematically mapping protein-metabolite interactions. MS-based approaches are largely restricted to metabolite classes (e.g., lipids), and while NMR spectroscopy shows promise for defining sophisticated interaction mechanisms (Nikolaev et al. 2016), it has lower throughput than MS. It has also become clear in recent work that some PTMs, such as acetylation and succinylation, can occur non-enzymatically in vitro through inherently reactive metabolites (Weinert et al. 2013; Wagner and Payne 2013). The identification of PTM sites or allosteric interactions which are highly conserved across species can help to select regulatory events which are more likely to be functionally relevant. The next step is to identify those regulatory events that are actually relevant to a particular phenotype (Kochanowski, Sauer, and Noor 2015), including efforts to address how post-translational regulatory events relate to other physiological measures, such as fluxes or metabolite concentrations.

A study in 2010 conducted in eukaryotic yeast cells which looked at purified proteins and their association with metabolites revealed that approximately 70% of ergosterol biosynthetic proteins and 20% of the protein kinases studied bound hydrophobic metabolites (X. Li et al. 2010). Extrapolating from this results, the authors speculated that >1200 soluble yeast proteins bind hydrophobic molecules in vivo. A recent study in 2018 achieved a comprehensive, systematic analysis of the bacterial proteome-metabolite interactome via the Limited Proteolysis-small molecule mapping (LiP-SMap) technique (Piazza et al. 2018). They looked at the binding of over 2,500 proteins with 20 common metabolites and demonstrated that ~25% of the proteome interacted with at least one of the evaluated metabolites, with preference exhibited for the “core

proteome" (Yang et al. 2015). Moreover, the fraction of the proteome that underwent structural variations upon metabolite binding was approximately one-third of the total proteome (620 proteins out of 2,565). This study indicates that the size and complexity of the metabolite-protein interactome is substantially larger than previously anticipated. Currently, MetProt contains xxx number of protein-metabolite interactions from xxx protein and xxx metabolites for xxx organisms. Despite the sizable nature of our database, this likely represents only a portion of biologically relevant interactions.

4.V.C.c. Using advance ML-based methods to predict protein-metabolite regulatory events that are relevant to disease phenotypes.

The increasingly complex protein-metabolite datasets that have been generated, such as those listed above, have spurred the development of powerful new computational capabilities for integrating proteomic and metabolomic datasets. An example of this is deep learning, which allows for deeper mechanistic inferences of biomedical significance, as well as better diagnostics and drugs (Guo, Peng, and Emili 2017). Several computational tools have been developed to predict the protein targets of small molecules (Koutsoukas et al. 2011), such as GUSAR, PASS INet, PharmMapper, and TarFisDock. However, these approaches and others used in recent studies are univariate and don't fully take an advantage of using big data. Advanced ML approaches to tackle this topic may be suitable, and we have plans to employ ML-based prediction methods to predict highly possible interactions between proteins and metabolites. In this project a cloud-based platform is developed to aggregate knowledge of protein-metabolite interfaces from multiple sources. We can leverage the massive datasets in MetProt as training data to predict other interactions, and use existing text-mining tools developed by Dr. Jiawei Han and Dr. Peipei Ping to validate the predicted interactions reported in both current and future biomedical literature. Moreover, rapid progress in the drug discovery

field is expected in the next few years, resulting in much needed new opportunities for improving sparse drug discovery pipelines. Our future plans for MetProt include using deep learning to find the associations between the protein-small molecule interactions and disease phenotypes, among others.

4.V. Conclusion.

We are establishing a cloud-based platform, MetProt (http://metaprot-env.us-west-2.elasticbeanstalk.com/), for quantifying, triaging and analyzing omics datasets. MetProt will enable us to find existing as well as novel connections between proteins and metabolites, annotate molecular functions, and provide biomedical insights.

REFERENCES

Adibi, Jennifer J., Russ Hauser, Paige L. Williams, Robin M. Whyatt, Antonia M. Calafat, Heather Nelson, Robert Herrick, and Shanna H. Swan. 2009. "Maternal Urinary Metabolites of Di-(2-Ethylhexyl) Phthalate in Relation to the Timing of Labor in a US Multicenter Pregnancy Cohort Study." *American Journal of Epidemiology* 169 (8): 1015–24.

Alcock, Lisa J., Michael V. Perkins, and Justin M. Chalker. 2018. "Chemical Methods for Mapping Cysteine Oxidation." *Chemical Society Reviews* 47 (1): 231–68.

Almoosa, Khalid F., Aditya Gupta, Claudia Pedroza, and Nelson B. Watts. 2014. "Low Testosterone Levels Are Frequent in Patients with Acute Respiratory Failure and Are Associated with Poor Outcomes." *Endocrine Practice: Official Journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists* 20 (10): 1057–63.

Andersson, Daniel C., Jérémy Fauconnier, Takashi Yamada, Alain Lacampagne, Shi-Jin Zhang, Abram Katz, and Håkan Westerblad. 2011. "Mitochondrial Production of Reactive Oxygen Species Contributes to the β-Adrenergic Stimulation of Mouse Cardiomycytes." *The Journal of Physiology* 589 (Pt 7): 1791–1801.

Arakaki, Adrian K., Jeffrey Skolnick, and John F. McDonald. 2008. "Marker Metabolites Can Be Therapeutic Targets as Well." *Nature* 456 (7221): 443.

Arakawa, Takatoshi, Yoshiaki Kawano, Yoko Katayama, Hiroshi Nakayama, Naoshi Dohmae, Masafumi Yohda, and Masafumi Odaka. 2009. "Structural Basis for Catalytic Activation of Thiocyanate Hydrolase Involving Metal-Ligated Cysteine Modification." *Journal of the American Chemical Society* 131 (41): 14838–43.

Armenian, Saro H., Sarah K. Gelehrter, Tabitha Vase, Rajkumar Venkatramani, Wendy Landier, Karla D. Wilson, Claudia Herrera, et al. 2014. "Carnitine and Cardiac Dysfunction in Childhood Cancer Survivors Treated with Anthracyclines." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 23 (6): 1109–14.

Bak, D. W., and E. Weerapana. 2015. "Cysteine-Mediated Redox Signalling in the Mitochondria." *Molecular bioSystems* 11 (3): 678–97.

Bechtel, Tyler J., and Eranthie Weerapana. 2017. "From Structure to Redox: The Diverse Functional Roles of Disulfides and Implications in Disease." *Proteomics* 17 (6). https://doi.org/10.1002/pmic.201600391.

Benjamin, Emelia J., Salim S. Virani, Clifton W. Callaway, Alanna M. Chamberlain, Alexander R. Chang, Susan Cheng, Stephanie E. Chiuve, et al. 2018. "Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association." *Circulation* 137 (12): e67–492.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 57 (1): 289–300.

Bennett, Bryson D., Elizabeth H. Kimball, Melissa Gao, Robin Osterhout, Stephen J. Van Dien, and Joshua D. Rabinowitz. 2009. “Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in Escherichia Coli.” *Nature Chemical Biology* 5 (8): 593–99.

Bhasi, Kavitha, Alan Forrest, and Murali Ramanathan. 2005. “SPLINDID: A Semi-Parametric, Model-Based Method for Obtaining Transcription Rates and Gene Regulation Parameters from Genomic and Proteomic Expression Profiles.” *Bioinformatics* 21 (20): 3873–79.

