AN OMICS-BASED INTERROGATION OF DISPARATE MICROBIAL SYSTEMS:

MULTI-OMICS ANALYSIS OF A BIO-MINING ARCHAEON AND

THE EFFECTS OF ARSENIC ON THE *E. COLI* LIPIDOME

by

Hunter Lee Fausset

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

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2023

©COPYRIGHT

by

Hunter Lee Fausset

2023

All Rights Reserved

DEDICATION

This dissertation is dedicated to my parents. Without your unwavering support, I would be nowhere near where I am today.

ACKNOWLEDGEMENTS

My sincerest thanks to Professor Brian Bothner, who is the ultimate mentor in every sense of the word, and to the Bothner lab for maintaining a sense of community through difficult times. To my committee, previous mentors, and all of the teachers through my two plus decades of schooling. Without the support from each of you, none of this would be possible. And finally, to all of the scientists, engineers, and big thinkers that have come before me; we truly are standing on the shoulders of giants.

TABLE OF CONTENTS

1. SYSETMS BIOLOGY "OMICS" INVESTIGATIONS WITH MASS SPECTROMETRY: BACKGROUND, EXPERIMENTAL APPROACHES , AND MODERN APPLICATIONS 1
Omics background: the history and theory behind omics-based scientific thought 1
Introduction 1
Genomics and Transcriptomics: Gateway to the Omics-Era 3
Proteomics: The Translation from Information to Biochemistry 6
Metabolomics, Lipidomics, and other small molecule analyses: Moving towards full phenotypic characterization 11
Multi-Omics and conclusion 15
Mass spectrometry and it's use in omics 16
Overview of Mass Spectrometry Theory 16
MS/MS fragmentation, peptide sequencing, and structural determination 18
References 23

2. A SHIFT BETWEEN MINERAL AND NONMINERAL SOURCES OF IRON AND SULFUR CAUSES PROTEOME-WIDE CHANGES IN M ETHANOSARCINA BARKERI 37
Contribution of Authors and Co-Authors 37
Manuscript Information 39
Abstract 40
Introduction 41
Results 43
Shotgun Proteomics and Statistical Analysis 43
Functional distribution analysis 45
Pairwise Analysis of Conditions 47
Ribosomal tuning 48
Methanogensis and Carbon metabolism 50
Metal Binding Proteins 52
Porphyrin biosynthesis 56
Quorum sensing 57
Discussion 58
Methods 64
Proteomics sample preparation 64
Statistical, functional, and pathway analyses 66
Data Availability 67
Acknowledgments: 67
Figures and Tables 68
References 84

TABLE OF CONTENTS CONTINUED

3. METABOLIC AND MEMBRANE ASSOCIATED CHANGES IN METHANOSARCINA BARKERI FACILITATE GROWTH ON PYRITE AS A SOLE SOURCE OF IRON AND SULFUR 87
Contribution of Authors and Co-Authors 87
Manuscript Information 89
Abstract 90
Introduction 91
Results 93
Global Untargeted Metabolomics and Lipidomics 97
Identifications and Metabolic Interpretation 96
Membrane composition and Lipidomics 101
Discussion 106
Methods 111
Cell cultivation 111
Cell lysis 112
Metabolomics sample preparation 112
Data processing 113
Figures and Tables 115
References 131

4. THE EFFECTS OF ARSENIC EXPOSURE ON THE LIPIDOME OF E.COLI: DETERMINING THE ULTIMATE FATE OF ARSENIC IN A MODEL MEMBER OF THE GUT MICROBIOME 134
Contribution of Authors and Co-Authors 134
Manuscript Information 136
Abstract 137
Introduction 138
Results 140
ICP-MS based arsenic biotracking 140
Global Lipidomics results 141
Intra-strain comparison 143
Discussion 145
Methods 149
Cell culturing and harvesting 149
Macromolecular As distribution 150
ICP-MS acquisition 150
LCMS acquisition 151

TABLE OF CONTENTS CONTINUED

Data processing and statistics 150
Identification pipeline 151
Figures and Tables 152
References 164

5. CONCLUDING REMARKS 170

6. REFERENCES CITED 174

LIST OF TABLES

Table Page

Chapter 2
1. Significantly Dysregulated proteins in various KEGG biochemical pathways 75

2. Ribosomal proteins showing significant dysregulation
between Fe(II)Cys vs Fe(II)/HS- 76

Chapter 3
1. All annotated PA, PI, PG, and PC 130

Chapter 4
1. Assigned identities to the features from the top 20 heatmap 159

2. Assigned identities from top 40 heatmaps (Aw3110 & K12).....163

LIST OF FIGURES

Figure Page

Chapter 1
1. Overview of omics techniques in systems biology 2

2. Overview schematic of a mass spectrometer 17

3. ESI+ MS/MS spectrum of arsenobetaine 19

4. Partial MS/MS spectrum of a peptide 20

5. Protein sequence of the methanol methyltransferase corrinoid protein 21

6. Top 200 heatmap (filtered by p-value) from the Methanosarcina barkeri metabolomics analysis 22

Chapter 2
1. Multivariate statistical analysis of proteomics data 68

2. Parallel hierarchical clustering (PHCA) heatmap of the top 35 proteins 69

3. S1 Box and whisker plots corresponding to the top 35 proteins 70

4. S2 PLSDA VIP plot of the top 15 proteins, filtered by VIP score 71

5. Proteomic functional distribution and biological process analyses 72

6. S3 Whole dataset functional distribution characterization 73

7. S4 Two-group proteomic statistical analyses. 74

8. Ribosomal protein and transcript regulatory patterns 77

9. Regulation of central carbon metabolism iron-sulfur binding proteins 78

10. S5 Central carbon metabolism transcriptional regulation 79

11. Metal binding protein differential analysis 80

12. S6 Hierarchical clustering heatmap of all iron-sulfur binding proteins 81

13. S7 Regulation of key metal binding proteins 82

LIST OF FIGURES CONTINUED

Figure

Page

14. S8 Regulatory patterns of the mRNA transcripts and proteins....83

Chapter 3

1. Global Multivariate statistics summary....115

2. S1 Two-group global multivariate statistics....116

3. S2 Two-group global multivariate statistics....117

4. Dysregulation of cysteine biosynthesis metabolites and related proteins....118

5. Methionine and sulfoethylcysteine dysregulation....119

6. Global multivariate statistics for membrane composition lipidomics....120

7. S5 Membrane lipidomics two group comparison....121

8. Membrane lipidomics two group comparison, FeS_2 vs Fe(II)/Cys....123

9. Dysregulation of identified Archaeol lipids....124

10. S9 Regulatory patterns of Archaeol synthesis proteins....125

11 Phosphoglycerolipid (PGL) dysregulation....126

12. S8 Abundance patterns of PC GPLs....127

13. Cysteinyl- carotenoid like lipids and their dysregulatory patterns....129

14. S9: Non-cysteine containing lipid//nonpolar molecule ID hits and their abundances....130

LIST OF FIGURES CONTINUED

Figure Page

Chapter 4

1. ICP-MS quantification of As in K12 *E.coli* macromolecular fractions.....152

2. All group global lipidomics multivariate analyses 153

3. S1 All feature PHCA heatmap.....154

4. S2: Top 20 PCHA heatmap.....155

5. Intra-strain global multivariate analyses.....157

6. Top 40 heatmap for the AW3110 intra-strain comparison 158

7. Top 40 heatmap for the K12 intra-strain comparison e 159

8. Dysregulated features in the AW3110 comparison 161

9. Dysregulated features in the K12 comparison.....162

10. Potential bio-arsenical features identified.....163

ABSTRACT

Systems biology represents the next frontier in the elucidation of biochemical mechanisms, disease states, and microorganisms. Rather than approaching individual parts of an organism, such as a specific protein, molecule, or mRNA, a systems biology or "omics" investigation seeks to characterize all proteins, molecules, or RNA simultaneously. This is crucial, because all macromolecules in a lifeform exist in dynamic equilibria with those around them; no one biological process occurs in a vacuum. Omics investigations have ballooned in usage over the last decades due to scientists realizing their power in characterizing complex biological phenomena. This has also been spurred on by advances in technologies enabling the robust elucidation of thousands of molecules at once, particularly benefitting from the modernization of mass spectrometry. This technique can be used to study any number of biological problems including those presented here; a multi-omics investigation into a mineral-eating methanogen and a lipidomic characterization of arsenic exposure in a key member of the gut microbiome, *E.coli*.

Methanosarcina barkeri, a widespread methanogen found in marine sediments, is able to reductively dissolve minerals such as pyrite (FeS2) to satisfy their iron and sulfur requirements. Presented here are two investigations containing transcriptomic, proteomic, metabolomic, and lipidomic analyses, performed in parallel on the same biomass. Together, these experiments suggest that the organism undergoes a significant phenotypic shift in response to changes in just two elements, Fe and S. Overall inferences are echoed in the small molecule analyses; the metabolomes and lipidome of the organism change similarly in to the proteome. Key sulfur equilibria are implied in the process, as are specific lipids, choline, and dethiobiotin.

A similar approach was applied to *E.coli* treated with arsenic, as a proxy for understanding the detoxifcation that takes place in the gut microbiome after ingestion. Marked lipidomic changes were observed in *E.coli* resulting from treatment, which were dependent both on species of arsenic as well as presence of the Ars operon. As a foundational study, this work answered some and generated many more hypotheses on the biochemical fate of As in microorganisms in the gut microbiome

SYSETMS BIOLOGY "OMICS" INVESTIGATIONS WITH MASS SPECTROMETRY: BACKGROUND, EXPERIMENTAL APPROACHES, AND MODERN APPLICATIONS

Omics background: the History and Theory Behind Omics-Based Scientific Thought

Introduction:

Omics analyses associated with Systems Biology universally seek to provide a comprehensive look at biological systems from a global perspective (1). Investigations of this genre aim to include many perspectives at once, as opposed to traditional biological experimentation which can be more reductionist. This approach is regularly necessary; biological systems are endlessly nuanced and scientific reasoning requires as much control around experimental variables as possible. The advent and development of modern technologies such as high-speed computing, precise and micro-scale electronics, precision measurement and machining, as well as an advanced globalized economy providing access to more solvents, reagents, and scientific tools than ever before have allowed for modern techniques to trend towards combinatorial, system spanning investigations. This is as true for omics as it is for other fields: computational simulations are enhancing or replacing many traditional experiments in physics and engineering, ground penetrating LIDAR is revolutionizing anthropology, and more than a dozen complex statistical evaluations, just one of which would have made up the bulk of a previous graduate thesis, are contained in this very document.

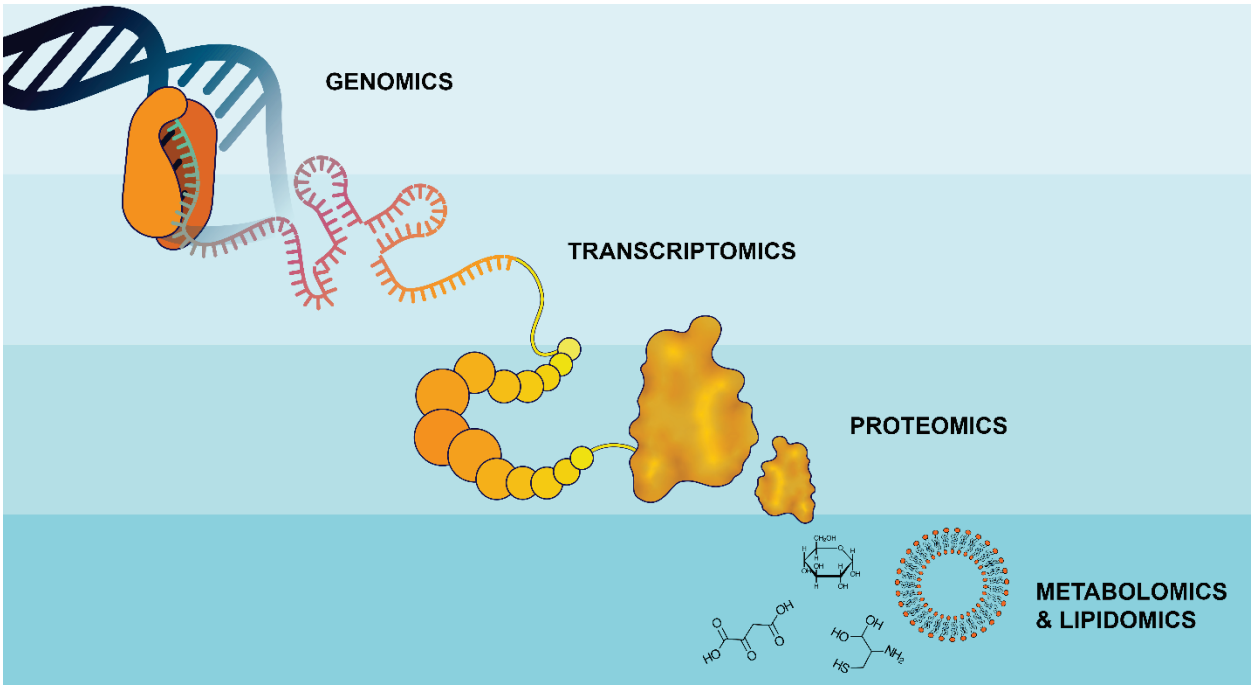


Figure 1: Overview of omics techniques in systems biology. Each level (e.g. Genomics) represents a different omics-approach that together comprise systems biology.

The onslaught of technological advancement has allowed for experimental thinking that mirrors the hierarchical nature of biomolecules and biochemistry. The familiar "central dogma of biology" defines this hierarchy: genes that reside within the DNA of an organism are first transcribed into another nucleic acid, mRNA, from which the proteins that carry out much of the biochemical reactions are translated. Systems biology has evolved to characterize each level of this biochemical stratification with techniques such as genomics, transcriptomics, and proteomics (1). The field has also extended beyond these into the respective downstream products of protein biochemistry: the metabolites, lipids, and other small molecules that make up the phenotype of a given organism (2). Each -omic analysis investigates all biomolecules of a given class and how they relate to one another and the system as a whole. The true power of systems biology lies in the combination of these techniques into integrated multi-omic systems biology approaches. By

assessing a biological process or phenomenon from all angles, the strength of inferences is increased allowing for a more complete understanding of the system at hand.

Genomics and Transcriptomics: Gateway to the Omics-Era

The creation of genomics represents the true beginning of systems biology-based investigations. It was first explicitly discussed as a scientific technique only after being coined by Dr. Thomas H. Roderick over drinks in Bar Harbor, ME in 1986, more than 13 years after the first gene was successfully sequenced (3–5). The term was originally created as a name for an up-and-coming journal he and his colleagues would eventually publish to showcase their work on human genetic material, and it still exists today as the journal Genomics. Roderick and his colleagues used the term to encapsulate their work in mapping all of the genes encapsulated in human cells, our *genome*.

A genome describes all of the genetic material contained within an organism. This includes the primary sequence of the nucleotides AGC and T as well as how they combine together to create genes, and information about the prevalence of those genes. *Gene* itself can have different definitions depending on who is asked, but is generally either based on biological heredity or downstream function; Mendelian genes are based on hereditary transmission, while a molecular gene is a sequence of DNA that can be transcribed into RNA (6–8). Regardless of the definition, the important concept for the sake of this dissertation is that it is a fundamental, describable unit. A *gene* is the distinct, measurable unit of a *genome* that can be used in a *genomic* analysis or comparison. A genome is often referred to in terms of the number of different genes it contains, how many copies of any given gene there are, or if there are any observed intra-gene sequence variations (5). For example, the genetic material of *Homo sapiens*

encodes 20,000 genes, rice has 41,000 genes, and the virus chicken pox possesses just 70 gene units (9). The number of genes represents a simple but pertinent example of what genomics fundamentally investigates. In this case, the genomes of different organisms are being compared based on the number of gene units (transcribable sequences of DNA) in their genetic material. This example also highlights how genomics, and all other types of omics analysis, are comparative in nature.

The major use case for modern genomics lies in the comparison of gene sets to one another or the analysis of mutation patterns in order to answer scientific questions. Scientists can and have used this approach to answer fundamental questions about biology, ranging from the first studies establishing the presence of genes, genomes, and chromosomes to today's consumer genomics platforms describing personal ancestry and potential health predispositions.

One extremely useful application of genomic technology is predictive diagnostics. A major driving force behind scientific development has always been understanding human disease, and the genome holds a wealth of information about disease risk and predispositions. Since its creation, genomics has been used to characterize the many correlations between one's genome and their risk to develop diseases. Alzheimer's, Parkinson's, and Huntington's diseases, several cancers, sickle cell anemia, Down syndrome, autism spectrum disorder, and multiple sclerosis (MS) have all been characterized as having some level of genetic predisposition. The latter ailment, MS, has been shown to have more than 100 distinct risk loci in the genome (10, 11). The genomic studies of MS have also suggested that helper T-cell differentiation is likely implicated in disease prognosis, based on the functional proximity of risk loci to related genes

(10). Understanding the relationship between disease states and the human genome has been and will continue to be paramount in the pursuit of better health on a societal scale.

The previous generalization of a static genome is broadly true, especially when comparing healthy cells within an individual, but certain conditions/diseases such as cancer, have been shown to cause specific mutations. Colon and rectal cancers are two prime examples. A high-impact 2012 paper, one to come from the Cancer Genome Atlas project, found 24 specific mutations in the genome of colon and rectal carcinomas relative to healthy cells (12). Genomics in this case, is identifying changes in the genome *within* an organism in order to provide targets for future pharmaceutical therapeutics. The connection between the genome and cancer is now well established, with specific genes, termed *oncogenes*, being identified as particularly risky in relation to the likelihood of developing cancer.

After the characterization of the first genomes, scientists shifted their attention to the next step in the dogma - the transcriptome. The first investigation resembling modern transcriptomics was in 1991, when 609 mRNA transcripts were (indirectly) sequenced from human brain cells (13). This represented a large breakthrough for future omics techniques.

DNA and RNA are both nucleotides and are characterized similarly through sequencing techniques, but the eventual inferences drawn from genomic and transcriptomics investigations are dissimilar. The genome is generally believed to be static for a given organism. Disregarding germ cells, all of the (non-cancerous) cells within an organism share the same genetic code; human hepatocytes, macrophages, epithelial, and neuronal cells alike all stem from the same DNA sequence. The transcriptome, in contrast, is extremely dynamic. Differences in DNA transcription underpin some of the phenotypic differences in the aforementioned cell types, for

example (14–17). An understanding of the transcriptome therefore doesn't just unlock another tier of the biochemical hierarchy; it also allows for comparisons within an organism in different biological states, between similar lifeforms, or for interrogations into the transcriptional mechanism itself. After its advent, advances in sequencing technology allowed scientists to evaluate the effects of stress and other stimuli on cells, evaluate disease states and how they impact transcriptional regulation (breast cancer paper), or better understand the process of transcription itself (18–20). By evaluating the abundance an entire class of biomolecules (in this case, mRNA), the field was able to make previously hidden inferences and identify key biological phenomena.

The introduction of the concepts of genome and transcriptome and their respective scientific studies represented a leap forward for systems biology research. New technologies had allowed scientists to probe entire classes of biomolecules at once and take the first comprehensive look at specific ongoings within an organism. The Nobel Prize - winning publication of the human genome in 2001 (21–23) and all subsequent findings from comparative genomics studies certainly represent large contributions to human progress on their own, but the inception of global biomolecule analysis into the scientific zeitgeist and subsequent invention of proteomics, metabolomics, lipidomics, and a host of other omics techniques continue to revolutionize the world to this day.

Proteomics: The Translation from Information to Biochemistry

The translation from mRNA to proteins represents a major fundamental shift in the central dogma of biology, in terms of the biochemical macromolecules themselves, their diversity in chemistry and localization, as well as the experimental considerations for their

analysis. Nucleic acids are the combination of just 5 building blocks (AGTC and U), whereas proteins are a sequence that contains some combination of 20 amino acids (AA). Proteins also differ from nucleic acids in that their primary purpose is to carry out the many life-giving chemical reactions contained within a cell (ribozymes being the major exception) (24, 25). These are the machines of life, built to carry out specific functions, be it enzymatic catalysis, structural support, nucleic acid replication, or other bio-syntheses. This diversity of function highlights the importance of understanding the dynamics of protein abundance and post translational modifications in biological science (26–28).

The scientific community has long understood the importance of protein structure and function. Early protein biochemists recognized their importance in nutrition, but primitive isolation and purification protocols limited their studies to abundant proteins in animal biproducts, blood, and egg-whites. In 1951, the first protein sequence was assigned to insulin, a lifesaving peptide for millions of diabetics worldwide, by Frederick Sanger (29). Proteins, of course, do not exist in a biochemical vacuum; their ultimate function is typically tied to the proteins surrounding them in biochemical pathways. This was recognized by scientists in the mid-1970's who sought to characterize numerous proteins from a biological system simultaneously. Patrick H. O'Farrell did so first in his application of 2-dimensional gel electrophoresis to proteins in the bacterium *Escherichia coli* in 1975 (30, 31). This new iteration of GE separated proteins in two-dimensions (2D-GE), providing a map of a cellular proteome with proteins separated by both size and isoelectric point (PI), a metric of the molecules charge-state tendency. This analysis was ahead of its time, however, as there was no good way for the scientists to *identify* the proteins they were looking at on their newfound proteomic maps. Edman

degradation was the contemporary protein sequencing technology but had low sensitivity, which disallowed widespread use in conjunction with 2D-GE (32–35). Thus, scientists at this time could visualize a proteome but had no good way of properly interrogating it. It would take the advent of modern technologies, such as mass spectrometry, for proteomics as a field to come into its own.

The modernization of proteomics followed developments in scientific technology, benefiting particularly from advances in mass spectrometry (MS) (36). The turn of the century saw new applications of this established technique, which allowed for identification of proteins from 2D-GE analyses through the sequencing of peptides resulting from digestion with a protease (usually trypsin) (37, 38). Using 2D-GE in combination with matrix assisted laser desorption ionization-mass spectrometry (MALDI-MS), Meisinger et al. mapped the *Saccharomyces cerevisiae* mitochondrial proteome, the first characterization of the protein fingerprint of a eukaryotic organelle (39). Now a widespread technique, this was among the first so-called “shotgun proteomics” investigations, named due to the manner in which the proteins are fragmented prior to analysis. The mapping of the mitochondrial proteome led to a new understanding of the processes involved in trafficking key enzymes from the cytosol into the powerhouse of the cell (40). Similar studies in this mid-2000s renaissance saw the first proteome maps of chloroplast membranes, human carcinomas, bodily fluids, and numerous microorganisms. (41–48).

Gel-free protein separation techniques also began to see the light during this period and constitute the majority of shotgun proteomics-based approaches today. A different separation technique, usually liquid chromatography (LC), is still needed to reduce the chemical complexity

of proteomic samples, but the principal concept remains the same (49, 50). Moving away from gel-based separation increased resolution and subsequent genome coverage (i.e., the number of potential proteins, as predicted in the genome, that are identified). Modern shotgun proteomics analyses track thousands to tens of thousands of proteins, their abundances, and variations in post translational modification. Compare this to previously cited gel-based analyses whose coverages max out in the low thousands, and average much lower. The ability for scientists to monitor this number of proteins, often times representing the vast majority of the entire genome, has aided in the characterization of several cancers, differences between tissues, and the extreme effects of SARS-CoV-2 on human cells.

One of the most cited proteomics studies in the 2010s was the Tissue-Based Map of the Human Proteome, published in 2015 by Uhlén and Pontén et al. (51). One of the most surprising findings was the relative ubiquity of proteins across human tissues; over 30% of proteins targeted by clinically approved therapeutics are expressed in each of the 42 tissues they examined. Similarly, half of all genes covered in the analysis (over 90% of the predicted protein-coding genome) are expressed in every tissue. Their data is available in an interactive format in the Human Protein Atlas, the proteomic analog to the Human Genome Atlas discussed above (52–60). Further studies have been published and are underway to increase the resolution of the proteomic map to single and subcellular granularity (61). Two concurrent, large-scale cancer studies utilized proteomics in a similar way to characterize therapeutic targets for hepatocellular carcinomas (62). Another study around this time characterized the early stages of Alzheimer disease, implicating energy metabolism in the microglia and astrocyte activation associated with disease severity (63). Already a burgeoning field at the time of the COVID-19 pandemic,

proteomics was well suited to characterize the cellular effects of viral infection and help clinicians understand exactly what was happening in their patients. Early studies discovered host proteome changes associated with viral infection in immune and epithelial cells, providing potential targets for future pharmaceuticals (64). Incredibly, by June, 2021 over 100 proteomic peer-reviewed studies on the virus and its interaction with human cells had been published or were in preprint (65). As an aside, it was incredible to witness the power of collective agreement and prioritization in the intellectual community. In the opinion of the author, future scientific policy should be shaped around what was gleaned from the pandemic. Subsequent proteomic studies fully mapped the protein-protein interactions (the "contactome") between viral and host proteins as well as proteomic contributions to long-COVID and overall disease severity (66–71). It is clear that proteomics is here to stay and will become commonplace in the continued elucidation of human disease in the coming decades.

Shotgun proteomics is now a mature technique. It has been and will remain a powerhouse well suited for the full characterization of systems biology. Each year sees further advancement in the instrumentation and computation technology requisite for a robust proteomics investigation. This is all well and good, but the characterization of protein abundances, post-translational modifications, and regulatory patterns still only represents a single class of biological macromolecule. Just like the first endeavors into protein biochemistry were limited by the scope of an isolated protein, modern shotgun proteomics can only provide inferences from the perspective of protein abundance. That all changes when this technique is combined with one in the suite of omics-techniques tailored to biological small molecules.

Metabolomics, Lipidomics, and other small molecule analyses:
Moving towards full phenotypic characterization

While proteomics is a well-established technique, small-molecule-omics such as metabolomics and lipidomics are still gaining momentum. This is due to many aspects, but ultimately originates in the physiochemical diversity inhabited by biological small-molecules. Nucleotides and proteins, while disparate in many ways, follow predictable formulas. The bases in nucleotides and the amino acids in proteins are all connected in a chemically consistent manner, allowing for their eventual sequencing and analysis. Small molecules follow no such paradigm. While some small molecule classes may behave similarly, such as hydrocarbon acyl chains, there is no fundamental pattern across the entire classification. Caffeine, cardiolipin, ricin, retinol, penicillin, and proline alike contribute to the biological systems where they originate but share very little in their physiochemical properties. This presents a challenge in that the phenotype of any given lifeform largely depends on its small-molecule profile. Many of the proteins discussed earlier, for example, exist solely to carry out conversions from one small molecule to another. All of these considerations have contributed to the slow advancement of small moleculeomics relative to proteomics and genomics. Until recently, there simply was not a methodology for capturing the many thousands of different molecules in a biological system to allow for their simultaneous characterization.

Small-molecule-omics techniques are generally separated by the specific sub-class of molecules they interrogate. Metabolomics, for example, quantifies *metabolites*, a general term for an intermediate or endpoint in a given metabolism. In practice this amounts to the analysis of soluble cytosolic small molecules (~ <1500Da) (72). Lipidomics specifically targets nonpolar hydrocarbons, sometimes focusing on those found in the cellular membrane. Glycomics,

exosomics, fluxomics, exposomics, ionomics, secretomics, and other categories of this field each seek to probe the small molecules of a system from a different perspective. While there is distinct overlap between many of these approaches (metabolomics will often pick up lipids and vice versa), they are individually catered to extract and analyze their respective molecules in the most efficient manner. This is usually prevalent throughout the entirety of the experimental process, from extraction protocols through acquisition and subsequent analysis. It would be arduous and irrelevant to detail each of the existing small molecule-omics subclasses. Therefore, the remainder of this thesis will focus on the two most common: metabolomics and lipidomics.

Thought processes surrounding metabolic profiles, the fundamental subject matter of metabolomics, have been around for a very long time. Sir Archibald Edward Garrod was a late 19th-early 20th-century physician who linked gout to the accumulation of uric acid, and also characterized specific changes in urine chemistry that led to his discovery of alkaptonuria (73–75). He was one of the first to realize that small molecules in human fluids or tissues are influenced by or can affect specific disease states. Over the course of the following century, this approach, like proteomics, took advantage of advances in MS, but also gained traction in the nuclear magnetic resonance spectroscopy (NMR) communities. NMR's inherent ability for structural characterization of small molecules provided the means for scientists to interrogate metabolism as never before (76). One major study showed that most cellular ATP is coordinated by magnesium in human muscle tissues using ^{31}P NMR (77). It took until 1998 for the term metabolomics to come to light, coined by Oliver Fiehn in his work on plant metabolism(78). The work of a contemporary scientist, Jeremy K Nicholson, also helped to establish the field. The turn of the century brought rapid expansion in the field, a large part of which was due to

advances in LC-MS based approaches led by Gary Siuzdak and others at the Scripps Research Institute(79). Their efforts saw the birth of METLIN, a database containing MS spectra and fragmentation patterns allowing for better identification of metabolites (80, 81).

Lipidomics, in many ways, is like metabolomics, especially when considering its development (82). The main difference in the techniques lies in the extraction protocols used to solubilize the respective compounds. Therefore, lipidomics and metabolomics coevolved during this time period, with reviews such as “The Emerging Field of Lipidomics” seeing publication in the mid-aughts (82). Like metabolomics, MS was the main technique for the systemic analysis of lipids. NMR was also used in early lipidomic-like experimentsv(83). Some early applications of the technique expanded the understanding of apoptosis by correlating lipid oxidation events to cell death (83–85). Brain lipids were also a popular subject for early lipidomics investigations (at least one source), as many neurological pathologies revolve around lipids or lipid-like molecules (86). As with many other techniques, early studies on disease states were followed by an expansion of the application into other biological fields, with modern lipidomics being prevalent in microbiology, nutrition, and neurology.

Modernization of chromatography systems as well as MS electrospray sources and the advent of ion mobility mass spectrometry (IMS) have all contributed to the more widespread use of metabolomics and lipidomics (87, 88). Increases in pressure and subsequent resolution of LC systems allow for better chromatographic separation than ever before. MS detectors are consistently faster and more accurate, ion optics and advanced IMS separations increase specificity, and advances in feature sorting algorithms mean modern metabolomics and lipidomics experiments can track more molecules from biological systems than ever before (87).

In 2019 a large scale metabolomic study taking advantage of these new technologies characterized the relationship between the gut microbiome, specifically the small molecules resulting from microbial metabolism, and autism spectrum disorder (ASD) (89). Germ-free mice were transplanted with microbiota from typical development or autistic donors, which carried with it the transference of autism-like behaviors for the latter group. Using metabolomics, the authors linked specific taxa and their metabolites to ASD in the mouse population(89). Similar recent studies have looked at the microbiome as it pertains to Alzheimer's disease, the ingestion of microplastics, and the effect of dietary fats on cancer progression (90–92). Further evidence that the field of metabolomics and lipidomics is rapidly accelerating can be seen in the number of tools, databases, and tutorials published in the last decade (93–104). Many of these represent the most cited sources over that time period. This also highlights the main shortcoming of this type of analysis: data processing, interpretation, and the generation of biologically relevant inferences. This will be discussed further in the next section, but the main limitation of MS based lipid and metabolomics is the lack of inherent identification. This technique excels at characterizing the metabolic profile from a statistical viewpoint, however, identifying specific molecules and biological inference remains very challenging. There are various ways around this pitfall, ranging from in silico-generated databases, predicted retention times, and fragmentation patterns, to the current paradigm of in-house standard libraries. Overcoming these challenges remains a priority for the metabolomics community and will see great strides in the coming decade, as it remains one of the final missing pieces to true system- wide, multi-omic investigations into biological entities.

Metabolomics, lipidomics, and other holistic views of cellular metabolism have come a long way since their inception. Early studies on gout and urine metabolites laid the groundwork for modern, transformational analyses of autism, cancer, and their relation to the gut microbiome. While often still hamstrung by its lack of interpretability, this type of analysis is accelerating in capability at a brisk clip. The community at large realizes the shortcomings and has made strides to standardize the endeavor. The m/q initiative, led by researchers at Pacific Northwest National Laboratories, is one such effort striving to allow for intrinsic identification of metabolites in current and future MS based approaches (105). This is crucial, as small molecules represent a major part of a true systems-wide analysis. Leading researchers have even called metabolomics the "Apogee of the omics-trilogy", exemplifying its importance for future multi-omics investigations (72). The truth is that a properly robust multi-omics analysis is much more than the sum of its parts. The goal for systems biology is to interrogate the entire system, and one cannot do so without uniting the disparate biological molecules into a single analytical framework.

Multi-Omics and conclusion

A deep scientific investigation into a biological system will take into account the hierarchical structure that exists. Therefore, to fully understand biological phenomena, scientists must cross-examine the genes, transcripts, proteins, and small molecules all at once. This is true for disease states, host-pathogen interactions, and single-celled organisms alike. To achieve what systems biology sets out to do, means the integration of all omics combined into an integrated predictive model. More than simply overlapping datasets or combining inferences made from multiple analyses, it should entail complex statistical operations performed on omics techniques

performed on the same biomass or biological replicate. Many of the studies cited earlier in this document employ more than one type of omics. Currently, this process is often prohibitively expensive or time-consuming, but the future lies in cross-technique integration and combinatory investigations.

Science must match the nuance and complexity of life with sophistication and intricacy of its own. Luckily, this has been happening for the better part of the last two decades, especially with respect to advances in mass spectrometry. The next section will outline modern use cases of mass spectrometry for proteomics, metabolomics, and lipidomics as a primer for understanding the presented scientific investigations.

Mass spectrometry and it's use in omics

Overview of Mass Spectrometry Theory

Many well written reviews on mass spectrometry have already been published by leaders in the field, so this section will serve as a brief introduction to the concepts relevant to the remainder of the dissertation (88, 106–116). A mass spectrometer ultimately measures the mass to charge ratio (m/z) of a charged molecule. The first part of the instrument, the source, handles placing a charge on said molecule. The central parts of the instrument typically contain electronics used to generate electromagnetic fields that manipulate the newly formed ions. This can result in the filtration of unwanted species and the separation of similar ions from one another, but the eventual goal is to pull the charged molecules to the detector at the end of the instrument, where they generate a measurable current. Based on previous calibrations and calculations, the resulting current is measured for a short period of time (a scan) and interpolated to a m/z value, resulting in an MS spectrum. There are hundreds of combinations of source type,

ion optics, mass analyzer, and accessory technologies that are used throughout the scientific community. However, the instruments used for proteomics, metabolomics, and lipidomics generally follow the same analytical framework.

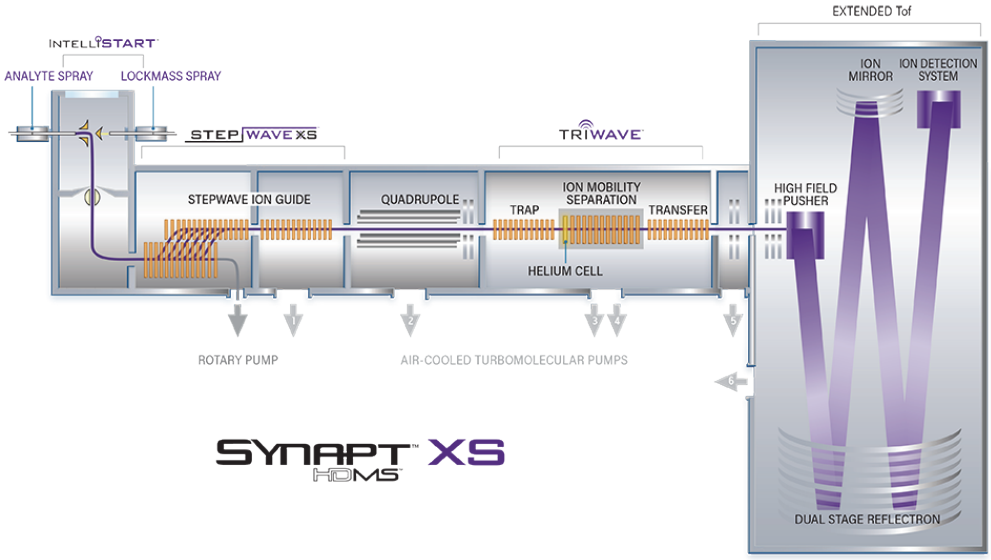


Figure 2: Overview schematic of a mass spectrometer (117).

Broadly speaking, discovery-based omics techniques that utilize MS are performed on time-of-flight (TOF) or orbitrap mass analyzers (118). TOF instruments separate ions by the time it takes them to transverse a given length of space, from which their m/z can be determined. Orbitraps are more complicated, using the precession frequencies of the ions orbiting a centralized charge plate to generate current. Three dimensions of frequency (x,y,z) are measured and Fourier transformed to produce m/z (119), resulting in ultra-high resolution and sensitivity. Gas-chromatography systems are sometimes used in this type of application, but they are less common now than closer to the advent of the technologies. These MS systems are usually simpler, using a quadrupole or octupole to filter ions prior to their detection (120). Each type of MS system has a best use case – TOF systems have higher mass accuracy while orbitraps excel

in resolution – so mass analyzers are often chosen to suit the analysis on hand. Modern proteomics experiments, for example, are typically performed on orbitrap analyzers. Just like mass analyzers, different sources are better suited for different applications, though the majority of proteomics, metabolomics, and lipidomics utilize electrospray ionization as their ionization source.

Electrospray ionization (ESI) is the gold standard for modern day omics, particularly those that also employ LC as a separation technique (121). This source operates on the principle of solvent-to-analyte charge transfer. Solvent, carrying molecules of interest, is sprayed through a charged capillary, forming nanodroplets. Droplets are then evaporated by hot, inert gas, causing charge to transfer from solvent to molecule while simultaneously launching the analytes into the gas phase. These charged, gas phase ions then enter the MS for further analysis. ESI is a "soft" ionization technique, gently coaxing molecules into a charged gaseous state. This, when performed correctly, maintains their original structure. This is even true for large protein complexes and viruses, which have been shown to be ionizable using this technique. The ability to maintain molecular structures in the gas phase is the main reason why ESI is so prevalent in systems biology today. Along with maintaining the original mass of the molecule for analysis, soft ionization also allows energy to be applied selectively to the molecules post-ionization. This is the basis of MS/MS, also called fragmentation, and will be outlined further in the section on peptide sequencing.

MS/MS fragmentation, peptide sequencing, and structural determination

A key capability of MS systems is tandem mass spectrometry, also called MS/MS. Here, ions in the gas phase are deliberately collided with inert gas at a given energy to induce breakage

of their chemical bonds(122). This occurs in a collision cell, a compartment in a MS flooded with N_2 or He. MS/MS allows users to garner more information about the chemical species than just their mass. The approach is used in proteomics to attain the amino acid sequence of peptides and in metabolomics to identify the structure of the analytes in question. An example of this is seen in the MS/MS spectrum of arsenobetaine, a common arsenical found in marine animals (Figure 1).

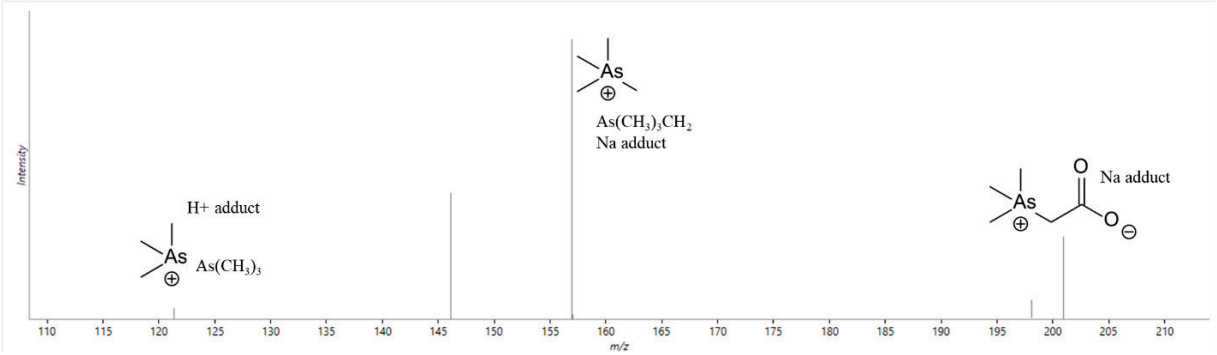


Figure 3: ESI+ MS/MS spectrum of arsenobetaine acquired on a TOF mass spectrometer. Shown are the structures of arsenobetaine and the fragment ions used to confirm its identity.

This type of ion manipulation is a core tenet of peptide sequencing in proteomics experiments. After digestion of cellular proteins with trypsin, a sample may contain tens to hundreds of thousands of individual tryptic peptides(36, 87). As discussed before, these peptides follow a regular pattern in their chemistry. Individual AA are linked by peptide bonds, and serial fragmentation of a peptide yields informative MS/MS spectra that allow for sequencing. A peptide can fragment at many chemical bonds, but at a consistent collision energy (the kinetic energy held by ions prior to their collision) they regularly break at the peptide bond(123). Shown in Figure 2 is a spectral representation of this phenomena. When a peptide bond fragments, it generates one of two ions; one with the charge on the carboxylic acid (y ions, blue) and one with

the charge on the amino acid (b ions, red). Since the only difference between the AA in a sequence is their respective R group, a peptide can be sequenced by measuring the mass difference between successively larger fragments of the peptide.

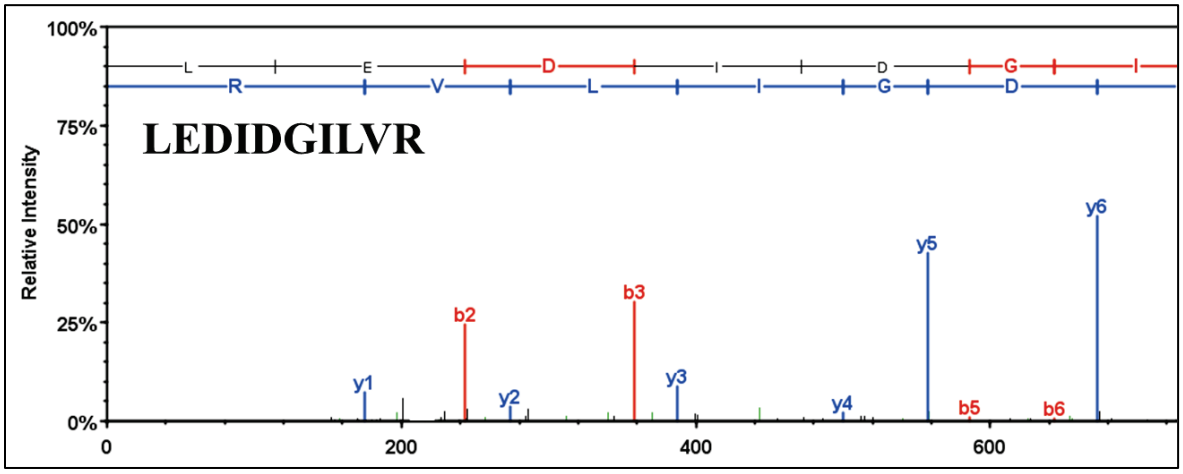


Figure 4: Partial MS/MS spectrum of the peptide with an AA sequence of LEDIDGILVR. Y ions are shown as blue peaks and b ions as red peaks. Mass differences resulting from AA R groups are shown across the top in the color corresponding to their ion type (y vs. b). y1 corresponds to the AA arginine, y2 corresponds to the dipeptide arginine + valine, y3 is RVL, etc..

The process of peptide sequencing is repeated for all detectible tryptic peptides from a sample. This generates a peptide library that is used to identify proteins. In brief, sequenced peptides are compared to a reference proteome (FASTA sequence all potential proteins, taken from the ORFs in the genome) from the organism in question. Proteins are then identified if they have proper coverage, i.e a large proportion of their sequence covered by measured peptides. The protein coverage map containing peptide outlined in Figure 2 can be seen in Figure 3(36). After a protein has been identified, the abundances of the top 3 most confident unique peptides (i.e. peptides not found in other proteins in the reference genome) are used as a proxy for the protein abundance. This is the value that generates all subsequent fold change (FC), significance (p-value), and other characteristics used to interpret the proteomics analysis.