Birk, Alexander V., Wesley M. Chao, Shaoyi Liu, Yi Soong, and Hazel H. Szeto. 2015. “Disruption of Cytochrome c Heme Coordination Is Responsible for Mitochondrial Injury during Ischemia.” *Biochimica et Biophysica Acta* 1847 (10): 1075–84.

Boersema, Paul J., Reinout Raijmakers, Simone Lemeer, Shabaz Mohammed, and Albert J. R. Heck. 2009. “Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics.” *Nature Protocols* 4 (4): 484–94.

Bogdanow, Boris, Henrik Zauber, and Matthias Selbach. 2016. “Systematic Errors in Peptide and Protein Identification and Quantification by Modified Peptides.” *Molecular & Cellular Proteomics: MCP* 15 (8): 2791–2801.

Bonate, Peter L. 2013. “The Effects of Active Metabolites on Parameter Estimation in Linear Mixed Effect Models of Concentration-QT Analyses.” *Journal of Pharmacokinetics and Pharmacodynamics* 40 (1): 101–15.

Braunwald, Eugene. 2015. “The War against Heart Failure: The Lancet Lecture.” *The Lancet* 385 (9970): 812–24.

Cheng, Mei-Ling, Chao-Hung Wang, Ming-Shi Shiao, Min-Hui Liu, Yu-Yen Huang, Cheng-Yu Huang, Chun-Tai Mao, Jui-Fen Lin, Hung-Yao Ho, and Ning-I Yang. 2015. “Metabolic Disturbances Identified in Plasma Are Associated with Outcomes in Patients with Heart Failure: Diagnostic and Prognostic Value of Metabolomics.” *Journal of the American College of Cardiology* 65 (15): 1509–20.

Chen, Tianlu, Guoxiang Xie, Xiaoying Wang, Jia Fan, Yunping Qiu, Xiaojiao Zheng, Xin Qi, et al. 2011. “Serum and Urine Metabolite Profiling Reveals Potential Biomarkers of Human Hepatocellular Carcinoma.” *Molecular & Cellular Proteomics: MCP* 10 (7): M110.004945.

Christin, Christin, Huub C. J. Hoefsloot, Age K. Smilde, B. Hoekman, Frank Suits, Rainer Bischoff, and Peter Horvatovich. 2013. “A Critical Assessment of Feature Selection Methods for Biomarker Discovery in Clinical Proteomics.” *Molecular & Cellular Proteomics: MCP* 12 (1): 263–76.

Chung, Heaseung Sophia, Christopher I. Murray, Vidya Venkatraman, Erin L. Crowgey, Peter P. Rainer, Robert N. Cole, Ryan D. Bomgarden, et al. 2015. “Dual Labeling Biotin Switch Assay to Reduce Bias Derived From Different Cysteine Subpopulations: A Method to Maximize S-Nitrosylation Detection.” *Circulation Research* 117 (10): 846–57.

Cohn, J. N., R. Ferrari, and N. Sharpe. 2000. “Cardiac Remodeling--Concepts and Clinical Implications: A Consensus Paper from an International Forum on Cardiac Remodeling. Behalf of an International Forum on Cardiac Remodeling.” *Journal of the American College of*

Cardiology 35 (3): 569–82.

Collins, Francis S., and Harold Varmus. 2015. “A New Initiative on Precision Medicine.” *The New England Journal of Medicine* 372 (9): 793–95.

“Comprehensive Natural Products II.” 2010. https://doi.org/10.1016/c2009-1-28362-6.

Cox, Jürgen, and Matthias Mann. 2008. “MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification.” *Nature Biotechnology* 26 (12): 1367–72.

Cox, Jürgen, Nadin Neuhauser, Annette Michalski, Richard A. Scheltema, Jesper V. Olsen, and Matthias Mann. 2011. “Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment.” *Journal of Proteome Research* 10 (4): 1794–1805.

Croft, David, Antonio Fabregat Mundo, Robin Haw, Marija Milacic, Joel Weiser, Guanming Wu, Michael Caudy, et al. 2014. “The Reactome Pathway Knowledgebase.” *Nucleic Acids Research* 42 (Database issue): D472–77.

Deutsch, Eric W., Attila Csordas, Zhi Sun, Andrew Jarnuczak, Yasset Perez-Riverol, Tobias Ternent, David S. Campbell, et al. 2017. “The ProteomeXchange Consortium in 2017: Supporting the Cultural Change in Proteomics Public Data Deposition.” *Nucleic Acids Research* 45 (D1): D1100–1106.

Djoussé, Luc, Pentti M. Rautaharju, Paul N. Hopkins, Eric A. Whitsel, Donna K. Arnett, John H. Eckfeldt, Michael A. Province, R. Curtis Ellison, and Investigators of the NHLBI Family Heart Study. 2005. “Dietary Linolenic Acid and Adjusted QT and JT Intervals in the National Heart, Lung, and Blood Institute Family Heart Study.” *Journal of the American College of Cardiology* 45 (10): 1716–22.

Doenst, Torsten, Tien Dung Nguyen, and E. Dale Abel. 2013. “Cardiac Metabolism in Heart Failure: Implications beyond ATP Production.” *Circulation Research* 113 (6): 709–24.

Dorn, Gerald W., 2nd, Jeffrey Robbins, and Peter H. Sugden. 2003. “Phenotyping Hypertrophy: Eschew Obfuscation.” *Circulation Research* 92 (11): 1171–75.

Dos Remedios, C. G., C. C. Liew, P. D. Allen, R. L. Winslow, J. E. Van Eyk, and M. J. Dunn. 2003. “Genomics, Proteomics and Bioinformatics of Human Heart Failure.” *Journal of Muscle Research and Cell Motility* 24 (4-6): 251–60.

Drews, Oliver, Osamu Tsukamoto, David Liem, John Streicher, Yibin Wang, and Peipei Ping. 2010. “Differential Regulation of Proteasome Function in Isoproterenol-Induced Cardiac Hypertrophy.” *Circulation Research* 107 (9): 1094–1101.

Duarte, Iola F., Cláudia M. Rocha, and Ana M. Gil. 2013. “Metabolic Profiling of Biofluids: Potential in Lung Cancer Screening and Diagnosis.” *Expert Review of Molecular Diagnostics* 13 (7): 737–48.

Dunbar, Sandra B., Olga A. Khavjou, Tamilyn Bakas, Gail Hunt, Rebecca A. Kirch, Alyssa R. Leib, R. Sean Morrison, et al. 2018. “Projected Costs of Informal Caregiving for Cardiovascular Disease: 2015 to 2035: A Policy Statement From the American Heart Association.” *Circulation*

137 (19): e558–77.

East, Mark A., Eric D. Peterson, Linda K. Shaw, Wendy A. Gattis, and Christopher M. O'Connor. 2004. "Racial Differences in the Outcomes of Patients with Diastolic Heart Failure." *American Heart Journal* 148 (1): 151–56.

Emwas, Abdul-Hamid M. 2015. "The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research." *Methods in Molecular Biology* 1277: 161–93.

Fabregat, Antonio, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, et al. 2018. "The Reactome Pathway Knowledgebase." *Nucleic Acids Research* 46 (D1): D649–55.