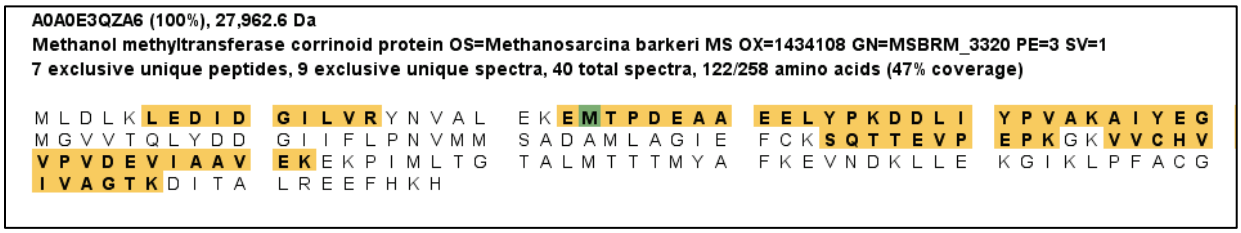


Figure 5: Protein sequence of the methanol methyltransferase corrinoid protein found in *Methanosarcina barkeri*. Peptides that were successfully sequenced in the proteomics experiment are highlighted in yellow. Information regarding protein identity such as coverage, total spectra, and unique peptides are listed below the protein name. Shown is a truncated version of the full protein sequence for clarity.

MS/MS fragmentation is also used in the metabolite ID phase of metabolomics investigations. One of the strengths of MS lies in its sensitivity, specificity, and overall coverage of metabolites from a biological system. A major weakness, as discussed earlier, is the lack of intrinsic identification. Figure 4 shows a typical statistical output from a metabolomics experiment. Here, the metabolomes of a methanogen grown in different growth conditions are compared. This is a top 200 heatmap, meaning only 200 of the many thousand MS features are present. It is clear, from this figure and those discussed in the relevant chapter, that there are widespread metabolic changes between groups. The trouble lies in the fact it is not immediately evident what each of the features, rows in the heatmap, actually are. They could be TCA intermediates, amino acids, vitamins, cofactors, or any of the yet-to-be characterized molecules in the organism. It is for this reason why MS/MS fragmentation must be used to identify features of interest.

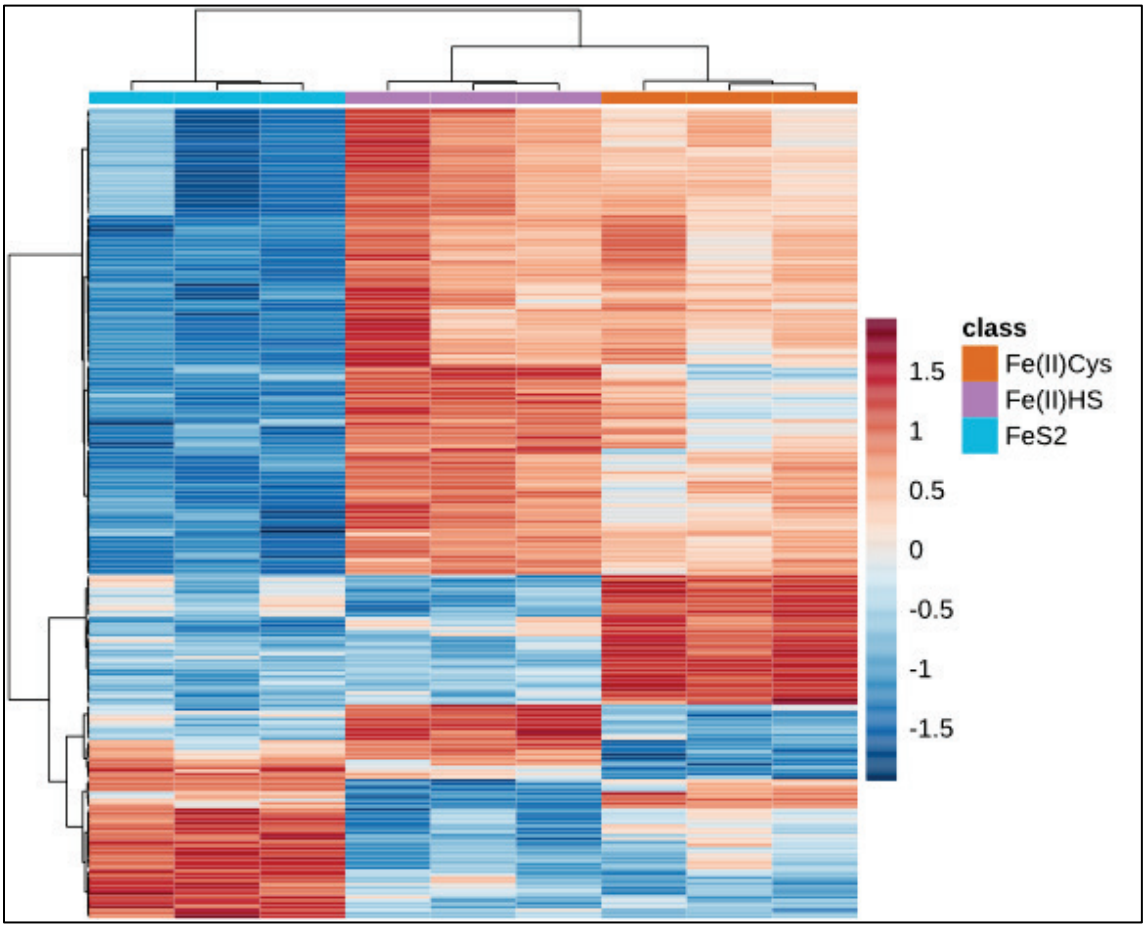


Figure 6: Top 200 heatmap (filtered by p-value) from the *Methanosarcina barkeri* metabolomics analysis found in Chapter 3. Columns represent individual samples, in this case biological replicate cultures of *M.barkeri*. Rows are MS features (m/z value). The intersection of rows and column contains information about the abundance of that feature in each group; red is above average and blue below average.

All things considered, mass spectrometry is well suited for systems wide analyses such as proteomics, metabolomics, and lipidomics. It has its pitfalls, but the incredible sensitivity, selectivity, versatile sample introduction, and detection speed found in modern mass spectrometers lead them to be the best option for systems analysis moving forward.

References

1. Horgan RP, Kenny LC. 2011. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician & Gynaecologist 13:189–195.

2. Fiehn O. 2002. Metabolomics — the link between genotypes and phenotypes. Functional Genomics 155–171.

3. Kuska B. 1998. Beer, Bethesda, and Biology: How "Genomics" Came Into Being. JNCI: Journal of the National Cancer Institute 90:93–93.

4. Yadav SP. 2007. The Wholeness in Suffix -omics, -omes, and the Word Om. J Biomol Tech 18:277.

5. Heather JM, Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107:1.

6. Orgogozo V, Peluffo AE, Morizot B. 2016. The "Mendelian Gene" and the "Molecular Gene": Two Relevant Concepts of Genetic Units. Curr Top Dev Biol 119:1–26.

7. Genetics. https://www.nigms.nih.gov/education/fact-sheets/Pages/genetics.aspx. Retrieved 4 March 2023.

8. 1909: The Word Gene Coined. https://www.genome.gov/25520244/online-education-kit-1909-the-word-gene-coined. Retrieved 4 March 2023.

9. Introduction to Genomics. https://www.genome.gov/About-Genomics/Introduction-to-Genomics. Retrieved 9 April 2023.

10. Sawcer S, Hellenthal G, Pirinen M, Spencer CCA, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Oturai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S, Giannoulatou E, D'Alfonso S, Blackburn H, Boneschi FM, Liddle J, Harbo HF, Perez ML, Spurkland A, Waller MJ, Mycko MP, Ricketts M, Comabella M, Hammond N, Kockum I, McCann OT, Ban M, Whittaker P, Kemppinen A, Weston P, Hawkins C, Widaa S, Zajicek J, Dronov S, Robertson N, Bumpstead SJ, Barcellos LF, Ravindrarajah R, Abraham R, Alfredsson L, Ardlie K, Aubin C, Baker A, Baker K, Baranzini SE, Bergamaschi L, Bergamaschi R, Bernstein A, Berthele A, Boggild M, Bradfield JP, Brassat D, Broadley SA, Buck D, Butzkueven H, Capra R, Carroll WM, Cavalla P, Celius EG, Cepok S, Chiavacci R, Clerget-Darpoux F, Clysters K, Comi G, Cossburn M, Cournu-Rebeix I, Cox MB, Cozen W, Cree BAC, Cross AH, Cusi D, Daly MJ, Davis E, De Bakker PIW, Debouverie M, D'Hooghe MB, Dixon K, Dobosi R, Dubois B, Ellinghaus D, Elovaara I, Esposito F, Fontenille C, Foote S, Franke A, Galimberti D, Ghezzi A, Glessner J, Gomez R, Gout O, Graham C, Grant SFA, Guerini FR, Hakonarson H, Hall P, Hamsten A, Hartung HP, Heard RN, Heath S, Hobart J, Hoshi M, Infante-Duarte C, Ingram G, Ingram W,

Islam T, Jagodic M, Kabesch M, Kermode AG, Kilpatrick TJ, Kim C, Klopp N, Koivisto K, Larsson M, Lathrop M, Lechner-Scott JS, Leone MA, Leppä V, Liljedahl U, Bomfim IL, Lincoln RR, Link J, Liu J, Lorentzen AR, Lupoli S, MacCiardi F, MacK T, Marriott M, Martinelli V, Mason D, McCauley JL, Mentch F, Mero IL, Mihalova T, Montalban X, Mottershead J, Myhr KM, Naldi P, Ollier W, Page A, Palotie A, Pelletier J, Piccio L, Pickersgill T, Piehl F, Pobywajlo S, Quach HL, Ramsay PP, Reunanen M, Reynolds R, Rioux JD, Rodegher M, Roesner S, Rubio JP, Rückert IM, Salvetti M, Salvi E, Santaniello A, Schaefer CA, Schreiber S, Schulze C, Scott RJ, Sellebjerg F, Selmaj KW, Sexton D, Shen L, Simms-Acuna B, Skidmore S, Sleiman PMA, Smestad C, Sørensen PS, Søndergaard HB, Stankovich J, Strange RC, Sulonen AM, Sundqvist E, Syvänen AC, Taddeo F, Taylor B, Blackwell JM, Tienari P, Bramon E, Tourbah A, Brown MA, Tronczynska E, Casas JP, Tubridy N, Corvin A, Vickery J, Jankowski J, Villoslada P, Markus HS, Wang K, Mathew CG, Wason J, Palmer CNA, Wichmann E, Plomin R, Willoughby E, Rautanen A, Winkelmann J, Wittig M, Trembath RC, Yaouanq J, Viswanathan AC, Zhang H, Wood NW, Zuvich R, Deloukas P, Langford C, Duncanson A, Oksenberg JR, Pericak-Vance MA, Haines JL, Olsson T, Hillert J, Ivinson AJ, De Jager PL, Peltonen L, Stewart GJ, Hafler DA, Hauser SL, McVean G, Donnelly P, Compston A. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011 476:7359 476:214–219.

11. Baecher-Allan C, Kaskow BJ, Weiner HL. 2018. Multiple Sclerosis: Mechanisms and Immunotherapy. Neuron 97:742–768.

12. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, Kovar CL, Lewis LR, Morgan MB, Newsham IF, Reid JG, Santibanez J, Shinbrot E, Trevino LR, Wu YQ, Wang M, Gunaratne P, Donehower LA, Creighton CJ, Wheeler DA, Gibbs RA, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, Stojanov P, McKenna A, Lander ES, Gabriel S, Ding L, Fulton RS, Koboldt DC, Wylie T, Walker J, Dooling DJ, Fulton L, Delehaunty KD, Fronick CC, Demeter R, Mardis ER, Wilson RK, Chu A, Chun HJE, Mungall AJ, Pleasance E, Gordon Robertson A, Stoll D, Balasundaram M, Birol I, Butterfield YSN, Chuah E, Coope RJN, Dhalla N, Guin R, Hirst C, Hirst M, Holt RA, Lee D, Li HI, Mayo M, Moore RA, Schein JE, Slobodan JR, Tam A, Thiessen N, Varhol R, Zeng T, Zhao Y, Jones SJM, Marra MA, Bass AJ, Ramos AH, Saksena G, Cherniack AD, Schumacher SE, Tabak B, Carter SL, Pho NH, Nguyen H, Onofrio RC, Crenshaw A, Ardlie K, Beroukhim R, Winckler W, Meyerson M, Protopopov A, Hadjipanayis A, Lee E, Xi R, Yang L, Ren X, Sathiamoorthy N, Chen PC, Haseley P, Xiao Y, Lee S, Seidman J, Chin L, Park PJ, Kucherlapati R, Todd Auman J, Hoadley KA, Du Y, Wilkerson MD, Shi Y, Liquori C, Meng S, Li L, Turman YJ, Topal MD, Tan D, Waring S, Buda E, Walsh J, Jones CD, Mieczkowski PA, Singh D, Wu J, Gulabani A, Dolina P, Bodenheimer T, Hoyle AP, Simons J V., Soloway M, Mose LE, Jefferys SR, Balu S, O'Connor BD, Prins JF, Chiang DY, Neil Hayes D, Perou CM, Hinoue T, Weisenberger DJ, Maglinte DT, Pan F, Berman BP, Van Den Berg DJ, Shen H, Triche T, Baylin SB, Laird PW, Getz G, Noble M, Voat D, Gehlenborg N, Dicara D, Zhang J, Zhang H, Wu CJ, Liu SY, Shukla S, Zhou L, Lin P, Park RW, Nazaire MD, Robinson J, Thorvaldsdottir H, Mesirov J, Thorsson V, Reynolds SM, Bernard B, Kreisberg R, Lin J, Iype L, Bressler R, Erkkilä T, Gundapuneni M, Liu Y, Norberg A, Robinson T, Yang D, Zhang W, Shmulevich I, De Ronde JJ, Schultz N, Cerami E, Ciriello G, Goldberg AP, Gross B, Jacobsen A, Gao J, Kaczkowski B, Sinha R, Arman Aksoy B, Antipin Y,

Reva B, Shen R, Taylor BS, Ladanyi M, Sander C, Akbani R, Zhang N, Broom BM, Casasent T, Unruh A, Wakefield C, Hamilton SR, Craig Cason R, Baggerly KA, Weinstein JN, Haussler D, Benz CC, Stuart JM, Benz SC, Zachary Sanborn J, Vaske CJ, Zhu J, Szeto C, Scott GK, Yau C, Ng S, Goldstein T, Ellrott K, Collisson E, Cozen AE, Zerbino D, Wilks C, Craft B, Spellman P, Penny R, Shelton T, Hatfield M, Morris S, Yena P, Shelton C, Sherman M, Paulauskis J, Gastier-Foster JM, Bowen J, Ramirez NC, Black A, Pyatt R, Wise L, White P, Bertagnolli M, Brown J, Chan TA, Chu GC, Czerwinski C, Denstman F, Dhir R, Dörner A, Fuchs CS, Guillem JG, Iacocca M, Juhl H, Kaufman A, Iii BK, Van Le X, Mariano MC, Medina EN, Meyers M, Nash GM, Paty PB, Petrelli N, Rabeno B, Richards WG, Solit D, Swanson P, Temple L, Tepper JE, Thorp R, Vakiani E, Weiser MR, Willis JE, Witkin G, Zeng Z, Zinner MJ, Zornig C, Jensen MA, Sfeir R, Kahn AB, Chu AL, Kothiyal P, Wang Z, Snyder EE, Pontius J, Pihl TD, Ayala B, Backus M, Walton J, Whitmore J, Baboud J, Berton DL, Nicholls MC, Srinivasan D, Raman R, Girshik S, Kigonya PA, Alonso S, Sanbhadti RN, Barletta SP, Greene JM, Pot DA, Shaw KRM, Dillon LAL, Buetow K, Davidsen T, Demchok JA, Eley G, Ferguson M, Fielding P, Schaefer C, Sheth M, Yang L, Guyer MS, Ozenberger BA, Palchik JD, Peterson J, Sofia HJ, Thomson. E. 2012. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012 487:7407 487:330–337.

13. Adams MD, Kelley JM, Gocayne JD, Dubnick MAK, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC. 1991. Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. Science (1979) 252:1651–1656.

14. Collin M, Bigley V. 2018. Human dendritic cell subsets: an update. Immunology 154:3–20.

15. Mazurais D, Darias M, Zambonino-Infante JL, Cahu CL. 2011. Transcriptomics for understanding marine fish larval development. Can J Zool 89:599–611.

16. Balmer N V., Leist M. 2014. Epigenetics and transcriptomics to detect adverse drug effects in model systems of human development. Basic Clin Pharmacol Toxicol 115:59–68.

17. Quan X jiang, Ramaekers A, Hassan BA. 2012. Transcriptional Control of Cell Fate Specification. Lessons from the Fly Retina. Curr Top Dev Biol 98:259–276.

18. Vengellur A, Woods BG, Ryan HE, Johnson RS, Lapres JJ. 2003. Gene expression profiling of the hypoxia signaling pathway in hypoxia-inducible factor 1alpha null mouse embryonic fibroblasts. Gene Expr 11:181–197.

19. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349.

20. Van't Veer LJ, Dai H, Van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, Van Der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS,

Bernards R, Friend SH. 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530–536.

21. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Hong ML, Dubois J, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, De La Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin E V., Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran J V., Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AFA, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson M V., Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Patrinos A, Morgan MJ. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

22. Craig Venter J, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, di Francesco V, Dunn P, Eilbeck K,

Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov G V., Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Yuan Wang Z, Wang A, Wang X, Wang J, Wei MH, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Lai Cheng M, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Ni Tint N, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander K V., Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Deslattes Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. 2001. The sequence of the human genome. Science (1979) 291:1304–1351.

23. Eaton L. 2002. Human genome pioneers are awarded Nobel prize. BMJ : British Medical Journal 325:791.

24. Doudna JA, Cech TR. 2002. The chemical repertoire of natural ribozymes. Nature 418:222–228.

25. Wilson DS, Szostak JW. 1999. In vitro selection of functional nucleic acids. Annu Rev Biochem 68:611–647.

26. Strahl BD, Allis CD. 2000. The language of covalent histone modifications. Nature 403:41–45.

27. He C, Klionsky DJ. 2009. Regulation mechanisms and signaling pathways of autophagy. Annu Rev Genet 43:67–93.

28. Mann M, Jensen ON. 2003. Proteomic analysis of post-translational modifications. Nat Biotechnol 21:255–261.

29. Stretton AOW. 2002. The First Sequence: Fred Sanger and Insulin. Genetics 162:527–532.

30. O'Farrell PZ, Goodman HM, O'Farrell PH. 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell 12:1133–1142.

31. O'Farrell PH. 1975. High Resolution Two-Dimensional Electrophoresis of Proteins. J Biol Chem 250:4007.

32. Edman P, Lindberg B, Waterman HI, Eliasson NA, Thorell B. 1956. On the Mechanism of the Phenyl Isothiocyanate Degradation of Peptides. Acta Chem Scand 10:761–768.

33. Laursen RA. 1971. Solid-Phase Edman Degradation: An Automatic Peptide Sequencer. Eur J Biochem 20:89–102.

34. Bellamy W, Takase M, Yamauchi K, Wakabayashi H, Kawase K, Tomita M. 1992. Identification of the bactericidal domain of lactoferrin. Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular 1121:130–136.

35. Mendez E, Lai CY. 1975. Regeneration of amino acids from thiazolinones formed in the Edman degradation. Anal Biochem 68:47–53.

36. Blackstock WP, Weir MP. 1999. Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol 17:121–127.

37. Li G, Waltham M, Anderson NL, Unsworth E, Treston A, Weinstein JN. 1997. Rapid mass spectrometric identification of proteins from two-dimensional polyacrylamide gels after in gel proteolytic digestion. Electrophoresis 18:391–402.

38. Bunkenborg J, Espadas G, Molina H. 2013. Cutting edge proteomics: Benchmarking of six commercial trypsins. J Proteome Res 12:3631–3641.

39. Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, Schönfisch B, Perschil I, Chacinska A, Guiard B, Rehling P, Pfanner N, Meisinger C. 2003. The proteome of Saccharomyces cerevisiae mitochondria. Proc Natl Acad Sci U S A 100:13207–13212.

40. Schmidt O, Pfanner N, Meisinger C. 2010. Mitochondrial protein import: from proteomics to functional mechanisms. Nature Reviews Molecular Cell Biology 2010 11:9 11:655–667.

41. Ferro M, Salvi D, Brugière S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N. 2003. Proteomics of the Chloroplast Envelope Membranes from Arabidopsis thaliana. Molecular & Cellular Proteomics 2:325–345.

42. Zakarias T, Bunkenborg J, Gronborg M, Molina H, Thuluvath PJ, Argani P, Goggins MG, Maitra A, Pandey A. 2004. A proteomic analysis of human bile. Mol Cell Proteomics 3:715–728.

43. Li Q, Li L, Rejtar T, Karger BL, Ferry JG. 2005. Proteome of Methanosarcina acetivorans part II: Comparison of protein levels in acetate- and methanol-grown cells. J Proteome Res 4:129–135.

44. Kao WC, Chen YR, Yi EC, Lee H, Tian Q, Wu KM, Tsai SF, Yu SSF, Chen YJ, Aebersold R, Chan SI. 2004. Quantitative Proteomic Analysis of Metabolic Regulation by Copper Ions in Methylococcus capsulatus (Bath). Journal of Biological Chemistry 279:51554–51560.

45. Wang W, Sun J, Nimtz M, Deckwer WD, Zeng AP. 2003. Protein identification from two-dimensional gel electrophoresis analysis of Klebsiella pneumoniae by combined use of mass spectrometry data and raw genome sequences. Proteome Sci 1:6.

46. Sowell SM, Norbeck AD, Lipton MS, Nicora CD, Callister SJ, Smith RD, Barofsky DF, Giovannoni SJ. 2008. Proteomic analysis of stationary phase in the marine bacterium "Candidatus pelagibacter ubique." Appl Environ Microbiol 74:4091–4100.

47. Lin JF, Xu J, Tian HY, Gao X, Chen QX, Gu Q, Xu GJ, Song J Da, Zhao FK. 2007. Identification of candidate prostate cancer biomarkers in prostate needle biopsy specimens using proteomic analysis. Int J Cancer 121:2596–2605.

48. Ahram M, Best CJM, Flaig MJ, Gillespie JW, Leiva IM, Chuaqui RF, Zhou G, Shu H, Duray PH, Linehan WM, Raffeld M, Ornstein DK, Zhao Y, Petricoin EF, Emmert-Buck MR. 2002. Proteomic analysis of human prostate cancer†. Mol Carcinog 33:9–15.

49. Shotgun proteomics: Integrating technologies to answer biological questions-Web of Science Core Collection. https://www.webofscience.com/wos/woscc/full-record/WOS:000184024600013. Retrieved 9 April 2023.

50. Yates JR, Ruse CI, Nakorchevsky A. 2009. Proteomics by mass spectrometry: Approaches, advances, and applications. Annu Rev Biomed Eng 11:49–79.

51. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson IM, Edlund K, Lundberg E, Navani S, Szigyarto CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, Von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, Von Heijne G, Nielsen J, Pontén F. 2015. Tissue-based map of the human proteome. Science (1979) 347.

52. Uhlén M, Björling E, Agaton C, Szigyarto CAK, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, Berglund L, Bergström K, Brumer H, Cerjan D, Ekström M, Elobeid A, Eriksson C, Fagerberg L, Falk R, Fall J, Forsberg M, Björklund MG, Gumbel K,

Halimi A, Hallin I, Hamsten C, Hansson M, Hedhammar M, Hercules G, Kampf C, Larsson K, Lindskog M, Lodewyckx W, Lund J, Lundeberg J, Magnusson K, Malm E, Nilsson P, Ödling J, Oksvold P, Olsson I, Öster E, Ottosson J, Paavilainen L, Persson A, Rimini R, Rockberg J, Runeson M, Sivertsson Å, Sköllermo A, Steen J, Stenvall M, Sterky F, Strömberg S, Sundberg M, Tegel H, Tourle S, Wahlund E, Waldén A, Wan J, Wernérus H, Westberg J, Wester K, Wrethagen U, Xu LL, Hober S, Pontén F. 2005. A human protein atlas for normal and cancer tissues based on antibody proteomics. Molecular and Cellular Proteomics 4:1920–1932.

53. Uhlén M, Karlsson MJ, Hober A, Svensson AS, Scheffel J, Kotol D, Zhong W, Tebani A, Strandberg L, Edfors F, Sjöstedt E, Mulder J, Mardinoglu A, Berling A, Ekblad S, Dannemeyer M, Kanje S, Rockberg J, Lundqvist M, Malm M, Volk AL, Nilsson P, Månberg A, Dodig-Crnkovic T, Pin E, Zwahlen M, Oksvold P, von Feilitzen K, Häussler RS, Hong MG, Lindskog C, Ponten F, Katona B, Vuu J, Lindström E, Nielsen J, Robinson J, Ayoglu B, Mahdessian D, Sullivan D, Thul P, Danielsson F, Stadler C, Lundberg E, Bergström G, Gummesson A, Voldborg BG, Tegel H, Hober S, Forsström B, Schwenk JM, Fagerberg L, Sivertsson Å. 2019. The human secretome. Sci Signal 12.

54. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, Sjöstedt E, Butler L, Odeberg J, Dusart P, Edfors F, Oksvold P, von Feilitzen K, Zwahlen M, Arif M, Altay O, Li X, Ozcan M, Mardonoglu A, Fagerberg L, Mulder J, Luo Y, Ponten F, Uhlén M, Lindskog C. 2021. A single–cell type transcriptomics map of human tissues. Sci Adv 7.

55. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, Wernerus H, Björling L, Ponten F. 2010. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol 28:1248–1250.

56. Uhlen M, Karlsson MJ, Zhong W, Tebani A, Pou C, Mikes J, Lakshmikanth T, Forsström B, Edfors F, Odeberg J, Mardinoglu A, Zhang C, von Feilitzen K, Mulder J, Sjöstedt E, Hober A, Oksvold P, Zwahlen M, Ponten F, Lindskog C, Sivertsson Å, Fagerberg L, Brodin P. 2019. A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. Science (1979) 366.

57. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, Sanli K, Von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnström H, Glimelius B, Sjöblom T, Edqvist PH, Djureinovic D, Micke P, Lindskog C, Mardinoglu A, Ponten F. 2017. A pathology atlas of the human cancer transcriptome. Science (1979) 357.

58. Pontén F, Jirström K, Uhlen M. 2008. The Human Protein Atlas - A tool for pathology. Journal of Pathology 216:387–393.

59. Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szigyarto CAK, Persson A, Ottosson J, Wernérus H, Nilsson P, Lundberg E, Sivertsson Å, Navani S, Wester K, Kampf C, Hobert S, Pontén F, Uhlén M. 2008. A genecentric human protein atlas for expression profiles based on antibodies. Molecular and Cellular Proteomics 7:2019–2027.

60. The Human Protein Atlas. https://www.proteinatlas.org/. Retrieved 9 April 2023.

61. Thul PJ, Akesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, Fall J, Gatto L, Gnann C, Hober S, Hjelmare M, Johansson F, Lee S, Lindskog C, Mulder J, Mulvey CM, Nilsson P, Oksvold P, Rockberg J, Schutten R, Schwenk JM, Sivertsson A, Sjöstedt E, Skogs M, Stadler C, Sullivan DP, Tegel H, Winsnes C, Zhang C, Zwahlen M, Mardinoglu A, Pontén F, Von Feilitzen K, Lilley KS, Uhlén M, Lundberg E. 2017. A subcellular map of the human proteome. Science (1979) 356.

62. Jiang Y, Sun A, Zhao Y, Ying W, Sun H, Yang X, Xing B, Sun W, Ren L, Hu B, Li C, Zhang L, Qin G, Zhang M, Chen N, Zhang M, Huang Y, Zhou J, Zhao Y, Liu M, Zhu X, Qiu Y, Sun Y, Huang C, Yan M, Wang M, Liu W, Tian F, Xu H, Zhou J, Wu Z, Shi T, Zhu W, Qin J, Xie L, Fan J, Qian X, He F, He F, Qian X, Qin J, Jiang Y, Ying W, Zhu Y, Zhu W, Wang Y, Yang D, Liu W, Liu Q, Yang X, Zhen B, Wu Z, Qian J, Hong T, Shen L, Xing B, Yang P, Shen H, Zhang L, Cheng S, Cai J, Zhao X, Sun Y, Xiao T, Mao Y, Chen X, Wu D, Chen L, Dong J, Deng H, Tan M, Wu Z, Zhao Q, Shen Z, Chen X, Gao Y, Sun W, Wang T, Liu S, Lin L, Zi J, Lou X, Zeng R, Wu Y, Cai S, Jiang B, Chen A, Li Z, Yang F, Chen X, Sun Y, Wang Q, Zhang Y, Wang G, Chen Z, Qin W, Li Z. 2019. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. Nature 567:257–261.

63. Johnson ECB, Dammer EB, Duong DM, Ping L, Zhou M, Yin L, Higginbotham LA, Guajardo A, White B, Troncoso JC, Thambisetty M, Montine TJ, Lee EB, Trojanowski JQ, Beach TG, Reiman EM, Haroutunian V, Wang M, Schadt E, Zhang B, Dickson DW, Ertekin-Taner N, Golde TE, Petyuk VA, De Jager PL, Bennett DA, Wingo TS, Rangaraju S, Hajjar I, Shulman JM, Lah JJ, Levey AI, Seyfried NT. 2020. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. Nat Med 26:769–780.

64. Messner CB, Demichev V, Wendisch D, Michalick L, White M, Freiwald A, Textoris-Taube K, Vernardis SI, Egger AS, Kreidl M, Ludwig D, Kilian C, Agostini F, Zelezniak A, Thibeault C, Pfeiffer M, Hippenstiel S, Hocke A, von Kalle C, Campbell A, Hayward C, Porteous DJ, Marioni RE, Langenberg C, Lilley KS, Kuebler WM, Mülleder M, Drosten C, Suttorp N, Witzenrath M, Kurth F, Sander LE, Ralser M. 2020. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. Cell Syst 11:11-24.e4.

65. Praissman JL, Wells L. 2021. Proteomics-Based Insights Into the SARS-CoV-2–Mediated COVID-19 Pandemic: A Review of the First Year of Research. Mol Cell Proteomics 20.

66. Kim DK, Weller B, Lin CW, Sheykhkarimli D, Knapp JJ, Dugied G, Zanzoni A, Pons C, Tofaute MJ, Maseko SB, Spirohn K, Laval F, Lambourne L, Kishore N, Rayhan A, Sauer M, Young V, Halder H, la Rosa NM de, Pogoutse O, Strobel A, Schwehn P, Li R, Rothballer ST, Altmann M, Cassonnet P, Coté AG, Vergara LE, Hazelwood I, Liu BB, Nguyen M, Pandiarajan R, Dohai B, Coloma PAR, Poirson J, Giuliana P, Willems L, Taipale M, Jacob Y, Hao T, Hill

DE, Brun C, Twizere JC, Krappmann D, Heinig M, Falter C, Aloy P, Demeret C, Vidal M, Calderwood MA, Roth FP, Falter-Braun P. 2022. A proteome-scale map of the SARS-CoV-2–human contactome. Nature Biotechnology 2022 41:1 41:140–149.

67. Brogna C, Cristoni S, Brogna B, Bisaccia DR, Marino G, Viduto V, Montano L, Piscopo M. 2023. Toxin-like Peptides from the Bacterial Cultures Derived from Gut Microbiome Infected by SARS-CoV-2—New Data for a Possible Role in the Long COVID Pattern. Biomedicines 11.

68. Bauer A, Pachl E, Hellmuth JC, Kneidinger N, Heydarian M, Frankenberger M, Stubbe HC, Ryffel B, Petrera A, Hauck SM, Behr J, Kaiser R, Scherer C, Deng L, Teupser D, Ahmidi N, Muenchhoff M, Schubert B, Hilgendorff A. 2023. Proteomics reveals antiviral host response and NETosis during acute COVID-19 in high-risk patients. Biochim Biophys Acta Mol Basis Dis 1869:166592.

69. Complement activation induces excessive T cell cytotoxicity in severe COVID-19-Web of Science Core Collection. https://www.webofscience.com/wos/woscc/full-record/WOS:000768266500001. Retrieved 2 April 2023.

70. Georg P, Astaburuaga-García R, Bonaguro L, Brumhard S, Michalick L, Lippert LJ, Kostevc T, Gäbel C, Schneider M, Streitz M, Demichev V, Gemünd I, Barone M, Tober-Lau P, Helbig ET, Hillus D, Petrov L, Stein J, Dey HP, Paclik D, Iwert C, Mülleder M, Aulakh SK, Djudjaj S, Bülow RD, Mei HE, Schulz AR, Thiel A, Hippenstiel S, Saliba AE, Eils R, Lehmann I, Mall MA, Stricker S, Röhmel J, Corman VM, Beule D, Wyler E, Landthaler M, Obermayer B, von Stillfried S, Boor P, Demir M, Wesselmann H, Suttorp N, Uhrig A, Müller-Redetzky H, Nattermann J, Kuebler WM, Meisel C, Ralser M, Schultze JL, Aschenbrenner AC, Thibeault C, Kurth F, Sander LE, Blüthgen N, Sawitzki B. 2022. Complement activation induces excessive T cell cytotoxicity in severe COVID-19. Cell 185:493-512.e25.

71. Bi X, Liu W, Ding X, Liang S, Zheng Y, Zhu X, Quan S, Yi X, Xiang N, Du J, Lyu H, Yu D, Zhang C, Xu L, Ge W, Zhan X, He J, Xiong Z, Zhang S, Li Y, Xu P, Zhu G, Wang D, Zhu H, Chen S, Li J, Zhao H, Zhu Y, Liu H, Xu J, Shen B, Guo T. 2022. Proteomic and metabolomic profiling of urine uncovers immune responses in patients with COVID-19. Cell Rep 38:110271.

72. Patti GJ, Yanes O, Siuzdak G. 2012. Metabolomics: the apogee of the omics trilogy. Nature Reviews Molecular Cell Biology 2012 13:4 13:263–269.

73. Hopkins FG. 1938. Archibald Edward Garrod. 1857-1936. Obituary Notices of Fellows of the Royal Society 2:224–228.

74. Prasad C, Galbraith PA. 2005. Sir Archibald Garrod and Alkaptonuria –'story of metabolic genetics'. Clin Genet 68:199–203.

75. Garrod AE. 1902. The Incidence of Alkaptonuria: A Study in Chemical Individuality. Lancet ii:1616–1620.

76. Dunn WB, Ellis DI. 2005. Metabolomics: Current analytical platforms and methodologies. TrAC Trends in Analytical Chemistry 24:285–294.

77. Hoult DI, Busby SJW, Gadian DG, Radda GK, Richards RE, Seeley PJ. 1974. Observation of tissue metabolites using 31P nuclear magnetic resonance. Nature 1974 252:5481 252:285–287.

78. Fiehn O, Kopka J, Dörmann P, … TA-N, 2000 undefined. Metabolite profiling for plant functional genomics. nature.com.

79. Scripps Research. https://www.scripps.edu/. Retrieved 9 April 2023.

80. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. 2005. METLIN: A metabolite mass spectral database. Ther Drug Monit 27:747–751.

81. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, Palermo A, Warth B, Hermann G, Koellensperger G, Huan T, Uritboonthai W, Aisporna AE, Wolan DW, Spilker ME, Benton HP, Siuzdak G. 2018. METLIN: A Technology Platform for Identifying Knowns and Unknowns. Anal Chem 90:3156–3164.

82. Prostaglandin OO, Fa0301] B Glycerolipids [. 2005. The emerging field of lipidomics. Nature Reviews Drug Discovery 2005 4:7 4:594–610.

83. Yashroy RC. 1990. Magnetic resonance studies of dynamic organisation of lipids in chloroplast membranes. J Biosci 15:281–288.

84. Kagan VE, Tyurin VA, Jiang J, Tyurina YY, Ritov VB, Amoscato AA, Osipov AN, Belikova NA, Kapralov AA, Kini V, Vlasova II, Zhao Q, Zou M, Di P, Svistunenko DA, Kurnikov I V., Borisenko GG. 2005. Cytochrome c acts as a cardiolipin oxygenase required for release of proapoptotic factors. Nature Chemical Biology 2005 1:4 1:223–232.

85. Borradaile NM, Han X, Harp JD, Gale SE, Ory DS, Schaffer JE. 2006. Disruption of endoplasmic reticulum structure and integrity in lipotoxic cell death. J Lipid Res 47:2726–2737.

86. Chan RB, Oliveira TG, Cortes EP, Honig LS, Duff KE, Small SA, Wenk MR, Shui G, Di Paolo G. 2012. Comparative lipidomic analysis of mouse and human brain with Alzheimer disease. Journal of Biological Chemistry 287:2678–2688.

87. Dettmer K, Aronov PA, Hammock BD. 2007. Mass spectrometry-based metabolomics. Mass Spectrom Rev 26:51–78.

88. Kanu AB, Dwivedi P, Tam M, Matz L, Hill HH. 2008. Ion mobility–mass spectrometry. Journal of Mass Spectrometry 43:1–22.

89. Sharon G, Cruz NJ, Kang DW, Gandal MJ, Wang B, Kim YM, Zink EM, Casey CP, Taylor BC, Lane CJ, Bramer LM, Isern NG, Hoyt DW, Noecker C, Sweredoski MJ, Moradian A, Borenstein E, Jansson JK, Knight R, Metz TO, Lois C, Geschwind DH, Krajmalnik-Brown R, Mazmanian SK. 2019. Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. Cell 177:1600-1618.e17.

90. MahmoudianDehkordi S, Arnold M, Nho K, Ahmad S, Jia W, Xie G, Louie G, Kueider-Paisley A, Moseley MA, Thompson JW, St John Williams L, Tenenbaum JD, Blach C, Baillie R, Han X, Bhattacharyya S, Toledo JB, Schafferer S, Klein S, Koal T, Risacher SL, Kling MA, Motsinger-Reif A, Rotroff DM, Jack J, Hankemeier T, Bennett DA, De Jager PL, Trojanowski JQ, Shaw LM, Weiner MW, Doraiswamy PM, van Duijn CM, Saykin AJ, Kastenmüller G, Kaddurah-Daouk R. 2019. Altered bile acid profile associates with cognitive impairment in Alzheimer's disease—An emerging role for gut microbiome. Alzheimer's & Dementia 15:76–92.

91. Qiao R, Sheng C, Lu Y, Zhang Y, Ren H, Lemos B. 2019. Microplastics induce intestinal inflammation, oxidative stress, and disorders of metabolome and microbiome in zebrafish. Science of The Total Environment 662:246–253.

92. Wan Y, Wang F, Yuan J, Li J, Jiang D, Zhang J, Li H, Wang R, Tang J, Huang T, Zheng J, Sinclair AJ, Mann J, Li D. 2019. Effects of dietary fat on gut microbiota and faecal metabolites, and their relationship with cardiometabolic risk factors: a 6-month randomised controlled-feeding trial. Gut 68:1417–1429.

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

94. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, Sayeeda Z, Lo E, Assempour N, Berjanskii M, Singhal S, Arndt D, Liang Y, Badran H, Grant J, Serra-Cayuela A, Liu Y, Mandal R, Neveu V, Pon A, Knox C, Wilson M, Manach C, Scalbert A. 2018. HMDB 4.0: The human metabolome database for 2018. Nucleic Acids Res 46:D608–D617.

95. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A. 2013. HMDB 3.0-The Human Metabolome Database in 2013. Nucleic Acids Res 41.

96. Xia J, Sinelnikov I V., Han B, Wishart DS. 2015. MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Res 43:W251–W257.

97. Wang M, Carver JJ, Phelan V V., Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik A V., Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot

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

98. Rohart F, Gautier B, Singh A, Lê Cao KA. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 13.

99. Wassenaar TA, Ingólfsson HI, Böckmann RA, Tieleman DP, Marrink SJ. 2015. Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. J Chem Theory Comput 11:2144–2155.

100. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CYJ, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A, Bacic A, Hill AF, Stroud DA, Ryan MT, Agbinya JI, Mariadason JM, Burgess AW, Mathivanan S. 2015. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 15:2597–2601.

101. Wassenaar TA, Ingólfsson HI, Böckmann RA, Tieleman DP, Marrink SJ. 2015. Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. J Chem Theory Comput 11:2144–2155.

102. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. 2017. MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res 45:W180–W188.

103. Pang Z, Chong J, Zhou G, De Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques PÉ, Li S, Xia J. 2021. MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49:W388–W396.

104. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, Vandergheynst J, Fiehn O, Arita M. 2015. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods 12:523–526.

105. m/q Initiative | PNNL. https://www.pnnl.gov/projects/mq-initiative. Retrieved 3 April 2023.

106. Domon B, Aebersold R. 2006. Mass spectrometry and protein analysis. Science (1979) 312:212–217.

107. Pulfer M, Murphy RC. 2003. Electrospray mass spectrometry of phospholipids. Mass Spectrom Rev 22:332–364.

108. Mass spectrometry and the age of the proteome-Web of Science Core Collection. https://www.webofscience.com/wos/woscc/full-record/WOS:000071585000001. Retrieved 9 April 2023.

109. Kanu AB, Dwivedi P, Tam M, Matz L, Hill HH. 2008. Ion mobility-mass spectrometry. Journal of Mass Spectrometry 43:1–22.

110. Aebersold R, Goodlett DR. 2001. Mass spectrometry in proteomics. Chem Rev 101:269–295.

111. Taylor PJ. 2005. Matrix effects: The Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. Clin Biochem 38:328–334.

112. Studying noncovalent protein complexes by electrospray ionization mass spectrometry-Web of Science Core Collection. https://www.webofscience.com/wos/woscc/full-record/WOS:A1997XM88300001. Retrieved 9 April 2023.

113. Smith RD, Loo JA, Edmonds CG, Barinaga CJ, Udseth HR. 1990. New Developments in Biochemical Mass Spectrometry: Electrospray Ionization. Anal Chem 62:882–899.

114. Dettmer K, Aronov PA, Hammock BD. 2007. Mass spectrometry-based metabolomics. Mass Spectrom Rev 26:51–78.

115. Marshall AG, Hendrickson CL, Jackson GS. FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS SPECTROMETRY: A PRIMER https://doi.org/10.1002/(SICI)1098-2787(1998)17:1.

116. Edition S. 2006. The Expanding Role of Mass Spectrometry in Biotechnology, Second Edition [Paperback] 257.

117. SYNAPT XS : Waters. (n.d.). Retrieved May 1, 2023, from https://ims.waters.com/the-synapt-xs/ .

118. Wiley WC, McLaren IH. 2004. Time-of-Flight Mass Spectrometer with Improved Resolution. Review of Scientific Instruments 26:1150.

119. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG. 2005. The Orbitrap: A new mass spectrometer. Journal of Mass Spectrometry 40:430–443.

A SHIFT BETWEEN MINERAL AND NONMINERAL SOURCES OF IRON AND SULFUR CAUSES PROTEOME-WIDE CHANGES IN METHANOSARCINA BARKERI

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: Hunter Fausset

Contributions: Main author. Experimental design, sample prep, data acquisition, statistical analysis, interpretation, writing, submission, editing.

Co-Author: Rachel L Spietz

Contributions: Cell culturing, intellectual assistance, writing, editing.