Fabregat, Antonio, Konstantinos Sidiropoulos, Phani Garapati, Marc Gillespie, Kerstin Hausmann, Robin Haw, Bijay Jassal, et al. 2016. "The Reactome Pathway Knowledgebase." *Nucleic Acids Research* 44 (D1): D481–87.

Fahy, Eoin, Shankar Subramaniam, Robert C. Murphy, Masahiro Nishijima, Christian R. H. Raetz, Takao Shimizu, Friedrich Spener, Gerrit van Meer, Michael J. O. Wakelam, and Edward A. Dennis. 2009. "Update of the LIPID MAPS Comprehensive Classification System for Lipids." *Journal of Lipid Research* 50 Suppl (April): S9–14.

Fahy, Eoin, Manish Sud, Dawn Cotter, and Shankar Subramaniam. 2007. "LIPID MAPS Online Tools for Lipid Research." *Nucleic Acids Research* 35 (Web Server issue): W606–12.
Fiaccavento, Roberta, Felicia Carotenuto, Marilena Minieri, Laura Masuelli, Alba Vecchini, Roberto Bei, Andrea Modesti, et al. 2006. "Alpha-Linolenic Acid-Enriched Diet Prevents Myocardial Damage and Expands Longevity in Cardiomyopathic Hamsters." *The American Journal of Pathology* 169 (6): 1913–24.

Folino, A., A. E. Sprio, F. Di Scipio, G. N. Berta, and R. Rastaldo. 2015. "Alpha-Linolenic Acid Protects against Cardiac Injury and Remodelling Induced by Beta-Adrenergic Overstimulation." *Food & Function* 6 (7): 2231–39.

Forrester, Michael T., and Jonathan S. Stamler. 2007. "A Classification Scheme for Redox-Based Modifications of Proteins." *American Journal of Respiratory Cell and Molecular Biology* 36 (2): 135–37.

Fu, Qiang, and Lingjun Li. 2005. "De Novo Sequencing of Neuropeptides Using Reductive Isotopic Methylation and Investigation of ESI QTOF MS/MS Fragmentation Pattern of Neuropeptides with N-Terminal Dimethylation." *Analytical Chemistry* 77 (23): 7783–95.

Furlotte, Nicholas A., and Eleazar Eskin. 2015. "Efficient Multiple-Trait Association and Estimation of Genetic Correlation Using the Matrix-Variate Linear Mixed Model." *Genetics* 200 (1): 59–68.

Gaggin, Hanna Kim, Annabel Angela Chen-Tournoux, Robert H. Christenson, Gheorghe Doros, Judd Eric Hollander, Phillip David Levy, John Tobias Nagurney, et al. 2017. "Rationale and Design of the ICON-RELOADED Study: International Collaborative of N-Terminal pro-B-Type Natriuretic Peptide Re-Evaluation of Acute Diagnostic Cut-Offs in the Emergency Department." *American Heart Journal* 192 (October): 26–37.

Gaggin, Hanna K., and James L. Januzzi Jr. 2013. “Biomarkers and Diagnostics in Heart Failure.” *Biochimica et Biophysica Acta* 1832 (12): 2442–50.

Ganti, Sheila, Sandra L. Taylor, Omran Abu Aboud, Joy Yang, Christopher Evans, Michael V. Osier, Danny C. Alexander, Kyoungmi Kim, and Robert H. Weiss. 2012. “Kidney Tumor Biomarkers Revealed by Simultaneous Multiple Matrix Metabolomics Analysis.” *Cancer Research* 72 (14): 3471–79.

García-Santamarina, Sarela, Susanna Boronat, Alba Domènech, José Ayté, Henrik Molina, and Elena Hidalgo. 2014. “Monitoring in Vivo Reversible Cysteine Oxidation in Proteins Using ICAT and Mass Spectrometry.” *Nature Protocols* 9 (5): 1131–45.

Gaudet, Pascale, Pierre-André Michel, Monique Zahn-Zabal, Aurore Britan, Isabelle Cusin, Marcin Domagalski, Paula D. Duek, et al. 2017. “The neXtProt Knowledgebase on Human Proteins: 2017 Update.” *Nucleic Acids Research* 45 (D1): D177–82.

Gerszten, Robert E., Aarti Asnani, and Steven A. Carr. 2011. “Status and Prospects for Discovery and Verification of New Biomarkers of Cardiovascular Disease by Proteomics.” *Circulation Research* 109 (4): 463–74.

Goldberg, Robert J., Julia Ciampa, Darleen Lessard, Theo E. Meyer, and Frederick A. Spencer. 2007. “Long-Term Survival after Heart Failure: A Contemporary Population-Based Perspective.” *Archives of Internal Medicine* 167 (5): 490–96.

Gonzalez, Daniel R., Adriana Treuer, Qi-An Sun, Jonathan S. Stamler, and Joshua M. Hare. 2009. “S-Nitrosylation of Cardiac Ion Channels.” *Journal of Cardiovascular Pharmacology* 54 (3): 188–95.

Graham, Stewart F., Olivier P. Chevallier, Dominic Roberts, Christian Hölscher, Christopher T. Elliott, and Brian D. Green. 2013. “Investigation of the Human Brain Metabolome to Identify Potential Markers for Early Diagnosis and Therapeutic Targets of Alzheimer’s Disease.” *Analytical Chemistry* 85 (3): 1803–11.

Gu, Liqing, and Renã A. S. Robinson. 2016. “A Simple Isotopic Labeling Method to Study Cysteine Oxidation in Alzheimer’s Disease: Oxidized Cysteine-Selective Dimethylation (OxcysDML).” *Analytical and Bioanalytical Chemistry* 408 (11): 2993–3004.

Guo, Hongbo, Hui Peng, and Andrew Emili. 2017. “Mass Spectrometry Methods to Study Protein-Metabolite Interactions.” *Expert Opinion on Drug Discovery* 12 (12): 1271–80.

Hall, John Edward, and Arthur C. Guyton. 2011. *Guyton and Hall Textbook of Medical Physiology*. Saunders.

Hastings, Janna, Paula de Matos, Adriano Dekker, Marcus Ennis, Bhavana Harsha, Namrata Kale, Venkatesh Muthukrishnan, et al. 2013. “The ChEBI Reference Database and Ontology for Biologically Relevant Chemistry: Enhancements for 2013.” *Nucleic Acids Research* 41 (Database issue): D456–63.

Haug, Kenneth, Reza M. Salek, Pablo Conesa, Janna Hastings, Paula de Matos, Mark Rijnbeek, Tejasvi Mahendraker, et al. 2013. “MetaboLights--an Open-Access General-Purpose

Repository for Metabolomics Studies and Associated Meta-Data." *Nucleic Acids Research* 41 (Database issue): D781–86.

Haugland, Rosaria P., and Wendy W. You. 2008. "Coupling of Antibodies with Biotin." *Methods in Molecular Biology* 418: 13–24.

Hoefer, Imo E., Sabine Steffens, Mika Ala-Korpela, Magnus Bäck, Lina Badimon, Marie-Luce Bochaton-Piallat, Chantal M. Boulanger, et al. 2015. "Novel Methodologies for Biomarker Discovery in Atherosclerosis." *European Heart Journal* 36 (39): 2635–42.