Co-Author: Savannah Cox

Contributions: Bioinformatic data processing writing, editing

Co-Author: Gwendolyn Cooper

Contributions: Bioinformatic data processing writing, editing

Co-Author: Scott Spurzem

Contributions: Sample preparation, bioinformatic data processing writing,

Co-Author: Monika Tokmina-Lukaszewska

Contributions: Data interpretation. writing, editing

Co-Author: Jennifer Dubois

Contributions: Data interpretation, writing, editing

Co-Author: Joan B. Broderick

Contributions: Data interpretation, writing, editing

Co-Author: Eric M. Shepard

Contributions: Data interpretation, writing, editing

Co-Author: Eric S. Boyd

Contributions: Grant writing, cell culturing, data interpretation, writing, editing

Co-Author: Brian Bothner

Contributions: Corresponding author, grant writing, data interpretation, writing, editing

Manuscript Information

Hunter Fausset, Rachel L. Spietz, Savannah Cox, Gwendolyn Cooper, Scott Spurzem, Monika Tokmina-Lukaszewska, Jennifer Dubois, Joan B. Broderick, Eric M. Shepard, Eric S. Boyd, Brian Bothner

Microbiology Spectrum

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
__x__ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

American Society for Microbiology
Submitted January 2023. Minor revisions underway, resubmission 04/23

Abstract:

Iron (Fe) and sulfur (S) are required elements for life, and changes in their availability can limit the ecological distribution and function of microorganisms. In anoxic environments, soluble Fe typically exists as ferrous iron (Fe(II)) and S as sulfide (HS-). These species exhibit a strong affinity towards one another, resulting in the formation of sedimentary pyrite (FeS2). Recent, paradigm shifting studies indicate that Fe and S in pyrite can be made bioavailable by methanogens through a reductive dissolution process. However, the impact of the utilization of FeS2, as opposed to canonical Fe and S sources, on the phenotype of cells is not fully understood. Here, shotgun proteomics was utilized to measure changes in the phenotype of Methanosarcina barkeri grown with FeS2, Fe(II)/HS-; or Fe(II)/cysteine. Shotgun proteomics tracked 1019 proteins, with the abundance of over 30% (307) changing significantly between growth conditions. Functional characterization and pathway analyses revealed these changes to be systemic, and largely tangential to Fe/S metabolism. As a final step, the proteomics data was viewed with respect to previously collected transcriptomics data to deepen the analysis. Presented here is evidence that M. barkeri adopts distinct phenotypes to exploit specific sources of Fe and S in their environment. This is supported by observed protein abundance changes across broad categories of cellular biology. DNA adjacent metabolism, central carbon metabolism / methanogenesis, metal trafficking, quorum sensing and porphyrin biosynthesis pathways are all features in the phenotypic differentiation.

Introduction:

With just one characterized exception, Lactococcus plantarum (1), all organisms require iron (Fe) and sulfur (S) as essential components of vitamins and a wide range of co-enzymes and co-factors. [Fe-S] clusters, when coordinated by a protein or protein complex, have a variety of roles in biological systems including electron transfer, substrate binding, and catalysis (2, 3). Therefore, it is not surprising that [Fe-S] clusters are widely distributed across the proteomes in all three domains of life, in processes as varied as photosynthesis, respiration, and fermentation (4–6).

Microorganisms typically acquire iron by direct transmembrane transport of ferrous iron (Fe(II)) or with the aid of extracellular metal chelators known as siderophores (7, 8). The mechanism of Fe uptake largely depends on the form of iron available to an organism. Ferric (Fe(III)) iron is the most abundant form in oxic conditions and Fe(II) in anoxic environments (8). Fe(III) has low solubility in oxic environments; however, secreted siderophores can chelate Fe(III) making it more bioavailable. Like Fe, the availability and form of S is dependent on environmental conditions. S can be acquired as inorganic (e.g., sulfate, sulfide (HS-)) or organic (e.g., cysteine (Fe(II)/Cys)) forms (9). In oxic environments, inorganic S largely exists as sulfate, whereas in anoxic environments HS- tends to predominate. Regardless of the form acquired, reduced forms of S are preferred as this is the oxidation state most often found in co-factors and amino acids. Intricate pathways have evolved to assimilate sulfate as HS- and to liberate S from organic sources for use in [Fe-S] clusters, amino acids, and other biomolecules (9).

Methanogens are anaerobic archaea that generate methane as a product of the metabolism of hydrogen and carbon dioxide, formate, methanol, methylamines, and/or acetate. Due to their

unique biology, these organisms require more Fe per cell and encode a higher number of [Fe-S] cluster binding motifs in encoded proteins when compared to other organisms (10–12). Methanogens are typically grown with Fe(II) as the Fe source and HS- or Fe(II)/Cys as a S source. However, Fe(II) and HS- readily react to form soluble [Fe-S] clusters that nucleate and ultimately form low solubility iron-sulfide minerals, such as pyrite (FeS2). As such, higher concentrations of environmental Fe(II) lead to S limitation and higher concentrations of environmental HS- can lead to Fe limitation. How methanogens obtain Fe and S under anoxic conditions that promotes FeS2 formation is a central question in understanding the ecology and function of these organisms (13).

Recently, the methanogens Methanococcus voltae, Methanococcus maripaludis, and Methanosarcina barkeri were shown to be capable of using FeS2 as their sole source of Fe and S through a reductive dissolution process (14). Reduction of FeS2 releases HS- into solution and generates pyrrhotite (Fe1-xS) on the mineral surface. Subsequent solubilization of Fe from Fe1-xS yields soluble Fe(II) that can react with HS- in solution to from aqueous [Fe-S] clusters (FeSaq), the presumed source of Fe and S for cells (14). Direct contact between FeS2 and cells is required for reduction, implying direct extracellular electron transfer (EET). This mechanism, as well as the process by which cells internalize and process FeSaq clusters, remains unelucidated.

The various forms of Fe and S that can be used to cultivate methanogens provides an opportunity to address such gaps in knowledge. For example, comparison of cells grown with excess HS- and limiting Fe(II), compared to growth with FeS2, could provide insight into how cells overcome Fe(II) limitation during growth. Likewise, comparison of cells grown with Fe(II)/Cys and Fe(II), when compared to growth with FeS2 could shed light into how Fe(II)/Cys

decomposition products (e.g., alanine) broadly impact carbon utilization pathways (15, 16). Here, we examine M. barkeri strain MS cells grown with HS-/Fe(II), Fe(II)/Cys/Fe(II), and FeS2 as their sole Fe and S sources. A total proteome analysis under the different conditions was completed to identify specific proteins and biological pathways they are important for growth on different sources of Fe and S. This data was then combined with previously published transcriptomics data (17) to identify changes on a multi-omics level.

Results:

Shotgun Proteomics and Statistical Analysis

To elucidate proteins involved in Fe assimilation, batch cultures of M. barkeri were grown on three different forms of Fe and S; Fe(II)/HS- (as FeCl2 and Na2S), FeS2 (as synthetic, framboidal pyrite), and Fe(II)/Cys (as FeCl2 and L-cysteine). Cultures (n = 4 per condition) were harvested mid-log phase, based on methane production and DNA yield as proxies for cell growth (17). The soluble intracellular protein fraction was digested with trypsin followed by label-free Liquid Chromatography Mass Spectrometry (LCMS). Peptide abundance data was processed with MASCOT (18) and Scaffold (Proteome Software, Inc., Portland, OR), before export for statistical analysis using Metaboanalyst 5.0 (19, 20). A total of 1019 proteins were used to compare samples. Multivariate analysis based on protein abundance was employed to distinguish proteins with a statistically significant difference between conditions. A total of 307 proteins were statistically significant in a one-way ANOVA (p-value < 0.05), representing 30.1% of the dataset (Figure 1A). Abundance changes ranged from a lower threshold of 2-fold to upwards of 180-fold. Principal component analysis (PCA) was performed to evaluate intra- and intergroup proteomic variability. Each growth condition clearly separated when ordinated by PCA, with

roughly equilateral distance between groups in the first two components, which together account for 45.4% of the total variability of the dataset (Figure 1B). The 95% confidence intervals of each group showed no overlap, indicating that the proteomic changes observed between each group were distinct. Hierarchical clustering analysis (HCA) including all proteins resulted in perfect group-wise clustering, with Fe(II)/Cys samples grouping away from Fe(II)/HS- and FeS2 (Figure 1C). This clustering pattern is indicative of a large shift in the overall proteome between groups. The parallel HCA (PHCA) shows the treatment group clustering orthogonal to the pattern of protein abundance. clustered by their regulatory patterns (i.e abundance in one condition relative to the others). Proteins that have consistent abundance changes within a group are clustered together in co-regulated blocks (Figure 1C).

To help focus the analysis, HCA was repeated using the 35 most significantly dysregulated proteins, filtered by p-value) (Figure 2). Five distinct regulatory patterns (clusters) are present in the filtered list. Cluster 1 includes an uncharacterized protein alongside a glutaredoxin family member, both of which are most abundant in the FeS2 condition followed by Fe(II)/HS- and are lowest in Fe(II)/Cys cells. The second cluster is the largest and contains di- and oligopeptide binding ABC transporters, as well as cell surface proteins. Two Iron (III) Dicitrate Binding Proteins that are adjacent in the genome have essentially identical regulation patterns in this cluster. Proteins in cluster 2 are distinctly downregulated in the Fe(II)/Cys condition. Cluster 3 contains a copper binding protein, as well as central carbon and methanogenesis enzymes (V-type ATP synthase and methyltransferase corrinoid activation protein). Clusters 4 and 5 contain secondary metabolism enzymes (L-threonine-O-3-phosphate decarboxylase, iron sulfur cluster assembly protein SufB) and central carbon enzymes (two

acetyl-CoA-decarbonylase subunits). Box and whisker plots displaying the regulatory pattern for these 35 proteins are in Figure S1.

Functional distribution analysis

To frame the statistical distance between the groups in terms of biological processes, partial least squares discriminant analysis (PLS-DA), a dimensional reduction statistical model, was performed. The top 15 proteins, filtered by VIP score (a metric of statistical weight for differentiating groups) mirror the heatmap; there is no clear pattern in their enzymatic activity, cellular location, or biological function (Figure S2). Predicted functions range from protein biosynthesis to transmembrane transport and are not expected to exhibit obvious interactions with either S or Fe, nor do they appear to be involved in their metabolism. To determine if the functional distribution seen in the dysregulated proteins spans the entirety of the dataset, the functions of all identified proteins were mapped and compared to those with statistical relevance. The Database for Annotation, Visualization and Integrated Discovery (DAVID) was employed to annotate and assign molecular function Gene Ontology (GO) numbers to each protein (21). Proteins were placed into one or more of the 24 functional categories. The proteome-wide distribution (all identified proteins) of molecular functions shows approximately 24% have transferase activity, 13% are hydrolases, and 9% are oxidoreductases. When filtered to only the significantly dysregulated proteins (p-value <0.05) from the ANOVA, the distribution remains markedly unchanged (Figure 3A). The similarity of the pie charts for total and dysregulated proteins implies that the regulated biological processes span a broad functional range, rather than specific pathways to accommodate growth on different Fe and S sources. In a previous study, Spietz et al. performed a transcriptomics analysis of M. barkeri MS under identical experimental

conditions (17). Using the transcriptomic data generated in that study, all 3000 transcripts were run through the DAVID annotation pipeline to assess their functional distribution. We then selected the 1000 most significantly regulated mRNAs and compared the functional distribution to the proteomics data reported here (Figure S3). The categories and ratios in the top 1000 transcripts match those in the top 300 proteins +/- 5% in each category (Figure S3). Thus, the two omics techniques agree that the scope of the biological changes occurring in M. barkeri when growing on FeS2, as compared to soluble Fe and S source, are nonspecific to Fe and S metabolism. Multi-omic investigations increase the strength of findings through intra-systemic corroboration, and this comparison represents one of the first to be done on a methanogenic archaeon.

The apparent systemic scope of dysregulation was further examined by categorizing the significantly dysregulated proteins into their respective biological process. This analysis combines the enzymatic functions from the previous section into the biological process they contribute to (e.g. transcription or translation), and also includes proteins with no enzymatic function (structural proteins, cell surface proteins, etc.). Five categories selected were DNA replication and transcription, energy metabolism, cellular transport, and structure. Uncharacterized proteins were grouped into a sixth category, representing approximately 15% of all significantly dysregulated features (Figure 3B). Mirroring the trend observed in the functional distribution analysis, adjustment of the M. barkeri Fe and S source leads to biological changes in proteins that span all biological processes. This, combined with the functional distribution and multivariate analysis, suggests that cells made large scale changes to their physiology based the source of Fe and S. Multivariate statistics show proteomic signatures, indicating that each growth

condition may represent a distinct phenotype, despite FeS2 and Fe(II)/HS- potentially sharing an effective Fe/S source.

Pairwise Analysis of Conditions

To characterize the biological variance between the phenotypes, pairwise comparative analyses were performed in order to link protein patterns with with specific environmental changes. The statistical analyses were split into three separate two-factor analyses (FeS2 vs Fe(II)/HS-, FeS2 vs Fe(II)/Cys, Fe(II)/Cys vs Fe(II)/HS-), wherein T-tests were performed on each protein to assess the significance of the difference in mean protein abundances for each growth conditional comparison. When assessed in this fashion, 390 of the 1037 proteins met the threshold of statistical significance ($p\text{-value} < 0.05$, $FC > 2$) in one or more of the comparisons. A Venn diagram of the three pairwise analyses was created to assess which proteins were unique to specific comparisons and which were shared between two or found in each comparison (Figure S4). Only 34 of the 390 proteins were statistically significant and shared in all three comparisons.

Proteins with significantly different abundances between treatment groups were entered into KEGG to map the biochemical pathways associated with the cellular changes. The top ten pathways and the number of proteins that were significant in each analysis are shown in Table 1. Notably, Fe(II)/Cys vs Fe(II)/HS- showed a greater degree of biochemical dysregulation than comparisons with FeS2 in nearly every category. Notable differences were present in proteins annotated as ribosomal, methane/carbon metabolism, quorum sensing, and amino acid related metabolism (all proteins annotated in the methane metabolism pathway also appear in carbon metabolism for M. barkeri, so the two categories were combined). The broad scope of proteomic

differences observed indicates the involved biological processes are nuanced and go beyond the aforementioned biological pathways. Because of this, only the most significant and well-annotated pathways will be used in this investigation to begin to characterize the phenotypes adopted in each growth condition.

Ribosomal tuning

Ribosomal protein regulation appears to play a large part in the phenotypic differences seen in changing Fe and S sources, with over 30 ribosomal proteins showing significant dysregulation. The FeS2 grown cells are in between the regulatory patterns of Fe(II)/HS- and Fe(II)/Cys conditions, with ribosomal protein abundance of Fe(II)/Cys > FeS2 > Fe(II)/HS-. Table 2 shows all significantly dysregulated ribosomal proteins from each two-group comparison, their fold change, and accession information. Interestingly, there was only one ribosomal protein significant in all three comparisons, 50S ribosomal protein L1. The other significant ribosomal proteins were not shared between the comparisons, indicating that the regulatory pattern is likely due to specific regulation events rather than a general response of up or down regulation (Figure 4). 50S ribosomal protein L1 is involved in translational regulation in addition to protein synthesis, providing additional evidence that translational regulation is crucial to the adjustment for growth on the different substrates. A difference in regulation between the two ribosomal subunits is apparent when comparing FeS2 to the other two growth conditions. All 50S ribosomal proteins followed the aforementioned hierarchical abundance pattern (Fe(II)/Cys > FeS2 > Fe(II)/HS); however, several 30S ribosomal proteins did not. No 30s subunit ribosomal proteins changed when comparing FeS2 to Fe(II)/HS-, but five 50S proteins did. The overall difference in ribosome biology was even more evident when comparing

Fe(II)/Cys to Fe(II)/HS-, there was large scale regulation, however the small and large subunits had similar patterns. 17 of 22 large subunit and 8 of 20 small subunit ribosomal proteins were more abundant in cells grown on Fe(II)/Cys, with none being downregulated.

Along with changes in protein concentration, the mRNA transcripts that encode ribosomal proteins also fluctuated between growth conditions. Ribosomal proteins are generally organized in operons and display similar regulatory patterns. Of particular interest are 13 ribosomal proteins that are adjacent in the genome (MSBRM_2579□MSBRM_2612). These 13 ribosomal proteins shared similar mRNA expression profiles, but their proteomic abundance differences were marked. The 50s ribosomal protein L23Ae (MSBRM_2610) stands out in particular, with its protein abundance 32-fold higher in Fe(II)/Cys than Fe(II)/HS-, but the transcripts encoding the protein were only upregulated 1.4-fold (Figure 4). Another large subunit ribosomal protein, L13Ae (MSBRM_477), had a similar regulatory pattern (31.5-fold more abundant in Fe(II)/Cys than Fe(II)/HS-); however, the transcripts decreased in abundance. Two other ribosomal proteins adjacent to L13Ae in the genome shared this pattern of regulation: L18e (MSBRM_0476) and S16e (MSBRM_0478). L23Ae binds to the 23S rRNA and is one of the proteins that rims the polypeptide exit tunnel on the exterior of the ribosome. L13Ae is a structural protein found towards the interior of the ribosome. It is interesting that these structural ribosomal proteins have such a strikingly higher abundance in cells grown with Fe(II)/Cys, especially considering the patterns seen in their respective transcriptional profiles. This could indicate they are involved in more than just ribosome formation and structure and may have moonlighting functions, as is observed in higher organisms (22, 23).

Methanogensis and Carbon metabolism

In M. barkeri, methanogenesis is central to cellular catabolism. Central carbon and methane metabolism showed complete overlap in KEGG when comparing the dysregulated proteins of interest, therefore, these two categories were combined for the remainder of the analysis. There were large absolute differences in enzymes involved in carbon metabolism between groups, with 24 found to be significant between Fe(II)/Cys and Fe(II)/HS-, while only seven differed between FeS2 and Fe(II)/HS-. The phenotype associated with growth on Fe(II)/Cys was particularly unique in its central carbon metabolism. When comparing central carbon enzymes between Fe(II)/Cys and Fe(II)/HS-, 15 were less abundant and nine were more abundant in Fe(II)/Cys. Six of the nine more abundant proteins have [Fe-S] clusters, and four are crucial in acetate driven methanogenesis. Many proteins in higher abundance in the Fe(II)/HS- phenotype are directly or tangentially involved in methanogenesis from CO2, and only three bind Fe in any capacity ([Fe-S] cluster or otherwise). Similar patterns were present when comparing Fe(II)/Cys to FeS2. 21 proteins were significantly different between these groups. The contrast in [Fe-S] cluster binding is greater in this case, with 6 of 10 the proteins in greater abundance in Fe(II)/Cys having [Fe-S] clusters. No central carbon proteins that had a greater abundance in FeS2 were annotated as binding [Fe-S] clusters or Fe. Interestingly, nine out of the 11 proteins were predicted to bind other metal ions, including cobalt, manganese, and magnesium.

Along these lines, FeS2 and Fe(II)/HS- experimental groups looked similar with respect to their central carbon and methanogenesis metabolism. Only seven proteins were significant, four of which were greater in Fe(II)/HS- and three in FeS2. Every protein found to be more abundant in Fe(II)/HS- is predicted to bind metal ions. Three of these have auxiliary functions in glycolysis/gluconeogenesis, indicating Fe(II)/HS- grown cells could be subtly modulating their

energetics relative to FeS2 cultured cells. The three proteins higher in abundance in FeS2 are all annotated as methanogenesis enzymes, though they appear to be tangential/secondary to this process. Each is involved in a separate pathway, whose endpoint are all cofactors used by other enzymes to generate methane and energy (methanofuran, coenzyme F420, and coenzyme B). Taken together, this provides evidence that the FeS2 and Fe(II)/HS- phenotypes are similar in their energetics, especially when compared to Fe(II)/Cys cultures, but still possess subtle differences related to methanogenesis.

In general cells grown on Fe(II)/Cys had a higher abundance of key central carbon enzymes, such as acetyl-CoA synthase and carbon monoxide dehydrogenase, relative to the other two groups (Figure 5A). This is a general trend, with exceptions such as aconitate hydratase which is upregulated in the Fe(II)/HS- condition (Figure 5A). This can potentially be attributed to the additional source of organic carbon in the Fe(II)/Cys relative to the others- Fe(II)/Cys is converted to alanine during sulfur extraction, which can be used downstream as an input to the central carbon cycle(24, 25). An intriguing pattern emerges when transcriptional regulation of these proteins is compared to protein abundance. The general trend in the proteomics dataset, an upregulation in Fe(II)/Cys grown cells relative to the other conditions, is mirrored in mRNA regulation patterns for acetyl-CoA synthase and carbon monoxide dehydrogenase. The transcriptomic regulation of aconitate hydratase, however, flipped entirely. Transcripts of this protein were lowest in abundance in the Fe(II)/HS- condition indicating a disconnect between the transcriptome and proteome with respect to aconitate metabolism(17) (Figure S5).

These data also indicate cells grown on Fe(II)/Cys have a higher abundance of [Fe-S] binding proteins in their central metabolism, while the other two groups trend towards proteins

that bind metal ions (often predicted to be other than Fe). When the proteomics dataset was filtered to only include [Fe-S] cluster containing proteins, a parallel HCA mirrored results from the entire dataset with more distinct clustering patterns (Figure 5B). Clusters formed based on protein function, with a strong upregulatory pattern of central carbon enzymes seen in Fe(II)/Cys, likely due to that condition's additional source of organic carbon. Other types of Fe-S binding proteins, like those involved in cofactor biosynthesis or methanogenesis, formed their own, mixed regulation clusters. An uptick in proteins that can perform requisite metabolic functions without [Fe-S] clusters could be metabolically advantageous for cells in [Fe-S] deficient conditions (Fe(II)/HS and FeS2¬). Metabolic enzymes that utilize metal ions are generally less efficient when compared squarely to those that contain [Fe-S] clusters, but in this case could cover some of the metabolic cost associated with the extraction and conversion of insoluble Fe and S from nFeSaq clusters.

Metal Binding Proteins

To further elucidate patterns in the regulation of metal binding proteins, a separate statistical analysis was performed that included only proteins predicted to bind metal or alkali metals- Fe, Co, Ni, Cu, Mg, K, Zn and Mo (107/1019, ~11% of the dataset). The DAVID annotation of these proteins revealed that 54 contain [Fe-S] clusters, while the other 53 bind Fe cations or other metal ions. These proteins ranged from the central carbon enzymes discussed above, to copper binding chaperones (CopZ), translation initiation factors, metal ion transporters such as the ferrous iron transporter (FeoB). An ANOVA found 59 of the 107 proteins as significant (p-value <0.05), a larger proportion (55%) of dysregulation compared to the unfiltered dataset (30%). Of the 59 proteins significantly dysregulated, 26 bind [Fe-S] clusters,

16 bind magnesium ions, and 6 bind zinc (the remaining bind Co (3), Ni (3), Cu (2), and Mn, Mo, and K (1)). A parallel HCA heatmap showed perfect clustering of experimental groups, with large clusters of proteins changing in each group (Figure 6A). The HCA clustered proteins into regulatory blocks, That tended to be composed of proteins with related functions. Cells grown with Fe(II)/Cys had two groups of upregulated proteins, one containing central carbon enzymes and the other metal ion binding proteins. Likewise, Fe(II)/HS- grown cells had two clusters: one containing methanogenesis enzymes while the other contained predominantly [Fe-S] cluster binding proteins that are not involved in carbon catabolism. The cluster unique to FeS2-grown cells contained upregulated proteins that are notably absent of [Fe-S] cluster binding proteins. The cluster was instead enriched with copper binding, cofactor biosynthesis, and methanogenesis enzymes that bind metal ions (Figure 6A). When the [Fe-S] cluster binding proteins were analyzed on their own, close to half (29/54) were both significant (p-value < 0.05) and in higher abundance (FC > 2) in Fe(II)/Cys when compared to Fe(II)/HS- and FeS2. This indicates that the [Fe-S] cluster pattern observed in carbon metabolism enzymes is not echoed throughout the remainder of the dataset - the FeS2 and Fe(II)/HS- phenotypes did not show system-wide downregulation of [Fe-S] binding proteins. This can be clearly seen in the heatmap of all [Fe-S] cluster binding proteins (Figure S6). In the Fe(II)/Cys growth condition, [Fe-S] binding proteins that were upregulated relative to the other groups are concentrated in the carbon metabolic pathway. This provides further evidence of a strategic metabolic shift by the FeS2 and Fe(II)/HS- phenotypes; they may be utilizing any available [Fe-S] clusters in other biochemical pathways.

There were several metal binding proteins of interest that showed peculiar regulation patterns. This was the case when comparing FeS2 to Fe(II)/HS-. A volcano plot of the two-group T-test showed a moderate to large statistical difference between the groups in their metal binding proteins (Figure 6B). This showed that 31 proteins change significantly, with 13 upregulated in FeS2 and 19 in Fe(II)/HS-. The iron sulfur assembly protein, SufB, that utilizes cationic Fe and HS- liberated from cysteine by IscS (cysteine desulfurase) (12) was upregulated in both Fe(II)/Cys and FeS2 relative to Fe(II)/HS-. SufB's partner, ATPase (SufC) showed insignificant dysregulation across all groups. The protein responsible for transporting cationic ferrous iron, FeoB, was also upregulated in FeS2, in this case relative to both Fe(II)/Cys and Fe(II)/HS- (Figure S7). Based on the identified proteins in this analysis, FeS2 grown cells appeared to be upregulating the canonical [Fe-S] cluster synthesis pathway relative to Fe(II)/HS-, and looks similar to Fe(II)/Cys in terms of certain metal binding proteins. This could indicate a difference in the abundance of Fe cations in the FeS2 growth condition relative to Fe(II)/HS-, whether it be through the reductive dissolution process or through active Fe(II) liberation from nFeSaq clusters.

Several proteins related to metal transport, scavenging, and chelation are also changing between the groups. Iron (III) dicitrate-binding protein, which binds the ferric iron siderophore dicitrate, shows upregulation in Fe(II)/HS- and FeS2. In E. coli, this protein is part of the 5-protein iron dicitrate transport operon, which provides the means for chelated iron (III) to be transported across the membrane. Despite the similarity in name, this protein shows no sequence similarity to FecR, the analogous periplasmic iron (III) dicitrate binding protein in E.coli (26). The protein contains an iron siderophore/cobalamin periplasmic binding domain (Prosite

PS50983) and shares sequence similarity with iron (III) dicitrate binding proteins across M. barkeri strains as well as other Methanosarcina species. While it is possible this protein binds iron dicitrate, it could also bind a chelated form of iron related to the nFeS clusters or their breakdown products, since it is at similar abundance in Fe(II)/HS- and FeS2 growth conditions. CobN-like chelatase BtuS, while less well characterized, appears to have a similar function in metalloporphyrin salvaging. A 2019 study found that a bacteria (B. fragilis) required two copies of BtuS to extract essential iron from heme and subsequently transport cationic iron using FeoAB (27). Like B. fragilis, we have identified two copies of BtuS (MSBRM_0518 and _1640), both of which are upregulated in Fe(II)/Cys and FeS2 relative to Fe(II)/HS-. Due to a lack of the entire operon seen in other organisms such as B.fragilis, it is likely the BtuS proteins here are involved in a reappropriation of iron from heme to other sources as opposed to importation. A related enzyme, siroheme decarboxylase, is involved in the synthesis of heme from iron-porphyrin complexes, and also shows an upregulatory pattern in FeS2 relative to Fe(II)/HS- (Figure S7). This could mean the FeS2 phenotype is deficient in heme, relative to the Fe(II)/HS- condition, and is compensating by upregulating related biosynthesis pathways.

As observed in other functional categories, transcriptional patterns show mixed correlation with protein abundance. FeoB, for example, has identical transcriptional and protein abundance regulation. Other crucial proteins involved in iron sulfur cluster assembly are not regulated similarly between the protein and transcriptional level- the transcripts representing SufB and IscS show marked differences to their respective proteins (Figure S8). In fact SufB essentially reverses the order of abundance in the experimental groups, with Fe(II)/HS- cells having the highest number of transcripts followed by Fe(II)/Cys and then FeS2. The rampant

disconnect between regulatory tiers (e.g protein vs transcript abundance) has prompted a deeper investigation- forthcoming metabolomic and lipidomic analyses will help to establish directionality of observed regulatory patterns.

Porphyrin biosynthesis

Porphyrins are key intermediates in the synthesis of many essential cofactors, such as vitamin B12 and coenzyme F420, in addition to heme and other metal chelators. The Fe(II)/HS- condition stands out when compared to the other groups for its regulation of porphyrin biosynthesis proteins, particularly those involved with cobalamin production. Fe(II)/HS- has 7 of 8 porphyrin synthesis proteins in lower abundance when compared to both Fe(II)/Cys and FeS2. Four of these proteins are shared between the two comparisons (CobB-like chelatase, siroheme dexarboxylase MSBRM_0433, L-threonine 3-o-phosphate decarboxylase, and porphobilnogen deaminase). These four are not adjacent in the related biochemical pathways, but porphobilibogen deaminase and the siroheme decarboxylase are in the same gene neighborhood (MSBRM_0428 and MSBRM_0433, respectively).

L-threonine 3-o-phosphate decarboxylase and adenosylcobinamide-phosphate synthase are directly adjacent in the genome (MSBRM_0840,0841) and both are much more abundant in Fe(II)/Cys than in Fe(II)/HS- and FeS2. These two proteins share D-1-aminopropan-2-ol O-phosphate as a product and substrate, respectively. This molecule is an intermediate between L-threonine and vitamin B12, which could be evidence for Fe(II)/Cys grown cells shuttling L-threonine (and other amino acids) towards vitamin B12 or other porphyrin containing cofactors. Considering cells grown on Fe(II)/Cys only have access to exogenous Fe2+, they could be upregulating chelators to better store and shuttle cationic Fe.

FeS2 and Fe(II)/Cys appear relatively similar in their porphyrin metabolisms. FeS2 has proteins in higher abundance related to the synthesis of heme, while Fe(II)/Cys (along with Fe(II)/HS-) has upregulated proteins connected to the interconversion of amino acids to vitamin B12.

Quorum sensing

Another important category that shows differences in regulation is quorum sensing. Quorum sensing, generally, is how microorganisms communicate with one another, whether it be intra or interspecies. Numerous ABC transporters are found to be changing between growth conditions, namely the periplasmic binding subunits of dipeptide, oligopeptide, and nickel transporters. All of these are annotated as being related to biofilm formation and virulence. Obvious patterns emerge in the data when comparing Fe(II)/Cys, Fe(II)/HS- and FeS2 groups. Fe(II)/Cys has much lower abundances of all ABC transporters involved in quorum sensing, with fold changes beyond 130-fold, when compared to FeS2. This pattern is echoed when comparing Fe(II)/Cys to Fe(II)/HS-: all quorum sensing ABC transporters are lower in abundance. There are a total of six separate proteins that are identified as Oppa, the periplasmic binding subunit of the oligopeptide ABC transporter. These six annotations have different sequences and are located away from one another on the genome, with a few exceptions (MSBRM_0219/0220, MSBRM_0900/0901). The pairs of Oppa genes that are adjacent share remarkably similar regulation patterns and have nearly identical sequences, while other proteins distant in the genome are regulated much differently and have dissimilar sequences. The difference in sequence likely means the proteins have different oligopeptide ligands. Oppa is usually lipid anchored in the membrane, but studies have reported that a soluble form actively binds peptides

in the cytosol of other organisms (28). Our collaborators have previously noted that Methanococcus voltae (a close cousin to M.barkeri) cells grown on mineralized pyrite show increased biofilm formation and clumping (29), and the proteomes of Fe(II)/HS- and FeS2 grown M.barkeri have upregulated proteins related to biofilm formation and signaling. Perhaps the nFeS clusters available in these growth conditions act as a stimulus for biofilm related processes, such as cell clumping and the production of extracellular polymeric substance, the latter of which has been observed in SEM images of M. barkeri adhered to pyrite (17).

Discussion:

Methanosarcina barkeri has recently been shown to satisfy its Fe and S requirements by reducing FeS2 and assimilating the reduction products (12, 13, 15–17, 29), breaking a longstanding paradigm that Fe and S are biologically inaccessible in FeS¬2. Our comparative proteomics investigation sought to elucidate this mechanism by identifying key proteins and biochemical pathways that are altered in response to different Fe and S environments. Three growth conditions (Fe(II)/HS-, FeS2, and Fe(II)/Cys) were employed to probe the effects of soluble vs insoluble Fe/S sources as well as the mechanism of exogenous [Fe-S] cluster formation, facilitated either biotically through EET or abiotically. We identified over 1000 proteins, 307 of which showed significant dysregulation in at least one growth condition. This large dysregulation ratio, paired with patterns observed in multivariate statistics, indicates the proteomes from each culturing condition are dramatically different (Figures 1-2). This was evident by a lack of overlap in the PCA, perfect clustering in both HCAs, and the diversity of function in top dysregulated proteins, the extent of biological dysregulation. In order to better characterize the degree of biochemical divergence, we employed functional distribution,

biological process, and pathway analyses utilizing gene ontology (GO) annotation and KEGG pathway maps.

Remarkably, the relative functional distrubutiondistribution of dysregulated proteins mirrored that of all identified proteins, as well as all identified transcripts (Figure 3). Like enzymatic function, each fundamental biological process was represented in the observed proteomic change. Proteins involved in DNA transcription/replication, anabolic and catabolic metabolisms, translational machinery, and ribosomes, as well as structural components all play a role in M. barkeri's ability to subsist in disparate Fe/S environments. To date, this is among the deepest systemic analyses of a methanogenic archaea, with transcriptomic and proteomic data coming together to suggest phenotypic divergence.

Biochemical pathway analysis is a useful tool to assess exogenous effects on a biological system in a holistic fashion consistent with the scope of proteomics. Dysregulated proteins were grouped into the pathways they constitute to identify patterns of regulation. KEGG's pathway mapper tool (30, 31) was utilized for this purpose, grouping proteins into ten relevant pathways (Table 1). For this analysis, protein abundance differences and statistical significance were taken from pairwise t-tests (as opposed to the 3-group ANOVA used to this point) in order to pair specific differences in environment to perturbed pathways. Mirroring results from the multivariate analysis, cells grown on Fe(II)/Cys appear the most different, having generally higher numbers of proteins in each pathway comparison. For example, there are 24 and 11 ribosomal proteins that are significantly changing in the Fe(II)/Cys group relative to Fe(II)/HS- and FeS2, but only four changing when comparing FeS2 to Fe(II)/HS-. Cofactor biosynthesis sticks out as important in the described processes, with more than a dozen proteins dysregulated

in each group comparison. The next step in elucidating differences between the phenotypes involves picking apart each pathway to put the numbers in Table 1 into biological context.

We started the characterization of each phenotype with the most obvious, and sometimes surprising, category- ribosomal proteins. The Fe(II)/Cys group clusters far away from both others in its ribosomal proteomic changes. An example lies in the ribosomal proteins L23Ae and L13, which elevate to 32-fold higher abundance in the Fe(II)/Cys group relative to Fe(II)/HS- grown cells. This pattern is shared in comparison to the FeS2 condition to a lesser extent (3 and 4-fold higher, respectively). Much of the ribosomal dysregulation appears operonic in nature, with 13 proteins in the MSBRM_0579 to _0612 range sharing protein and transcriptional regulation patterns (Figure 4). There are even differences between ribosomal subunits when comparing FeS2 to Fe(II)/HS- grown cells. It is peculiar, especially given the predicted "structural ribosomal protein" annotation of L23Ae and L13. These are some of the most drastically changing proteins in the entire dataset. Moonlighting activity of analogous ribosomal proteins has been observed in other organisms, notably humans and metazoans (22, 23). Perhaps they serve similar ribosome-independent functions in M.barkeri.

As the name would suggest, methanogenesis is often seen as the center point for metabolism in M.barkeri and related methanogens. Thus, it is unsurprising to see this category changing in what we are classifying as a phenotypic shift. Like the previous category, Fe(II)/Cys grown cells appear worlds apart when compared to both other growth conditions. Table 1 denotes 24 and 21 proteins being significantly different in Fe(II)/Cys compared to Fe(II)/HS- and FeS2, respectively (Figure 5A). A pattern in Fe-S cluster binding appears when peering into central carbon/methanogenesis pathway changes imparted by Fe/S source differences. Notably,

cultures of Fe(II)/Cys have higher abundances of carbon metabolism proteins that bind Fe-S clusters, while the other two groups upregulate metal-ion binding proteins. Fe(II)/Cys grown cells also appear to be upregulating central carbon enzymes, which can be explained by their access to an additional source of organic carbon. Since it is the only source of sulfur, Cys is quickly stripped of its sulfur by SufS. This likely generates an excess of alanine, which can be directly converted to pyruvate by an aminotransferase. Therefore, an upregulation of central carbon enzymes would be required to process this influx. There also appears to be changes in methanogenesis source between Fe(II)/HS- and Fe(II)/Cys grown cells which could be a direct result of the metal binding patterns. Interestingly, the FeS2 and Fe(II)/HS- groups look similar in their energetics. This is in contrast to what we expected given the FeS2 group must facilitate additional electron transfers in the reduction of FeS2 to pyrrhotite.

Considering the experimental design, and the organism of interest, it is of little surprise that major proteomics differences between the phenotypes were found in proteins that bind metal. The proportion and importance of metal binding proteins in the dataset warranted a separate statistical analysis. 11% of the proteins identified are predicted to have metal binding domains, whether it be metal ions such as Co2+ or Zn2+, or the ever-important Fe-S clusters. 59 of these show statistical significance (approximately 20% of the significantly dysregulated proteins in the whole dataset bind metals). The upregulation of Fe-S cluster binding proteins is shockingly absent when looking at all metal binding proteins (Figure 5B). Only half of all Fe-S cluster binding proteins have a higher abundance in Fe(II)/Cys relative to the other two groups, rendering the pattern heavily concentrated in central carbon metabolism. In most of the other categories, one group usually sticks out while the other two look similar. That is not the case

with metal binding proteins. Figure 6A shows clusters of proteins with specific attributes changing in each growth condition. Despite both having exogenous insoluble Fe-S clusters, the Fe(II)/HS- and FeS2 groups show moderate differences in their metal binding profiles (Figure 6B). Key enzymes in the Fe-S cluster synthesis pathways are regulated in this regard including ferrous iron transporter FeoB, cysteine desulfurase SufS and iron sulfur cluster assembly protein SufB. This represents the major components of the route from exogenous Fe(II)aq to biologically relevant iron sulfur clusters, and all are higher abundance in FeS2 than in Fe(II)/HS-. Interestingly, the Fe(II)/Cys and FeS2 groups look very similar in their regulation of SufS, the enzyme that frees S from Fe(II)/Cys for use in Fe-S clusters. This could indicate a difference in the route from exogenous FeSn clusters to Fe-S cluster cofactors. A metalloporphyrin salvage enzyme (BtuS) likely related to stripping iron from heme typically seen in iron limitation in other organisms shows upregulation in Fe(II)/Cys and FeS2. Neither group shows stress indicators or attenuated growth indicative of iron limitation, so it is likely related to heme processing and interconversion to other forms of iron.

Other categories that help explain the statistical differences between groups, such as quorum sensing and porphyrin biosynthesis echo the results from the remainder of the investigation. Peptide transporters that function in biofilm formation, virulence, and other cell-cell communications are found ballooning in abundance over 130-fold in some comparisons. Cell surface proteins with PKD domains that mediate extracellular matrix interactions are also regulated in response to Fe and S environment. The extracellular matrix clearly plays a role in cellular-mineral interactions ((17), so the further characterization of these CSP and transporters

will aid in elucidating the communicative portion of M.barkeri's ability to clump together and thrive on the surface of minerals.

This study provides the groundwork for the full elucidation of the mechanism utilized by Methanosarcina barkeri to access nutrients (Fe, S) from of inert, Earth abundant FeS minerals. This work builds upon many previous analyses, including an analogous investigation on the proteome of Methanococcus voltae (32). Steward et al showed a greater degree of dysregulation (40% of identified proteins) in the comparison of M. voltae grown on FeS2 compared to Fe(II)/HS-. Similar to the observations in this study, M. voltae show dysregulation in proteins ranging in function from ribosomal structure components to metal binding and cellular energetics. Several key proteins show similarity in their regulatory patterns, including the increased abundance of FeoB in FeS2 exposed cells. Comparable multivariate statistics on the proteomes of two organisms from different orders, in the same environmental conditions, indicate these findings likely underpin the fundamental biology of methanogenic archaea; Fe and S environmental availability alone can dictate the observed phenotype of these lifeforms.

Presented here is a bounty of enzymes and proteins primed for further investigation via knockout studies, direct assays, and more. Top targets include: ribosomal proteins changing by more than 30-fold, which are likely to be interacting with the genome or RNAs in a regulatory fashion, peptide transporters related to cell-cell or cell-environment communication, and metal binding proteins like FeoB, SufS, and the iron (III) dicitrate binding protein. The proteomic evidence presented here suggests that the three environments provided induce unique and distinct phenotypes of M.barkeri. Transcriptomics data presents an interesting, mixed corroboratory relationship with the proteomics analysis. The abundance of some proteins precisely mirror the

pattern seen in the transcripts while others seem either completely flipped or entirely unrelated. Forthcoming metabolomic and lipidomic studies on this system, in these exact growth conditions, will present a holistic view of methanogen biology and will provide grounds to present a model of Fe-S acquisition dynamics in M.barkeri.

Methanosarcina barkeri has evolved to survive in a wide range of habitats, spanning multitudes of Fe-S equilibria. It is a clear advantage to have a broad toolset for extracting and utilizing these elements, and this study suggests the methanogen does so from the ground up; changing its entire cellular makeup to compensate for changes in just two elements. Understanding how this organism ticks will provide future scientists a foundation for developing novel bio-mining technologies, and will also provide a clear window into the biology of the early earth where many of the discussed biological processes evolved.

Methods

Cell cultivation

M. barkeri strain MS was obtained from the American Type Culture Collection (ATCC-43569). Cells were grown following previously described methods (Spietz et al 2022 Frontiers) in anoxic, low salinity medium and without added Fe or S. Cultures for proteomics analysis were grown with three different Fe and S sources: 1) 2 mM cysteine and 20 µM FeCl2, 2) 2 mM Na2S and 20 µM FeCl2, or 3) 2 mM synthetic nanoparticulate FeS2 prepared as previously described (Payne et al 2021, Spietz et al 2022). When cultures reached mid-log phase, as determined by CH4 and DNA production monitored using previously described methods (Spietz et al 2022), cells were harvested under anoxic headspace by centrifugation at 4,600 x g for 30 minutes at 4 °C, then supernatant was decanted in an anaerobic chamber before submerging the cell pellet in

liquid N2 to flash freeze. Frozen cell pellets were stored at -80 °C until further processing for proteomics.

Proteomics sample preparation

Cultures were prepared for ongoing multi-omics analysis where proteins, metabolites, and lipids were extracted in parallel from the same cell pellets. Cell Lysis: Cell pellets were resuspended in 5mL 50mM ammonium bicarbonate (NH4HCO3) with protease inhibitor mix (Complete Mini EDTA Free Protease Inhibitor Cocktail, Roche) ph 8. Samples were lysed using an ultrasonic homogenizer at 40% power output for 15 minutes. Resulting lysate was centrifuged at 4,000G for 15 minutes. The supernatant was isolated for soluble protein preparation. Protein concentration was estimated using nanodrop spectrophotometer (Thermo-Fisher), and volume equivalent to 100µg of protein was aliquoted. Reduction and Alkylation: DTT was added to a final concentration of 5mM and incubated at room temperature for 30 minutes. Iodoacetamide (IAA) was added to a final concentration of 15mM and placed into darkness for 30 minutes at room temperature. Filter assisted tryptic digest: A filter assisted trypsin digestion was performed on a 3kDa spin filter at 1:20 m/m trypsin:protein overnight at 37C. Trypsin was quenched through addition of TFA (0.1%) and spun to filter tryptic peptides.

Shotgun Proteomics data Acquisition

Tryptic peptides were separated by reverse phase XSelect CSH C18 2.5 um resin (Waters) on an in-line 150 x 0.075 mm column using an UltiMate 3000 RSLCnano system (Thermo). Peptides were eluted using a 90 min gradient from 98:2 to 65:35 buffer A:B ratio. (Buffer A = 0.1% formic acid, 0.5% acetonitrile, Buffer B = 0.1% formic acid, 99.9% acetonitrile). Eluted peptides were ionized by electrospray (2.4 kV) followed by mass

spectrometric analysis on an Orbitrap Eclipse Tribrid mass spectrometer (Thermo-Fisher). MS data were acquired using the FTMS analyzer in profile mode at a resolution of 120,000 over a range of 375 to 1200 m/z. Following HCD activation, MS/MS data were acquired using the ion trap analyzer in centroid mode and normal mass range with a normalized collision energy of 30%. Proteins were identified by database search using MaxQuant (Max Planck Institute) label-free quantification with a parent ion tolerance of 2.5 ppm and a fragment ion tolerance of 0.5 Da. Scaffold Q+S (Proteome Software) was used to verify MS/MS based peptide and protein identifications. Protein identifications were accepted if they could be established with less than 1.0% false discovery and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [Anal. Chem. 75: 4646-58 (2003)]. All protein identifications utilized a reference genome acquired from NCBI-Bioproject (accession # PRJNA230939).