Hou, Yunlong, Juan M. Adrian-Segarra, Manfred Richter, Natalia Kubin, Jaeyoung Shin, Isabella Werner, Thomas Walther, et al. 2015. "Animal Models and 'Omics' Technologies for Identification of Novel Biomarkers and Drug Targets to Prevent Heart Failure." *BioMed Research International* 2015 (July): 212910.

Hsu, Jue-Liang, Sheng-Yu Huang, Nan-Haw Chow, and Shu-Hui Chen. 2003. "Stable-Isotope Dimethyl Labeling for Quantitative Proteomics." *Analytical Chemistry* 75 (24): 6843–52.

Hua, Yinan, Yingmei Zhang, and Jun Ren. 2012. "IGF-1 Deficiency Resists Cardiac Hypertrophy and Myocardial Contractile Dysfunction: Role of microRNA-1 and microRNA-133a." *Journal of Cellular and Molecular Medicine* 16 (1): 83–95.

Husaini, Baqar A., George A. Mensah, Douglas Sawyer, Van A. Cain, Zahid Samad, Pamela C. Hull, Robert S. Levine, and Uchechukwu K. A. Sampson. 2011. "Race, Sex, and Age Differences in Heart Failure-Related Hospitalizations in a Southern State: Implications for Prevention." *Circulation. Heart Failure* 4 (2): 161–69.

Jaffrey, S. R., and S. H. Snyder. 2001. "The Biotin Switch Method for the Detection of S-Nitrosylated Proteins." *Science's STKE: Signal Transduction Knowledge Environment* 2001 (86): l1.

Jové, Mariona, Manuel Portero-Otín, Alba Naudí, Isidre Ferrer, and Reinald Pamplona. 2014. "Metabolomics of Human Brain Aging and Age-Related Neurodegenerative Diseases." *Journal of Neuropathology and Experimental Neurology* 73 (7): 640–57.

Kang, Hyun Min, Jae Hoon Sul, Susan K. Service, Noah A. Zaitlen, Sit-Yee Kong, Nelson B. Freimer, Chiara Sabatti, and Eleazar Eskin. 2010. "Variance Component Model to Account for Sample Structure in Genome-Wide Association Studies." *Nature Genetics* 42 (4): 348–54.

Kang, Hyun Min, Noah A. Zaitlen, Claire M. Wade, Andrew Kirby, David Heckerman, Mark J. Daly, and Eleazar Eskin. 2008. "Efficient Control of Population Structure in Model Organism Association Mapping." *Genetics* 178 (3): 1709–23.

Kang, Jian, Ling Zhu, Jingli Lu, and Xiaojian Zhang. 2015. "Application of Metabolomics in Autoimmune Diseases: Insight into Biomarkers and Pathology." *Journal of Neuroimmunology* 279 (February): 25–32.

Kaplan. 2008. *Kaplan CSET: California Subject Examination for Teachers*. Kaplan Publishing.

Kim, Min-Sik, Sneha M. Pinto, Derese Getnet, Raja Sekhar Nirujogi, Srikanth S. Manda, Raghothama Chaerkady, Anil K. Madugundu, et al. 2014. "A Draft Map of the Human

Proteome." *Nature* 509 (7502): 575–81.

Kirby, Andrew, Hyun Min Kang, Claire M. Wade, Chris Cotsapas, Emrah Kostem, Buhm Han, Nick Furlotte, et al. 2010. "Fine Mapping in 94 Inbred Mouse Strains Using a High-Density Haplotype Resource." *Genetics* 185 (3): 1081–95.

Kochanowski, Karl, Uwe Sauer, and Elad Noor. 2015. "Posttranslational Regulation of Microbial Metabolism." *Current Opinion in Microbiology* 27 (October): 10–17.

Kohr, Mark J., Angel Aponte, Junhui Sun, Marjan Gucek, Charles Steenbergen, and Elizabeth Murphy. 2012. "Measurement of S-Nitrosylation Occupancy in the Myocardium with Cysteine-Reactive Tandem Mass Tags: Short Communication." *Circulation Research* 111 (10): 1308–12.

Kohr, Mark J., Alicia M. Evangelista, Marcella Ferlito, Charles Steenbergen, and Elizabeth Murphy. 2014. "S-Nitrosylation of TRIM72 at Cysteine 144 Is Critical for Protection against Oxidation-Induced Protein Degradation and Cell Death." *Journal of Molecular and Cellular Cardiology* 69 (April): 67–74.

Kolwicz, Stephen C., Jr, Suneet Purohit, and Rong Tian. 2013. "Cardiac Metabolism and Its Interactions with Contraction, Growth, and Survival of Cardiomyocytes." *Circulation Research* 113 (5): 603–16.

Kolwicz, Stephen C., Jr, and Rong Tian. 2011. "Glucose Metabolism and Cardiac Hypertrophy." *Cardiovascular Research* 90 (2): 194–201.

Koutsoukas, Alexios, Benjamin Simms, Johannes Kirchmair, Peter J. Bond, Alan V. Whitmore, Steven Zimmer, Malcolm P. Young, et al. 2011. "From in Silico Target Prediction to Multi-Target Drug Design: Current Databases, Methods and Applications." *Journal of Proteomics* 74 (12): 2554–74.

Kramer, Philip A., Jicheng Duan, Wei-Jun Qian, and David J. Marcinek. 2015. "The Measurement of Reversible Redox Dependent Post-Translational Modifications and Their Regulation of Mitochondrial and Skeletal Muscle Function." *Frontiers in Physiology* 6 (November): 347.

Krenek, Peter, Jana Kmecova, Dana Kucerova, Zuzana Bajuszova, Peter Musil, Andrea Gazova, Peter Ochodnicky, Jan Klimas, and Jan Kyselovic. 2009. "Isoproterenol-Induced Heart Failure in the Rat Is Associated with Nitric Oxide-Dependent Functional Alterations of Cardiac Function." *European Journal of Heart Failure* 11 (2): 140–46.

Kuhn, Michael, Christian von Mering, Monica Campillos, Lars Juhl Jensen, and Peer Bork. 2008. "STITCH: Interaction Networks of Chemicals and Proteins." *Nucleic Acids Research* 36 (Database issue): D684–88.

Kuhn, Michael, Damian Szklarczyk, Sune Pletscher-Frankild, Thomas H. Blicher, Christian von Mering, Lars J. Jensen, and Peer Bork. 2014. "STITCH 4: Integration of Protein-Chemical Interactions with User Data." *Nucleic Acids Research* 42 (Database issue): D401–7.

Lam, Maggie P. Y., Ding Wang, Edward Lau, David A. Liem, Allen K. Kim, Dominic C. M. Ng, Xiangbo Liang, et al. 2014. "Protein Kinetic Signatures of the Remodeling Heart Following Isoproterenol Stimulation." *The Journal of Clinical Investigation* 124 (4): 1734–44.

Lau, Edward, Quan Cao, Maggie P. Y. Lam, Jie Wang, Dominic C. M. Ng, Brian J. Bleakley, Jessica M. Lee, et al. 2018. “Integrated Omics Dissection of Proteome Dynamics during Cardiac Remodeling.” *Nature Communications* 9 (1): 120.

Lau, Edward, Quan Cao, Dominic C. M. Ng, Brian J. Bleakley, T. Umut Dincer, Brian M. Bot, Ding Wang, et al. 2016. “A Large Dataset of Protein Dynamics in the Mammalian Heart Proteome.” *Scientific Data* 3 (March): 160015.

Liem, David A., Ali Nsair, Shaun P. Setty, Martin Cadeiras, Ding Wang, Robb Maclellan, Chris Lotz, et al. 2014. “Molecular- and Organelle-Based Predictive Paradigm Underlying Recovery by Left Ventricular Assist Device Support.” *Circulation. Heart Failure* 7 (2): 359–66.

Li, Haomin, Nobel C. Zong, Xiangbo Liang, Allen K. Kim, Jeong Ho Choi, Ning Deng, Ivette Zelaya, Maggie Lam, Huilong Duan, and Peipei Ping. 2013. “A Novel Spectral Library Workflow to Enhance Protein Identifications.” *Journal of Proteomics* 81 (April): 173–84.

Lim, Amareth, Tatiana Prokaeva, Mark E. McComb, Lawreen H. Connors, Martha Skinner, and Catherine E. Costello. 2003. “Identification of S-Sulfonation and S-Thiolation of a Novel Transthyretin Phe33Cys Variant from a Patient Diagnosed with Familial Transthyretin Amyloidosis.” *Protein Science: A Publication of the Protein Society* 12 (8): 1775–85.

Li, Ru, Jiqing Huang, and Juergen Kast. 2015. “Identification of Total Reversible Cysteine Oxidation in an Atherosclerosis Model Using a Modified Biotin Switch Assay.” *Journal of Proteome Research* 14 (5): 2026–35.

Liu, Limin, Yun Yan, Ming Zeng, Jian Zhang, Martha A. Hanes, Gregory Ahearn, Timothy J. McMahon, et al. 2004. “Essential Roles of S-Nitrosothiols in Vascular Homeostasis and Endotoxic Shock.” *Cell* 116 (4): 617–28.

Liu, Tzu-Yu, Hector H. Huang, Diamond Wheeler, Yichen Xu, James A. Wells, Yun S. Song, and Arun P. Wiita. 2017. “Time-Resolved Proteomics Extends Ribosome Profiling-Based Measurements of Protein Synthesis Dynamics.” *Cell Systems* 4 (6): 636–44.e9.

Li, Xiyan, Tara A. Gianoulis, Kevin Y. Yip, Mark Gerstein, and Michael Snyder. 2010. “Extensive in Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses.” *Cell* 143 (4): 639–50.

Li, Xiyan, and Michael Snyder. 2011. “Metabolites as Global Regulators: A New View of Protein Regulation: Systematic Investigation of Metabolite-Protein Interactions May Help Bridge the Gap between Genome-Wide Association Studies and Small Molecule Screening Studies.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 33 (7): 485–89.

Lloyd-Jones, Donald M. 2010. “Cardiovascular Risk Prediction: Basic Concepts, Current Status, and Future Directions.” *Circulation* 121 (15): 1768–77.

Lotz, Christopher, Amanda J. Lin, Caitlin M. Black, Jun Zhang, Edward Lau, Ning Deng, Yueju Wang, et al. 2014. “Characterization, Design, and Function of the Mitochondrial Proteome: From Organs to Organisms.” *Journal of Proteome Research* 13 (2): 433–46.

Lowther, W. Todd, and Alexina C. Haynes. 2011. “Reduction of Cysteine Sulfinic Acid in Eukaryotic, Typical 2-Cys Peroxiredoxins by Sulfiredoxin.” *Antioxidants & Redox Signaling* 15 (1): 99–109.

Mancini, Donna, and Paolo C. Colombo. 2015. “Left Ventricular Assist Devices: A Rapidly Evolving Alternative to Transplant.” *Journal of the American College of Cardiology* 65 (23): 2542–55.

Mann, Matthias. 2006. “Functional and Quantitative Proteomics Using SILAC.” *Nature Reviews. Molecular Cell Biology* 7 (12): 952–58.

Maulik, Subir Kumar, and Santosh Kumar. 2012. “Oxidative Stress and Cardiac Hypertrophy: A Review.” *Toxicology Mechanisms and Methods* 22 (5): 359–66.

McMurray, John J. V. 2010. “Clinical Practice. Systolic Heart Failure.” *The New England Journal of Medicine* 362 (3): 228–38.

Menazza, Sara, Angel Aponte, Junhui Sun, Marjan Gucek, Charles Steenbergen, and Elizabeth Murphy. 2015. “Molecular Signature of Nitroso-Redox Balance in Idiopathic Dilated Cardiomyopathies.” *Journal of the American Heart Association* 4 (9): e002251.

Mensah, George A., Ali H. Mokdad, Earl S. Ford, Kurt J. Greenlund, and Janet B. Croft. 2005. “State of Disparities in Cardiovascular Health in the United States.” *Circulation* 111 (10): 1233–41.

Milacic, Marija, Robin Haw, Karen Rothfels, Guanming Wu, David Croft, Henning Hermjakob, Peter D’Eustachio, and Lincoln Stein. 2012. “Annotating Cancer Variants and Anti-Cancer Therapeutics in Reactome.” *Cancers* 4 (4): 1180–1211.

Milo, Ron. 2013. “What Is the Total Number of Protein Molecules per Cell Volume? A Call to Rethink Some Published Values.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 35 (12): 1050–55.

Morrow, David A., James A. de Lemos, Marc S. Sabatine, and Elliott M. Antman. 2003. “The Search for a Biomarker of Cardiac Ischemia.” *Clinical Chemistry* 49 (4): 537–39.

Murakami, T., M. Nojiri, H. Nakayama, M. Odaka, M. Yohda, N. Dohmae, K. Takio, T. Nagamune, and I. Endo. 2000. “Post-Translational Modification Is Essential for Catalytic Activity of Nitrile Hydratase.” *Protein Science: A Publication of the Protein Society* 9 (5): 1024–30.

Murray, Christopher I., Hea Seung Chung, Helge Uhrigshardt, and Jennifer E. Van Eyk. 2013. “Quantification of Mitochondrial S-Nitrosylation by CysTMT^6 Switch Assay.” *Methods in Molecular Biology* 1005: 169–79.

Murray, Christopher I., Helge Uhrigshardt, Robert N. O’Meally, Robert N. Cole, and Jennifer E. Van Eyk. 2012. “Identification and Quantification of S-Nitrosylation by Cysteine Reactive Tandem Mass Tag Switch Assay.” *Molecular & Cellular Proteomics: MCP* 11 (2): M111.013441.

Murray, Christopher I., and Jennifer E. Van Eyk. 2012. “Chasing Cysteine Oxidative Modifications: Proteomic Tools for Characterizing Cysteine Redox Status.” *Circulation. Cardiovascular Genetics* 5 (5): 591.

Neubauer, Stefan. 2007. “The Failing Heart--an Engine out of Fuel.” *The New England Journal of Medicine* 356 (11): 1140–51.

Nierman, D. M., and J. I. Mechanick. 1999. “Hypotestosteronemia in Chronically Critically Ill Men.” *Critical Care Medicine* 27 (11): 2418–21.

Nikolaev, Yaroslav V., Karl Kochanowski, Hannes Link, Uwe Sauer, and Frederic H-T Allain. 2016. “Systematic Identification of Protein-Metabolite Interactions in Complex Metabolite Mixtures by Ligand-Detected Nuclear Magnetic Resonance Spectroscopy.” *Biochemistry* 55 (18): 2590–2600.