Statistical, functional, and pathway analyses

Data from Scaffold Q + S (Proteome software) represent extracted protein intensities from the top 3 (based on sequence confidence) peptides for each protein. Raw data was exported from scaffold and uploaded to Metaboanalyst 5.0 (ref)for statistical analysis. Data underwent a quantile normalization followed by a log10 transformation. Data is shown as auto-scaled values (mean centered, divided by standard deviation of each variable). ANOVA significance threshold was set to p-value<0.05, pairwise t test significance threshold was increased to p<0.1. Heatmaps employed Ward's method of HCA and utilized Euclidian distances. DAVID(21, 33)and KEGG(30, 31, 34)were used for protein characterization, categorization, and pathway analysis.

Data availability:

Data is available in the MassIVE repository (Accession info will be added when upload is approved)

Acknowledgments:

We thank the IDEA National Resource for Quantitative Proteomics, University of Arkansas Medical School, Little Rock, AR

Proteomics, MetabolomicsMetabolomics, and Mass Spectrometry Facility, Montana State University, Bozeman MT

Figures and Tables

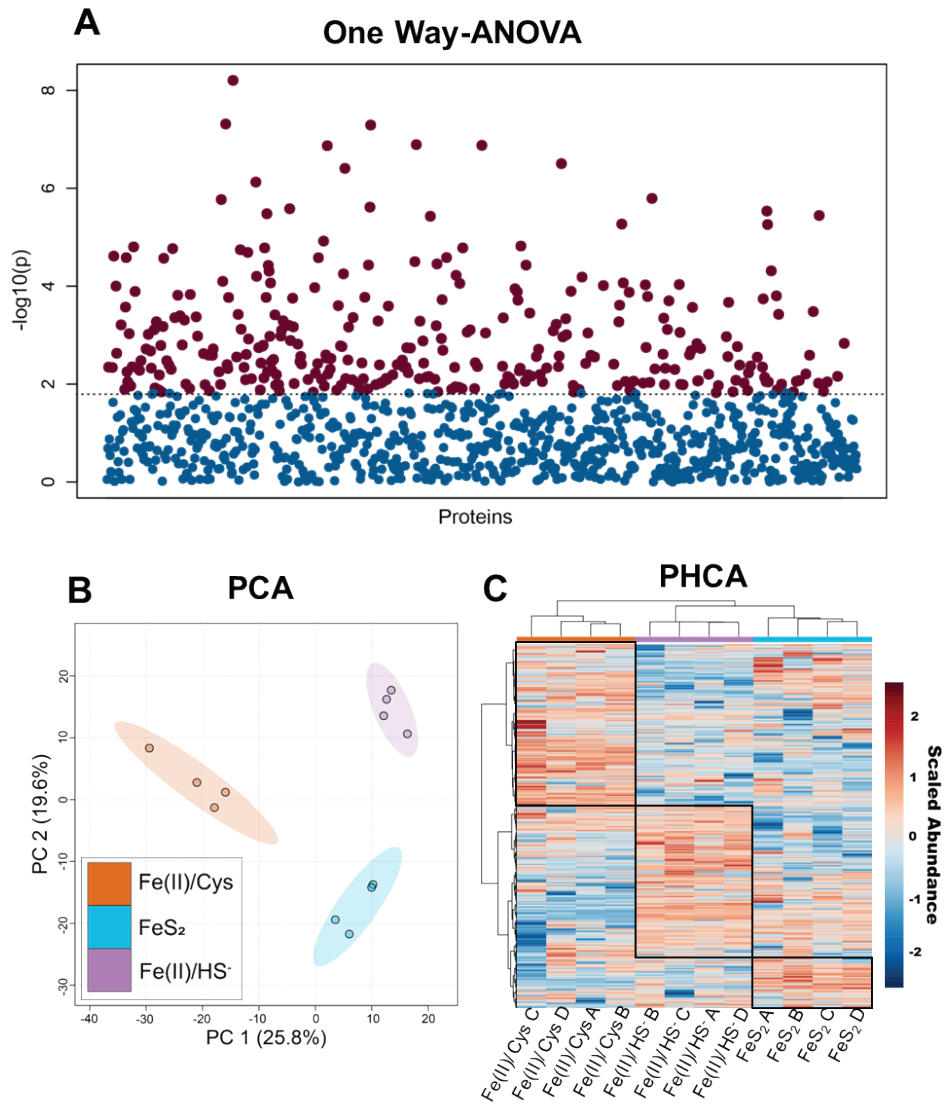


Figure 1: Multivariate statistical analysis of proteomics data. A: One-way ANOVA of protein abundance for 1019 identified proteins. Of the total, 307 proteins meet the criteria for statistical significance p-value <0.05) (red). The remaining 712 proteins do not meet this criteria (blue). B: Principal Component analysis- PC1 (25.8% of variation) and PC2 (19.6%) coordinates of PCA between cysteine (Fe(II)/Cys, orange), pyrite (FeS2, blue), and sulfide (Fe(II)/HS-, purple) grown cells. Colored ovals represent 95% confidence interval. C: Parallel hierarchical clustering analysis (PHCA) heatmap. Blocks of coregulated proteins are outlined by black boxes. Dendrogram of samples (x axis) compared with dendrogram of proteins (y axis). A-D indicate biological replicate cultures of each condition. Red and blue indicate abundance above or below average for each protein.

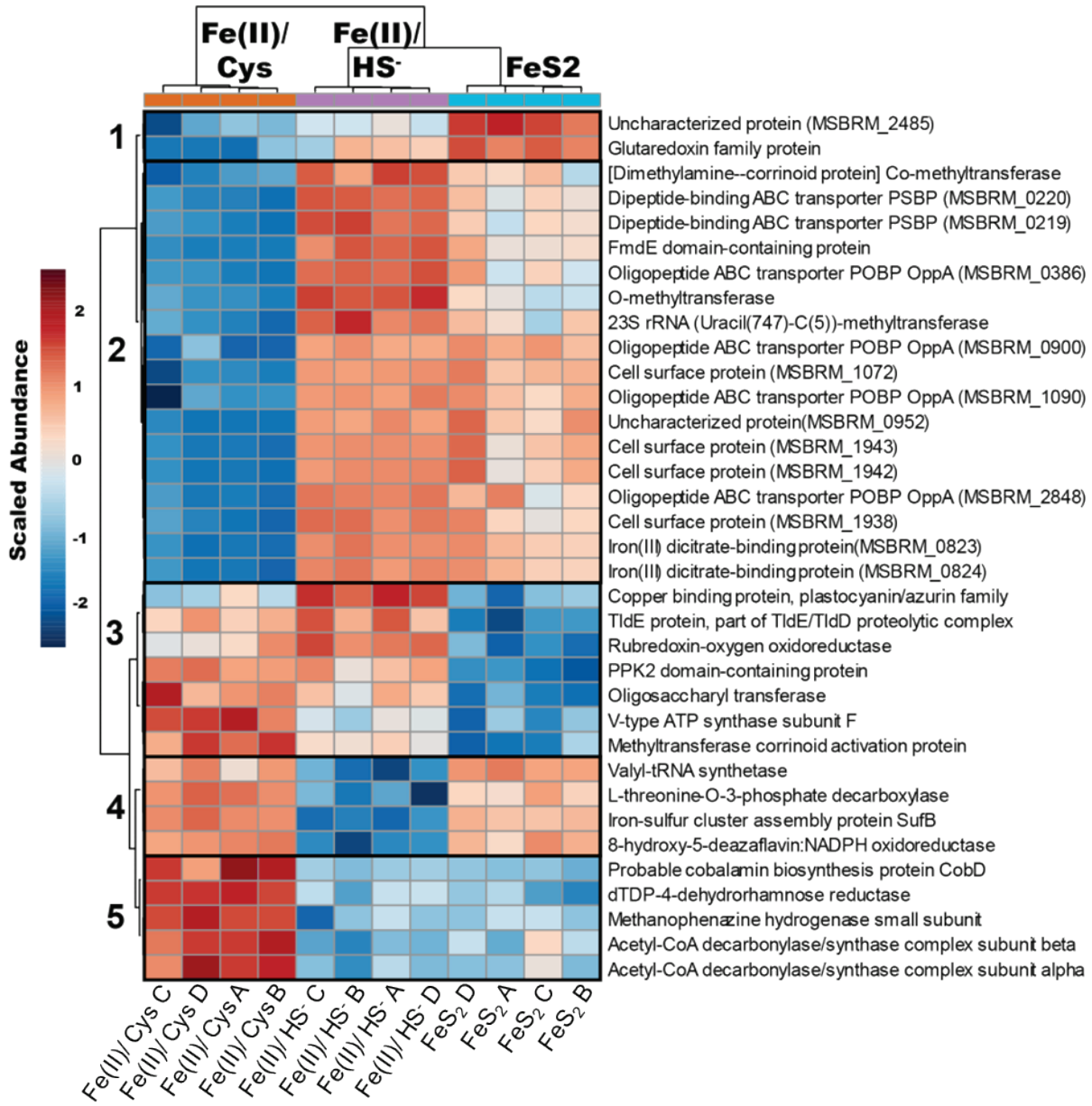


Figure 2. Parallel hierarchical clustering (PHCA) heatmap of the top 35 proteins (filtered by lowest p-value, ANOVA). Dendrograms on each axis employ Euclidian distance and a ward clustering algorithm. Y axis dendrogram clusters proteins by abundance, while the x axis clusters samples (biological replicates marked A-D) based on protein abundance patterns. Each block displays the relative, auto-scaled abundance in each sample (red> average, blue < average). Proteins fall into 1 of 5 distinct clusters (1-5). (PSBC= periplasmic substrate-binding component, POBP = periplasmic oligopeptide-binding protein)

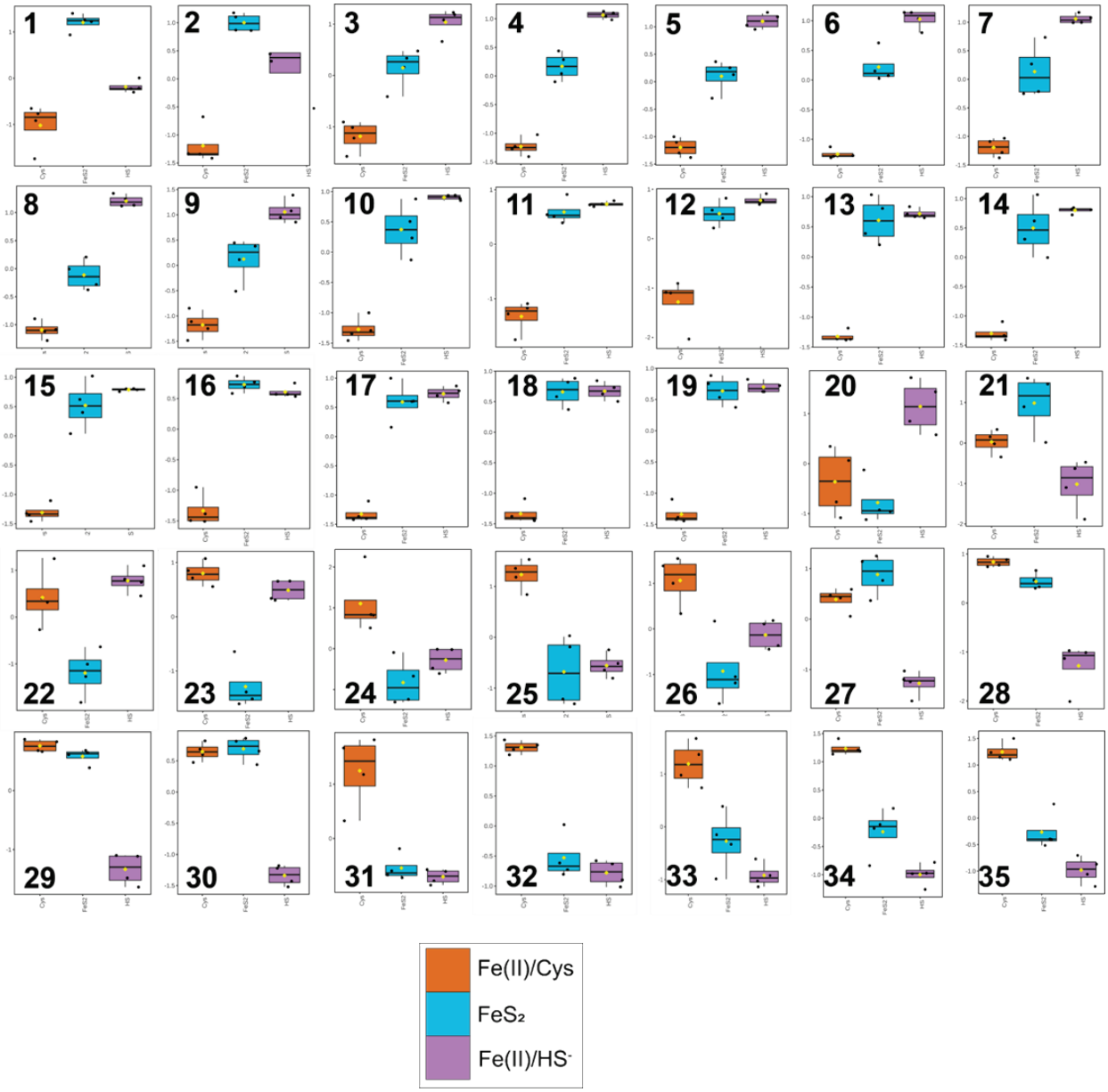


Figure S1: Box and whisker plots corresponding to the top 35 proteins (filtered by p-value). Abundance Values are auto-scaled

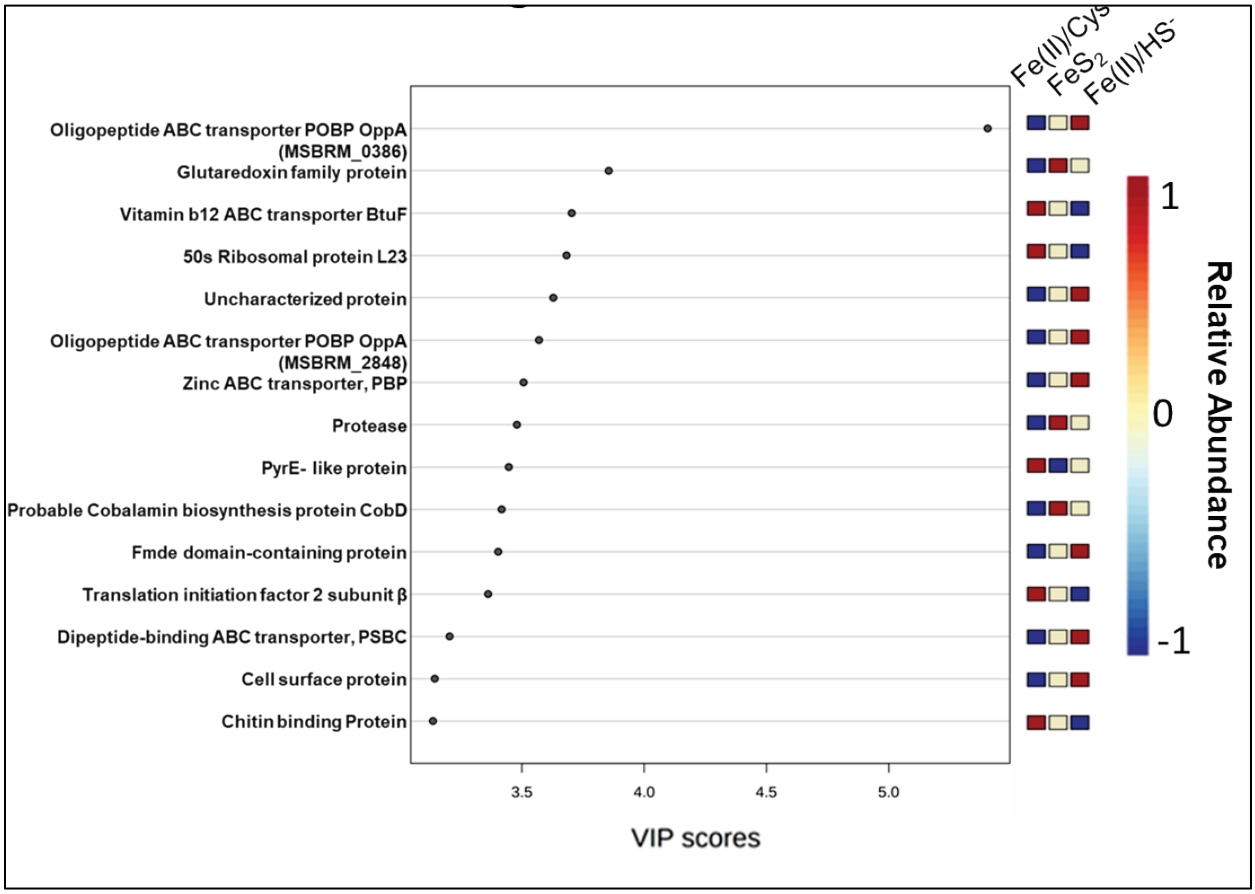
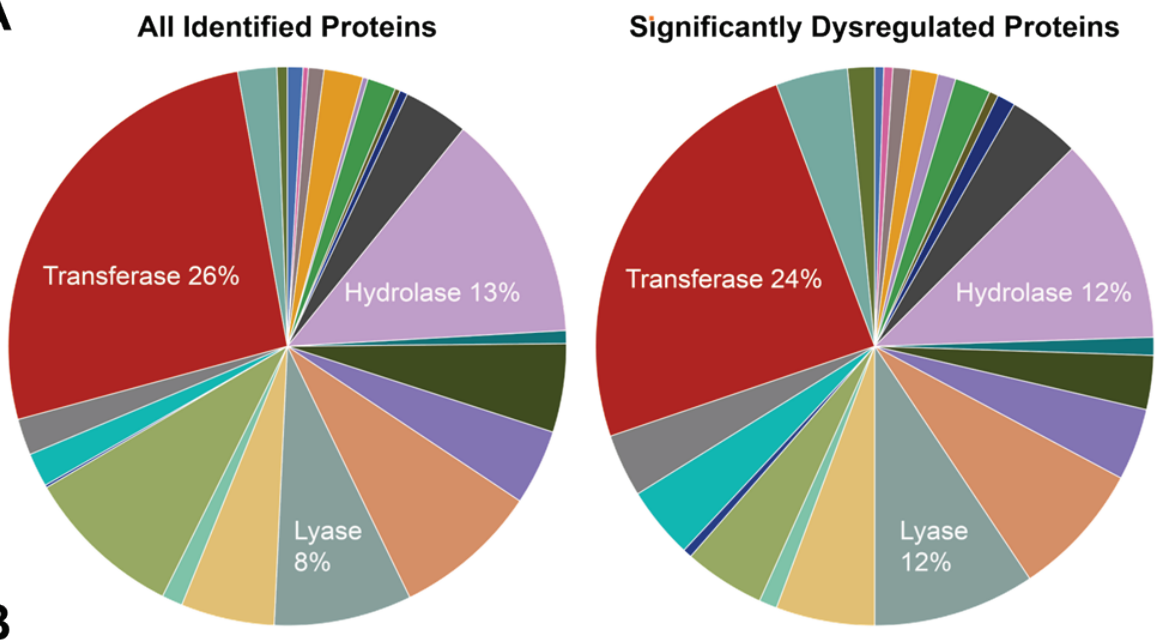


Figure S2: PLSDA VIP plot of the top 15 proteins, filtered by VIP score. Legend on the right displays the relative abundance of the protein in each group.

A



B

DNA Replication/Transcription	Translation/Chaperone	Energy Metabolism	Secondary Metabolism	Cell transport/Structural/Misc	Uncharacterized
14	45	47	93	32	36

Figure 3. Proteomic functional distribution and biological process analyses. A: Pie charts of the distribution of enzymatic functions in all identified proteins (left) and 307 significantly dysregulated proteins (right) from the ANOVA. DAVID was used for protein annotation, displayed are keyword (KW) numbers associated with specific enzymatic activity. B: Table of biological processes of dysregulated proteins. Significantly dysregulated proteins were grouped into categories based on the biological process they contribute to. Displayed are the number of proteins that correspond to each category which are also statistically significant.

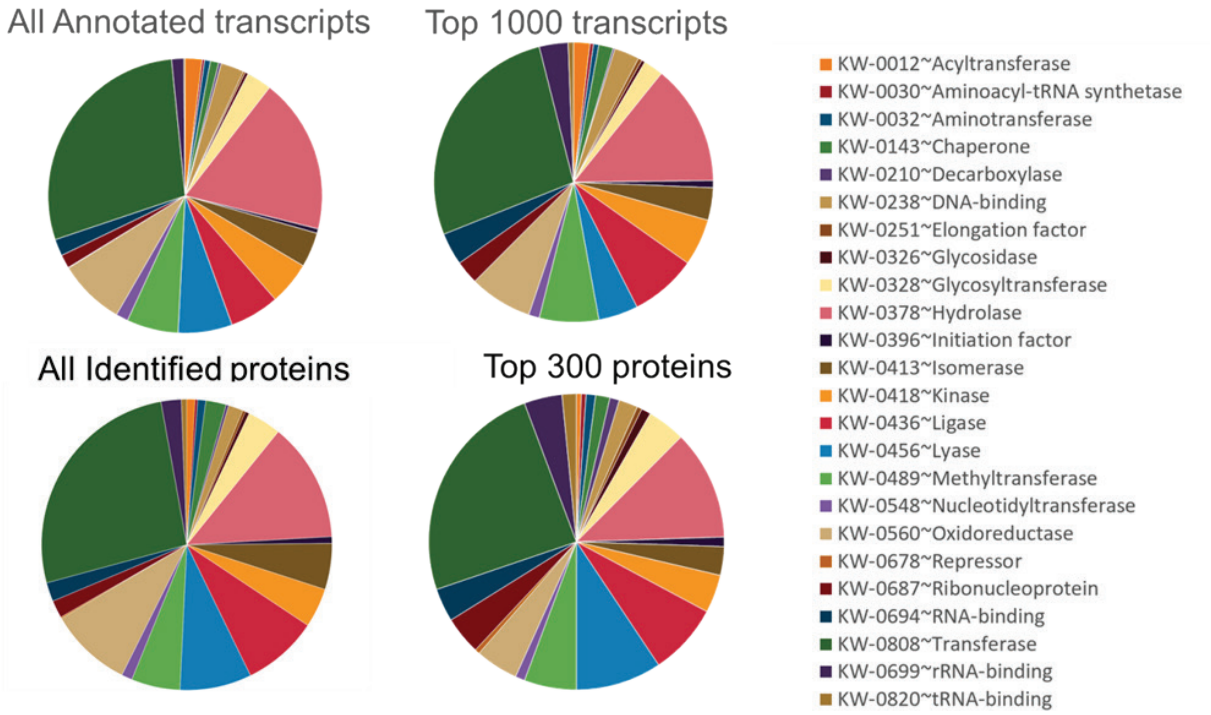


Figure S3: Whole dataset functional distribution characterization. A: Functions of encoded proteins of all detected transcripts (3000), B: top (filtered by p-value) 1000 transcripts, C: All identified proteins and D: Top 300 proteins-filtered by p-value. Transcript number in B was selected to reflect the top 1/3 of the dataset so as to allow comparison to significant proteins (307/1019).