O’Donnell, John P., Heather M. Marsh, Holger Sondermann, and Carolyn S. Sevier. 2018. “Disrupted Hydrogen-Bond Network and Impaired ATPase Activity in an Hsc70 Cysteine Mutant.” *Biochemistry* 57 (7): 1073–86.

Orchard, Sandra, Mais Ammari, Bruno Aranda, Lionel Breuza, Leonardo Briganti, Fiona Broackes-Carter, Nancy H. Campbell, et al. 2014. “The MIntAct Project--IntAct as a Common Curation Platform for 11 Molecular Interaction Databases.” *Nucleic Acids Research* 42 (Database issue): D358–63.

Orozco, Luz D., Marco Morselli, Liudmilla Rubbi, Weilong Guo, James Go, Huwenbo Shi, David Lopez, et al. 2015. “Epigenome-Wide Association of Liver Methylation Patterns and Complex Metabolic Traits in Mice.” *Cell Metabolism* 21 (6): 905–17.

Pastore, Anna, and Fiorella Piemonte. 2013. “Protein Glutathionylation in Cardiovascular Diseases.” *International Journal of Molecular Sciences* 14 (10): 20845–76.

Patil, Mallikarjun, Keyur A. Sheth, Adarsh C. Krishnamurthy, and Harshad Devarbhavi. 2013. “A Review and Current Perspective on Wilson Disease.” *Journal of Clinical and Experimental Hepatology* 3 (4): 321–36.

Piazza, Ilaria, Karl Kochanowski, Valentina Cappelletti, Tobias Fuhrer, Elad Noor, Uwe Sauer, and Paola Picotti. 2018. “A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication.” *Cell* 172 (1-2): 358–72.e23.

Piquereau, Jerome, Fanny Caffin, Marta Novotova, Christophe Lemaire, Vladimir Veksler, Anne Garnier, Renee Ventura-Clapier, and Frederic Joubert. 2013. “Mitochondrial Dynamics in the Adult Cardiomyocytes: Which Roles for a Highly Specialized Cell?” *Frontiers in Physiology* 4 (May): 102.

Qin, Fuzhong, Deborah A. Siwik, Steve Lancel, Jingmei Zhang, Gabriela M. Kuster, Ivan Luptak, Lei Wang, et al. 2013. “Hydrogen Peroxide-Mediated SERCA Cysteine 674 Oxidation Contributes to Impaired Cardiac Myocyte Relaxation in Senescent Mouse Heart.” *Journal of the American Heart Association* 2 (4): e000184.

Qin, Fuzhong, Deborah A. Siwik, David R. Pimentel, Robert J. Morgan, Andreia Biolo, Vivian H. Tu, Y. James Kang, Richard A. Cohen, and Wilson S. Colucci. 2014. “Cytosolic H2O2 Mediates Hypertrophy, Apoptosis, and Decreased SERCA Activity in Mice with Chronic Hemodynamic Overload.” *American Journal of Physiology. Heart and Circulatory Physiology* 306 (10): H1453–63.

Rau, Christoph D., Jessica Wang, Rozeta Avetisyan, Milagros C. Romay, Lisa Martin, Shuxun Ren, Yibin Wang, and Aldons J. Lusis. 2015. “Mapping Genetic Contributions to Cardiac Pathology Induced by Beta-Adrenergic Stimulation in Mice.” *Circulation. Cardiovascular Genetics* 8 (1): 40–49.

Reest, Jiska van der, Sergio Lilla, Liang Zheng, Sara Zanivan, and Eyal Gottlieb. 2018. “Proteome-Wide Analysis of Cysteine Oxidation Reveals Metabolic Sensitivity to Redox Stress.” *Nature Communications* 9 (1): 1581.

Reisz, Julie A., Erika Bechtold, S. Bruce King, Leslie B. Poole, and Cristina M. Furdui. 2013. “Thiol-Blocking Electrophiles Interfere with Labeling and Detection of Protein Sulfenic Acids.” *The FEBS Journal* 280 (23): 6150–61.

Ribeiro, Marco, Sameer Singh, and Carlos Guestrin. 2016. “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier.” *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. https://doi.org/10.18653/v1/n16-3020.

Riesmeyer, Jeffrey S., Daniel E. Salazar, Govinda J. Weerakkody, Lan Ni, Rebecca E. Wrishko, C. Steven Ernest 2nd, Junxiang Luo, et al. 2012. “Relationship between Exposure to Prasugrel Active Metabolite and Clinical Outcomes in the TRITON-TIMI 38 Substudy.” *Journal of Clinical Pharmacology* 52 (6): 789–97.

Roger, Véronique L. 2013. “Epidemiology of Heart Failure.” *Circulation Research* 113 (6): 646–59.

Roux, Aurélie, Dominique Lison, Christophe Junot, and Jean-François Heilier. 2011. “Applications of Liquid Chromatography Coupled to Mass Spectrometry-Based Metabolomics in Clinical Chemistry and Toxicology: A Review.” *Clinical Biochemistry* 44 (1): 119–35.

Sag, Can M., Celio X. C. Santos, and Ajay M. Shah. 2014. “Redox Regulation of Cardiac Hypertrophy.” *Journal of Molecular and Cellular Cardiology* 73 (August): 103–11.

Sansbury, Brian E., Angelica M. DeMartino, Zhengzhi Xie, Alan C. Brooks, Robert E. Brainard, Lewis J. Watson, Andrew P. DeFilippis, et al. 2014. “Metabolomic Analysis of Pressure-Overloaded and Infarcted Mouse Hearts.” *Circulation. Heart Failure* 7 (4): 634–42.

Seth, Divya, and Jonathan S. Stamler. 2015. “SNOs Differ: Methodological and Biological Implications.” *Circulation Research* 117 (10): 826–29.

Shah, Amil M., Brian Claggett, Nancy K. Sweitzer, Sanjiv J. Shah, Inder S. Anand, Eileen O’Meara, Akshay S. Desai, et al. 2014. “Cardiac Structure and Function and Prognosis in Heart Failure with Preserved Ejection Fraction: Findings from the Echocardiographic Study of the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) Trial.” *Circulation. Heart Failure* 7 (5): 740–51.

Shah, Sanjiv J., Daniel H. Katz, Senthil Selvaraj, Michael A. Burke, Clyde W. Yancy, Mihai Gheorghiade, Robert O. Bonow, Chiang-Ching Huang, and Rahul C. Deo. 2015. “Phenomapping for Novel Classification of Heart Failure with Preserved Ejection Fraction.” *Circulation* 131 (3): 269–79.

Shah, Svati H., William E. Kraus, and Christopher B. Newgard. 2012. “Metabolomic Profiling for the Identification of Novel Biomarkers and Mechanisms Related to Common Cardiovascular Diseases: Form and Function.” *Circulation* 126 (9): 1110–20.

Shimizu, Ippei, and Tohru Minamino. 2016. “Physiological and Pathological Cardiac Hypertrophy.” *Journal of Molecular and Cellular Cardiology* 97 (August): 245–62.

Singh, R. B., M. A. Niaz, J. P. Sharma, R. Kumar, V. Rastogi, and M. Moshiri. 1997. “Randomized, Double-Blind, Placebo-Controlled Trial of Fish Oil and Mustard Oil in Patients with Suspected Acute Myocardial Infarction: The Indian Experiment of Infarct Survival--4.” *Cardiovascular Drugs and Therapy / Sponsored by the International Society of Cardiovascular Pharmacotherapy* 11 (3): 485–91.

Soininen, Pasi, Antti J. Kangas, Peter Würtz, Teemu Suna, and Mika Ala-Korpela. 2015. “Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics.” *Circulation. Cardiovascular Genetics* 8 (1): 192–206.

Souders, Colby A., Thomas K. Borg, Indroneal Banerjee, and Troy A. Baudino. 2012. “Pressure Overload Induces Early Morphological Changes in the Heart.” *The American Journal of Pathology* 181 (4): 1226–35.

Starr, Cecie, Christine Evers, and Lisa Starr. 2009. *Biology Today and Tomorrow with Physiology*. Cengage Learning.

Stegemann, Christin, Raimund Pechlaner, Peter Willeit, Sarah R. Langley, Massimo Mangino, Ursula Mayr, Cristina Menni, et al. 2014. “Lipidomics Profiling and Risk of Cardiovascular Disease in the Prospective Population-Based Bruneck Study.” *Circulation* 129 (18): 1821–31.

Stomberski, Colin T., Douglas T. Hess, and Jonathan S. Stamler. 2019. “Protein S-Nitrosylation: Determinants of Specificity and Enzymatic Regulation of S-Nitrosothiol-Based Signaling.” *Antioxidants & Redox Signaling* 30 (10): 1331–51.

Straube, Jasmin, Alain-Dominique Gorse, PROOF Centre of Excellence Team, Bevan Emma Huang, and Kim-Anh Lê Cao. 2015. “A Linear Mixed Model Spline Framework for Analysing Time Course ‘Omics’ Data.” *PloS One* 10 (8): e0134540.

Streng, Alexander S., Douwe de Boer, Jolanda van der Velden, Marja P. van Dieijen-Visser, and Will K. W. H. Wodzig. 2013. “Posttranslational Modifications of Cardiac Troponin T: An Overview.” *Journal of Molecular and Cellular Cardiology* 63 (October): 47–56.

Sun, Haipeng, Kristine C. Olson, Chen Gao, Domenick A. Prosdocimo, Meiyi Zhou, Zhihua Wang, Darwin Jeyaraj, et al. 2016. “Catabolic Defect of Branched-Chain Amino Acids Promotes Heart Failure.” *Circulation* 133 (21): 2038–49.

Szklarczyk, Damian, Alberto Santos, Christian von Mering, Lars Juhl Jensen, Peer Bork, and Michael Kuhn. 2016. “STITCH 5: Augmenting Protein-Chemical Interaction Networks with Tissue and Affinity Data.” *Nucleic Acids Research* 44 (D1): D380–84.

Taylor, Anne L., Susan Ziesche, Clyde Yancy, Peter Carson, Ralph D’Agostino Jr, Keith Ferdinand, Malcolm Taylor, et al. 2004. “Combination of Isosorbide Dinitrate and Hydralazine in

Blacks with Heart Failure." *The New England Journal of Medicine* 351 (20): 2049–57.

Tchagang, Alain B., Kevin V. Bui, Thomas McGinnis, and Panayiotis V. Benos. 2009. "Extracting Biologically Significant Patterns from Short Time Series Gene Expression Data." *BMC Bioinformatics* 10 (August): 255.

Tolstikov, Vladimir, Alexander Nikolayev, Sucai Dong, Genshi Zhao, and Ming-Shang Kuo. 2014. "Metabolomics Analysis of Metabolic Effects of Nicotinamide Phosphoribosyltransferase (NAMPT) Inhibition on Human Cancer Cells." *PloS One* 9 (12): e114019.

Troncoso, Rodrigo, Cristián Ibarra, Jose Miguel Vicencio, Enrique Jaimovich, and Sergio Lavandero. 2014. "New Insights into IGF-1 Signaling in the Heart." *Trends in Endocrinology and Metabolism: TEM* 25 (3): 128–37.

Tsueng, Ginger, Benjamin M. Good, Peipei Ping, Erica Golemis, Israel Hanukoglu, Andre J. van Wijnen, and Andrew I. Su. 2016. "Gene Wiki Reviews-Raising the Quality and Accessibility of Information about the Human Genome." *Gene* 592 (2): 235–38.

Twerenbold, Raphael, Allan Jaffe, Tobias Reichlin, Miriam Reiter, and Christian Mueller. 2012. "High-Sensitive Troponin T Measurements: What Do We Gain and What Are the Challenges?" *European Heart Journal* 33 (5): 579–86.

UniProt Consortium. 2015. "UniProt: A Hub for Protein Information." *Nucleic Acids Research* 43 (Database issue): D204–12.

Venes, Donald. 2009. *Taber's Cyclopedic Medical Dictionary*. F A Davis Company.

Vizcaíno, Juan Antonio, Attila Csordas, Noemi Del-Toro, José A. Dianes, Johannes Griss, Ilias Lavidas, Gerhard Mayer, et al. 2016. "2016 Update of the PRIDE Database and Its Related Tools." *Nucleic Acids Research* 44 (22): 11033.

Wagner, Gregory R., and R. Mark Payne. 2013. "Widespread and Enzyme-Independent Nε-Acetylation and Nε-Succinylation of Proteins in the Chemical Conditions of the Mitochondrial Matrix." *The Journal of Biological Chemistry* 288 (40): 29036–45.

Wang, Gang, Yi Zhou, Feng-Jie Huang, Hui-Dong Tang, Xu-Hua Xu, Jia-Jian Liu, Ying Wang, et al. 2014. "Plasma Metabolite Profiles of Alzheimer's Disease and Mild Cognitive Impairment." *Journal of Proteome Research* 13 (5): 2649–58.

Wang, Jie, Kristeen A. Pareja, Chris A. Kaiser, and Carolyn S. Sevier. 2014. "Redox Signaling via the Molecular Chaperone BiP Protects Cells against Endoplasmic Reticulum-Derived Oxidative Stress." *eLife* 3 (July): e03496.

Wang, Jie, and Carolyn S. Sevier. 2016. "Formation and Reversibility of BiP Protein Cysteine Oxidation Facilitate Cell Survival during and Post Oxidative Stress." *The Journal of Biological Chemistry* 291 (14): 7541–57.

Wang, Sheng-Bing, Vidya Venkatraman, Erin L. Crowgey, Ting Liu, Zongming Fu, Ronald Holewinski, Mark Ranek, David A. Kass, Brian O'Rourke, and Jennifer E. Van Eyk. 2018. "Protein -Nitrosylation Controls Glycogen Synthase Kinase 3β Function Independent of Its Phosphorylation State." *Circulation Research* 122 (11): 1517–31.

Wang, Thomas J. 2011. “Assessing the Role of Circulating, Genetic, and Imaging Biomarkers in Cardiovascular Risk Prediction.” *Circulation* 123 (5): 551–65.

Weinert, Brian T., Vytautas Iesmantavicius, Sebastian A. Wagner, Christian Schölz, Bertil Gummesson, Petra Beli, Thomas Nyström, and Chunaram Choudhary. 2013. “Acetyl-Phosphate Is a Critical Determinant of Lysine Acetylation in E. Coli.” *Molecular Cell* 51 (2): 265–72.

Weinstock-Guttman, Bianca, Robert Zivadinov, Jun Qu, Diane Cookfair, Xiaotao Duan, Eunjin Bang, Niels Bergsland, et al. 2011. “Vitamin D Metabolites Are Associated with Clinical and MRI Outcomes in Multiple Sclerosis Patients.” *Journal of Neurology, Neurosurgery, and Psychiatry* 82 (2): 189–95.

Wishart, David S., Timothy Jewison, An Chi Guo, Michael Wilson, Craig Knox, Yifeng Liu, Yannick Djoumbou, et al. 2013. “HMDB 3.0--The Human Metabolome Database in 2013.” *Nucleic Acids Research* 41 (Database issue): D801–7.

Wishart, David S., Craig Knox, An Chi Guo, Roman Eisner, Nelson Young, Bijaya Gautam, David D. Hau, et al. 2009. “HMDB: A Knowledgebase for the Human Metabolome.” *Nucleic Acids Research* 37 (Database issue): D603–10.

Wishart, David S., Dan Tzur, Craig Knox, Roman Eisner, An Chi Guo, Nelson Young, Dean Cheng, et al. 2007. “HMDB: The Human Metabolome Database.” *Nucleic Acids Research* 35 (Database issue): D521–26.

Woo, Hyun Ae, Woojin Jeong, Tong-Shin Chang, Kwang Joo Park, Sung Jun Park, Jeong Soo Yang, and Sue Goo Rhee. 2005. “Reduction of Cysteine Sulfinic Acid by Sulfiredoxin Is Specific to 2-Cys Peroxiredoxins.” *The Journal of Biological Chemistry* 280 (5): 3125–28.

WRITING COMMITTEE MEMBERS, Clyde W. Yancy, Mariell Jessup, Biykem Bozkurt, Javed Butler, Donald E. Casey Jr, Mark H. Drazner, et al. 2013. “2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines.” *Circulation* 128 (16): e240–327.

Wu, Chunlei, Xuefeng Jin, Ginger Tsueng, Cyrus Afrasiabi, and Andrew I. Su. 2016. “BioGPS: Building Your Own Mash-up of Gene Annotations and Expression Profiles.” *Nucleic Acids Research* 44 (D1): D313–16.

Wu, Chunlei, Ian Macleod, and Andrew I. Su. 2013. “BioGPS and MyGene.info: Organizing Online, Gene-Centric Information.” *Nucleic Acids Research* 41 (Database issue): D561–65.

Würtz, Peter, Aki S. Havulinna, Pasi Soininen, Tuulia Tynkkynen, David Prieto-Merino, Therese Tillin, Anahita Ghorbani, et al. 2015. “Metabolite Profiling and Cardiovascular Event Risk: A Prospective Study of 3 Population-Based Cohorts.” *Circulation* 131 (9): 774–85.

Xia, Jianguo, Rupasri Mandal, Igor V. Sinelnikov, David Broadhurst, and David S. Wishart. 2012. “MetaboAnalyst 2.0--a Comprehensive Server for Metabolomic Data Analysis.” *Nucleic Acids Research* 40 (Web Server issue): W127–33.

Xia, Jianguo, Nick Psychogios, Nelson Young, and David S. Wishart. 2009. “MetaboAnalyst: A Web Server for Metabolomic Data Analysis and Interpretation.” *Nucleic Acids Research* 37 (Web Server issue): W652–60.

Xia, Jianguo, Igor V. Sinelnikov, Beomsoo Han, and David S. Wishart. 2015. “MetaboAnalyst 3.0--Making Metabolomics More Meaningful.” *Nucleic Acids Research* 43 (W1): W251–57.

Xia, Jianguo, and David S. Wishart. 2011a. “Metabolomic Data Processing, Analysis, and Interpretation Using MetaboAnalyst.” *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* Chapter 14 (June): Unit 14.10.
———. 2011b. “Web-Based Inference of Biological Patterns, Functions and Pathways from Metabolomic Data Using MetaboAnalyst.” *Nature Protocols* 6 (6): 743–60.

Xu, Jing, Yanhua Chen, Ruiping Zhang, Yongmei Song, Jianzhong Cao, Nan Bi, Jingbo Wang, et al. 2013. “Global and Targeted Metabolomics of Esophageal Squamous Cell Carcinoma Discovers Potential Diagnostic and Therapeutic Biomarkers.” *Molecular & Cellular Proteomics: MCP* 12 (5): 1306–18.

Yang, Laurence, Justin Tan, Edward J. O’Brien, Jonathan M. Monk, Donghyuk Kim, Howard J. Li, Pep Charusanti, et al. 2015. “Systems Biology Definition of the Core Proteome of Metabolism and Expression Is Consistent with High-Throughput Data.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (34): 10810–15.

Zhang, Aihua, Hui Sun, Ping Wang, Ying Han, and Xijun Wang. 2012. “Modern Analytical Techniques in Metabolomics Analysis.” *The Analyst* 137 (2): 293–300.

Zhang, Aihua, Hui Sun, Xiuhong Wu, and Xijun Wang. 2012. “Urine Metabolomics.” *Clinica Chimica Acta; International Journal of Clinical Chemistry* 414 (December): 65–69.

Zhang, Aihua, Hui Sun, Guangli Yan, Ping Wang, and Xijun Wang. 2016. “Mass Spectrometry-Based Metabolomics: Applications to Biomarker and Metabolic Pathway Research.” *Biomedical Chromatography: BMC* 30 (1): 7–12.

Zhang, Guo-Xing, Shoji Kimura, Akira Nishiyama, Takatomi Shokoji, Matlubur Rahman, Li Yao, Yukiko Nagai, Yoshihide Fujisawa, Akira Miyatake, and Youichi Abe. 2005. “Cardiac Oxidative Stress in Acute and Chronic Isoproterenol-Infused Rats.” *Cardiovascular Research* 65 (1): 230–38.

Zhang, Guo-Xing, Koji Ohmori, Yukiko Nagai, Yoshihide Fujisawa, Akira Nishiyama, Youichi Abe, and Shoji Kimura. 2007. “Role of AT1 Receptor in Isoproterenol-Induced Cardiac Hypertrophy and Oxidative Stress in Mice.” *Journal of Molecular and Cellular Cardiology* 42 (4): 804–11.

Zhang, Rongli, Douglas T. Hess, James D. Reynolds, and Jonathan S. Stamler. 2016. “Hemoglobin S-Nitrosylation Plays an Essential Role in Cardioprotection.” *The Journal of Clinical Investigation* 126 (12): 4654–58.

Zong, Nobel C., Haomin Li, Hua Li, Maggie P. Y. Lam, Rafael C. Jimenez, Christina S. Kim, Ning Deng, et al. 2013. “Integration of Cardiac Proteome Biology and Medicine by a Specialized Knowledgebase.” *Circulation Research* 113 (9): 1043–53.

Zong, Nobel, Peipei Ping, Edward Lau, Howard Jh Choi, Dominic Cm Ng, David Meyer, Caiyun Fang, et al. 2014. “Lysine Ubiquitination and Acetylation of Human Cardiac 20S Proteasomes.” *Proteomics. Clinical Applications* 8 (7-8): 590–94.