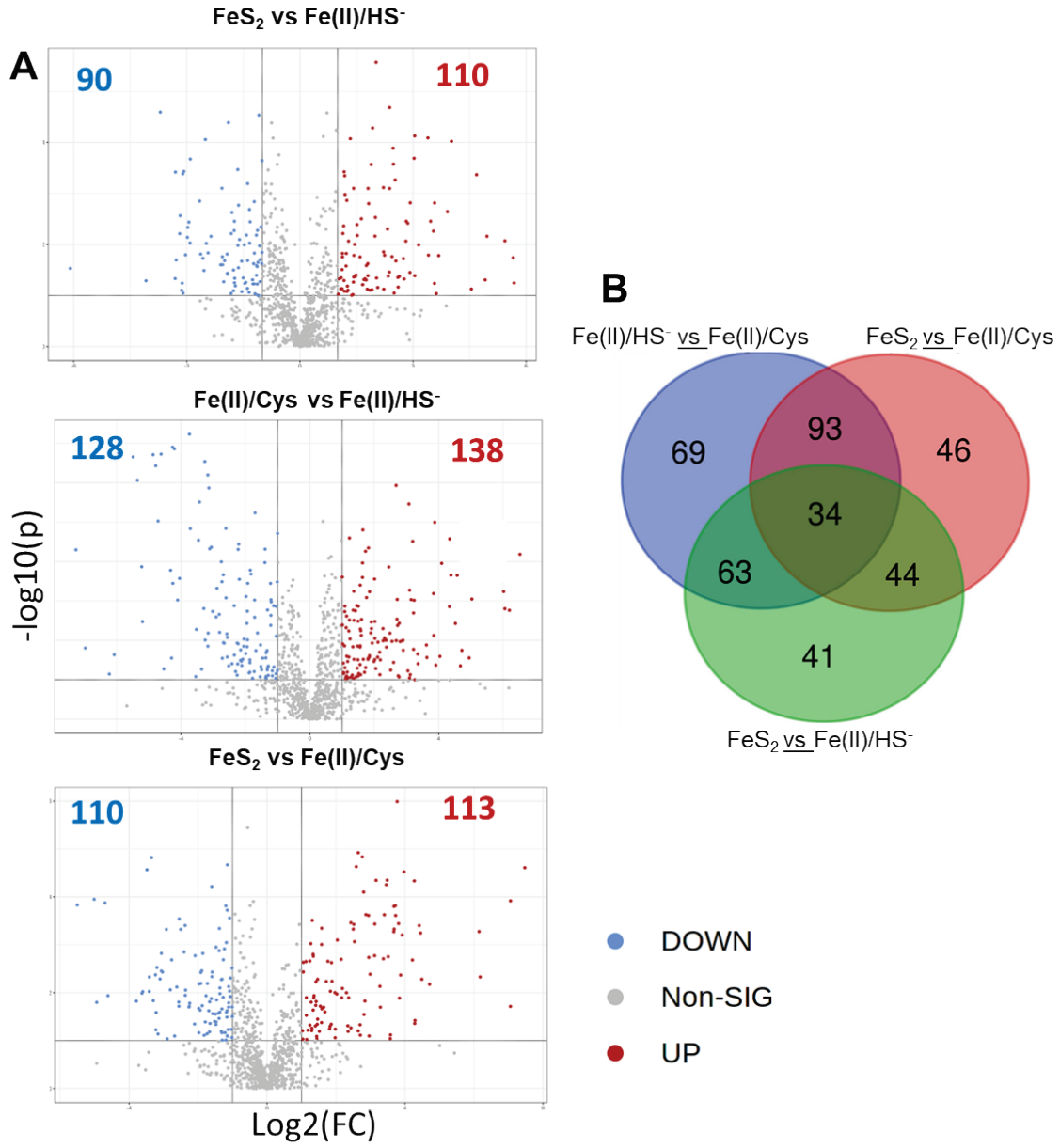


Figure S4: Two-group proteomic statistical analyses. A: volcano plots comparing fold change (Log2(FC)) with statistical significance (-log10(p)) for each comparison. Blue/leftward proteins have lower abundance in specified comparison direction (e.g FeS2 vs Fe(II)/HS, 90 proteins have lower abundance in FeS2 when compared to Fe(II)/HS). B: Venn diagram comparing significant proteins in each comparison. 34 proteins were found to be significantly changing (p-value<0.05) by at least 2 fold in all three, two group comparisons. 69 proteins were found to be uniquely changing when comparing Fe(II)/Cys to Fe(II)/ HS, 46 when comparing FeS2 to Fe(II)/ Cys, and 41 when comparing FeS2 to Fe(II)/ HS.

	# of mapped, significnatly dysregulated proteins		
Pathway Annotation	**FeS2 vs Fe(II)/HS-**	**Fe(II)/Cys vs Fe(II)/HS-**	**FeS2 vs Fe(II)/Cys**
General Metabolic Pathways	48	80	60
Biosynthesis of secondary metabolites	23	34	24
Cofactor Biosynthesis	21	22	14
Microbial metabolism in diverse environments	11	29	26
Amino Acid Biosynthesis	9	16	12
Porphyrin metabolism	8	8	5
Methane/carbon metabolism	7	24	21
Ribosome	4	24	11
Quorum sensing	4	11	11
Phe,Tyr,Trp Biosynthesis	1	8	4

Table 1.Significantly Dysregulated proteins in various KEGG biochemical pathways

Fe(II)/Cys vs Fe(II)/HS-			
Common ID	Uniprot accession	Locus Tag	FC (Fe(II)/Cys vs Fe(II)/HS-)
LSU ribosomal protein L18e	A0A0E3QSK3	MSBRM_0476	5.984
LSU ribosomal protein L13Ae (L13p)	A0A0E3QSE2	MSBRM_0477	31.533
SSU ribosomal protein S16e (S9p)	A0A0E3LMR4	MSBRM_0478	4.0007
SSU ribosomal protein S5e (S7p)	A0A0E3QSN1	MSBRM_0892	2.8255
SSU ribosomal protein S20e (S10p)	A0A0E3QTM0	MSBRM_0895	2.0995
LSU ribosomal protein L7Ae	A0A0E3QS87	MSBRM_1141	2.5834
LSU ribosomal protein L37Ae	A0A0E3QWH0	MSBRM_2032	2.5233
SSU ribosomal protein S9e (S4p)	A0A0E3QVW3	MSBRM_2579	2.0064
SSU ribosomal protein S18e (S13p)	A0A0E3QVN2	MSBRM_2580	2.3713
LSU ribosomal protein L7e (L30p)	A0A0E3QVP0	MSBRM_2592	2.8989
SSU ribosomal protein S2e (S5p)	A0A0E3LNY7	MSBRM_2593	2.0107
LSU ribosomal protein L5e (L18p)	A0A0E3QXU4	MSBRM_2594	4.8026
LSU ribosomal protein L9e (L6p)	A0A0E3QVX4	MSBRM_2597	2.6412
SSU ribosomal protein S15Ae (S8p)	A0A0E3LNY8	MSBRM_2598	2.0548
LSU ribosomal protein L11e (L5p)	A0A0E3QXV0	MSBRM_2600	2.3044
LSU ribosomal protein L23e (L14p)	A0A0E3LNY9	MSBRM_2603	4.8725
LSU ribosomal protein L35e (L29p)	A0A0E3QVQ1	MSBRM_2605	5.7881
LSU ribosomal protein L8e (L2p)	A0A0E3QXP2	MSBRM_2609	2.0968
LSU ribosomal protein L23Ae (L23p)	A0A0E3QVX9	MSBRM_2610	32.074
LSU ribosomal protein L3e (L3p)	A0A0E3QXV8	MSBRM_2612	2.102
LSU ribosomal protein L31e	A0A0E3QYR2	MSBRM_3025	3.3258
LSU ribosomal protein P0 (L10p)	A0A0E3QZ29	MSBRM_3175	3.7944
LSU ribosomal protein L10Ae (L1p)	A0A0E3QZ01	MSBRM_3176	6.8026
LSU ribosomal protein L12e (L11p)	A0A0E3QZ72	MSBRM_3177	2.9381
FeS2 vs Fe(II)/HS-			
Common ID	Uniprot accession	Locus Tag	FC (FeS2 vs Fe(II)/HS-)
LSU ribosomal protein L7e (L30p)	A0A0E3QVP0	MSBRM_2592	2.1247
LSU ribosomal protein L5e (L18p)	A0A0E3QXU4	MSBRM_2594	2.4504
LSU ribosomal protein L35e (L29p)	A0A0E3QVQ1	MSBRM_2605	2.8466
LSU ribosomal protein L10Ae (L1p)	A0A0E3QZ01	MSBRM_3176	2.243
FeS2 vs Fe(II)/Cys			
Common ID	Uniprot accession	Locus Tag	FC (FeS2 vs Fe(II)/HS-)
LSU ribosomal protein L18e	A0A0E3QSK3	MSBRM_0476	0.33557
SSU ribosomal protein S16e (S9p)	A0A0E3LMR4	MSBRM_0478	0.48208
LSU ribosomal protein L37Ae	A0A0E3QWH0	MSBRM_2032	0.28885
LSU ribosomal protein L15e	A0A0E3QUH1	MSBRM_2041	0.08126
SSU ribosomal protein S3Ae	A0A0E3QVD5	MSBRM_2372	0.49734
LSU ribosomal protein L23e (L14p)	A0A0E3LNY9	MSBRM_2603	0.29601
LSU ribosomal protein L8e (L2p)	A0A0E3QXP2	MSBRM_2609	0.45782
LSU ribosomal protein L23Ae (L23p)	A0A0E3QVX9	MSBRM_2610	0.29568
LSU ribosomal protein L31e	A0A0E3QYR2	MSBRM_3025	0.40885
LSU ribosomal protein P0 (L10p)	A0A0E3QZ29	MSBRM_3175	0.45276
LSU ribosomal protein L10Ae (L1p)	A0A0E3QZ01	MSBRM_3176	0.36141

Table 2. Ribosomal proteins showing significant dysregulation between Fe(II)Cys vs Fe(II)/HS-

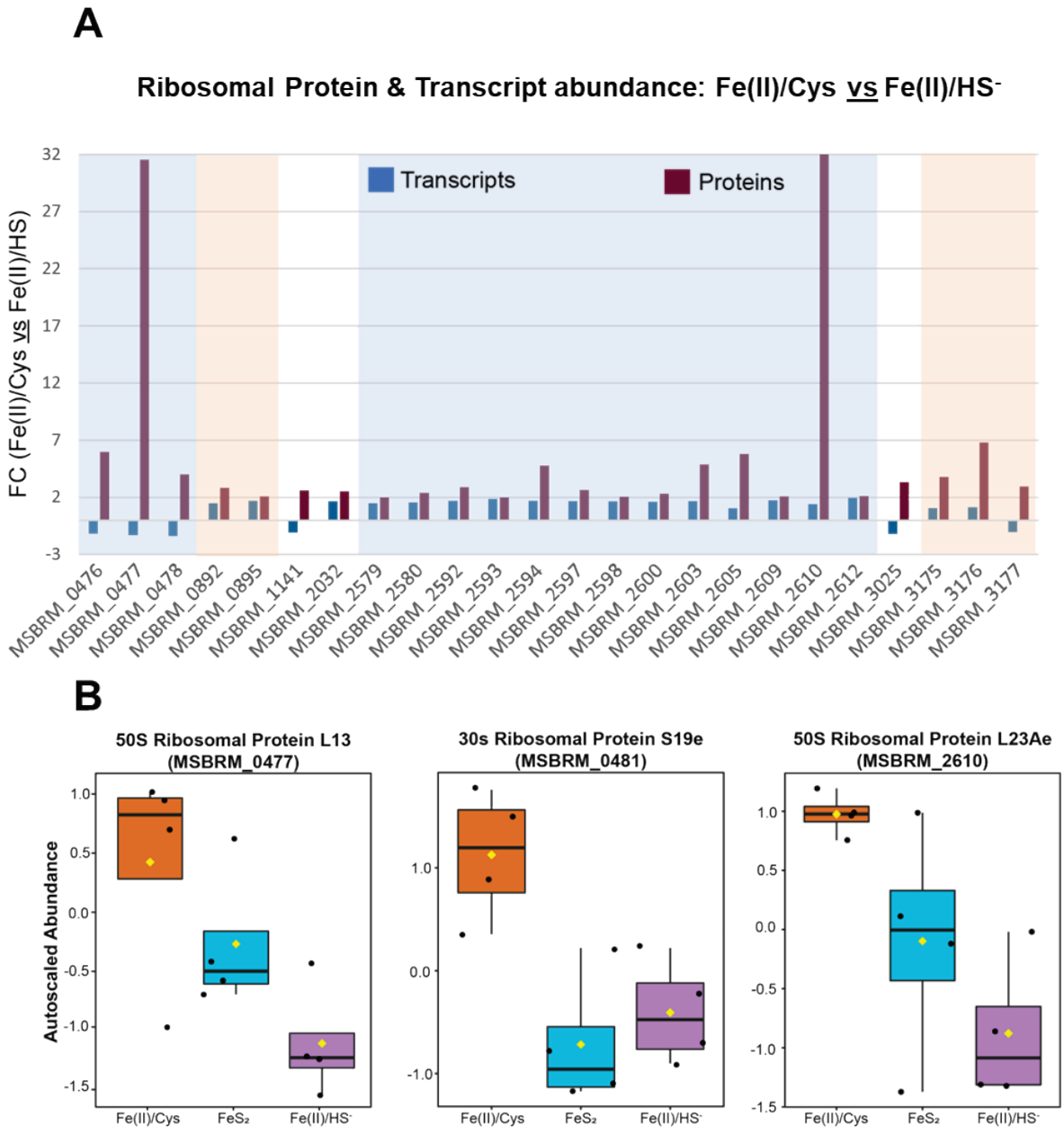


Figure 4. Ribosomal protein and transcript regulatory patterns. A: Protein (red) and transcript (blue) FC of all significant ribosomal proteins when comparing Cys to HS. Gene operons are color blocked. B: Box and whisker plots showing auto-scaled abundance measurements of three ribosomal proteins displaying interesting regulatory patterns (30s RP S19e, 50S RP L13Ae, 50S ribosomal protein L23Ae).

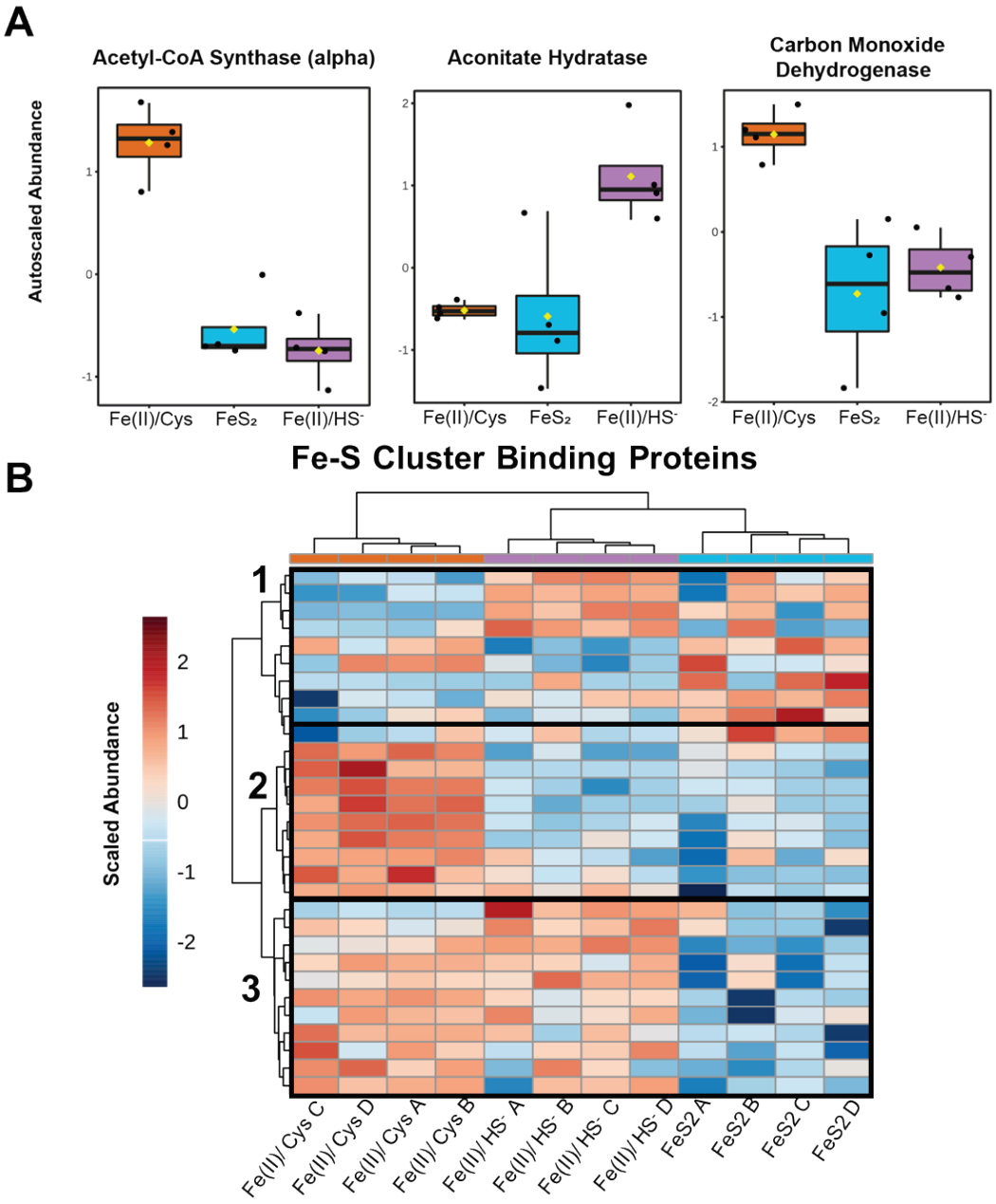


Figure 5: Regulation of central carbon metabolism iron-sulfur binding proteins A: Box and whisker plots showing auto-scaled abundance of three relevant carbon metabolism enzymes. Proteins were selected to highlight differential regulation patterns. B: Hierarchical clustering heatmap of the top 30 (p value) iron-sulfur binding proteins. Clusters 1-3 from the HCA represent 3 classes of enzymes: 1: Cofactor assembly and miscellaneous Fe-S cluster binding proteins, 2: carbon metabolism related enzymes. 3: methanogenesis and cofactor biosynthesis proteins.

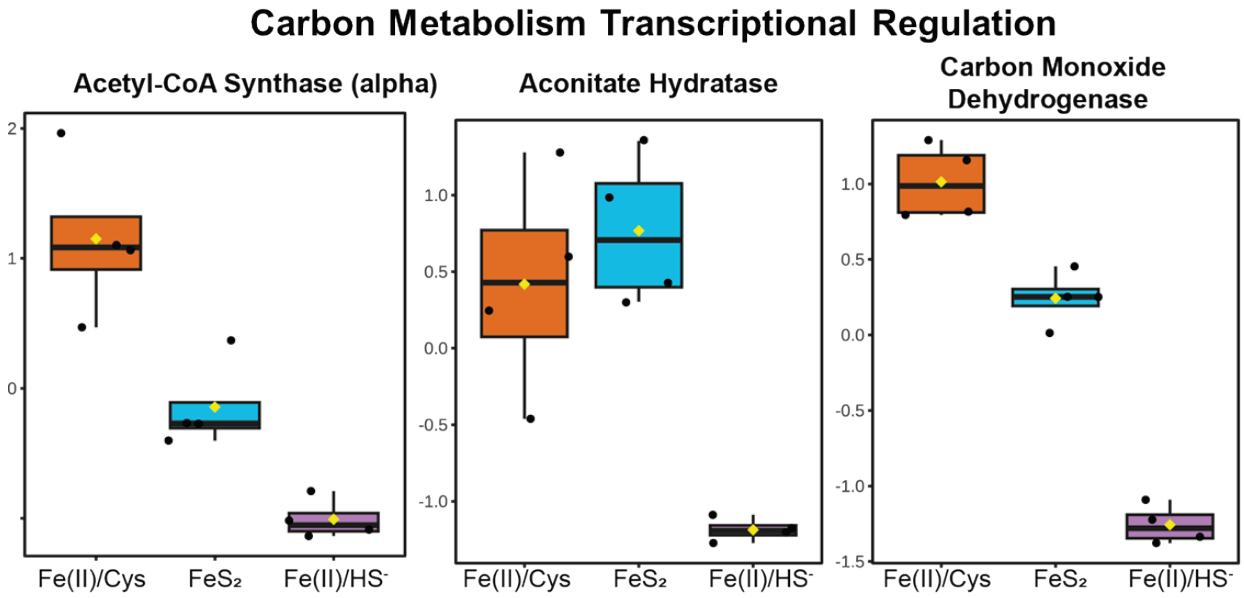


Figure S5: Central carbon metabolism transcriptional regulation. Regulatory patterns of the transcripts of 3 carbon metabolism enzymes. Box and whisker plots show scaled abundance in each condition.

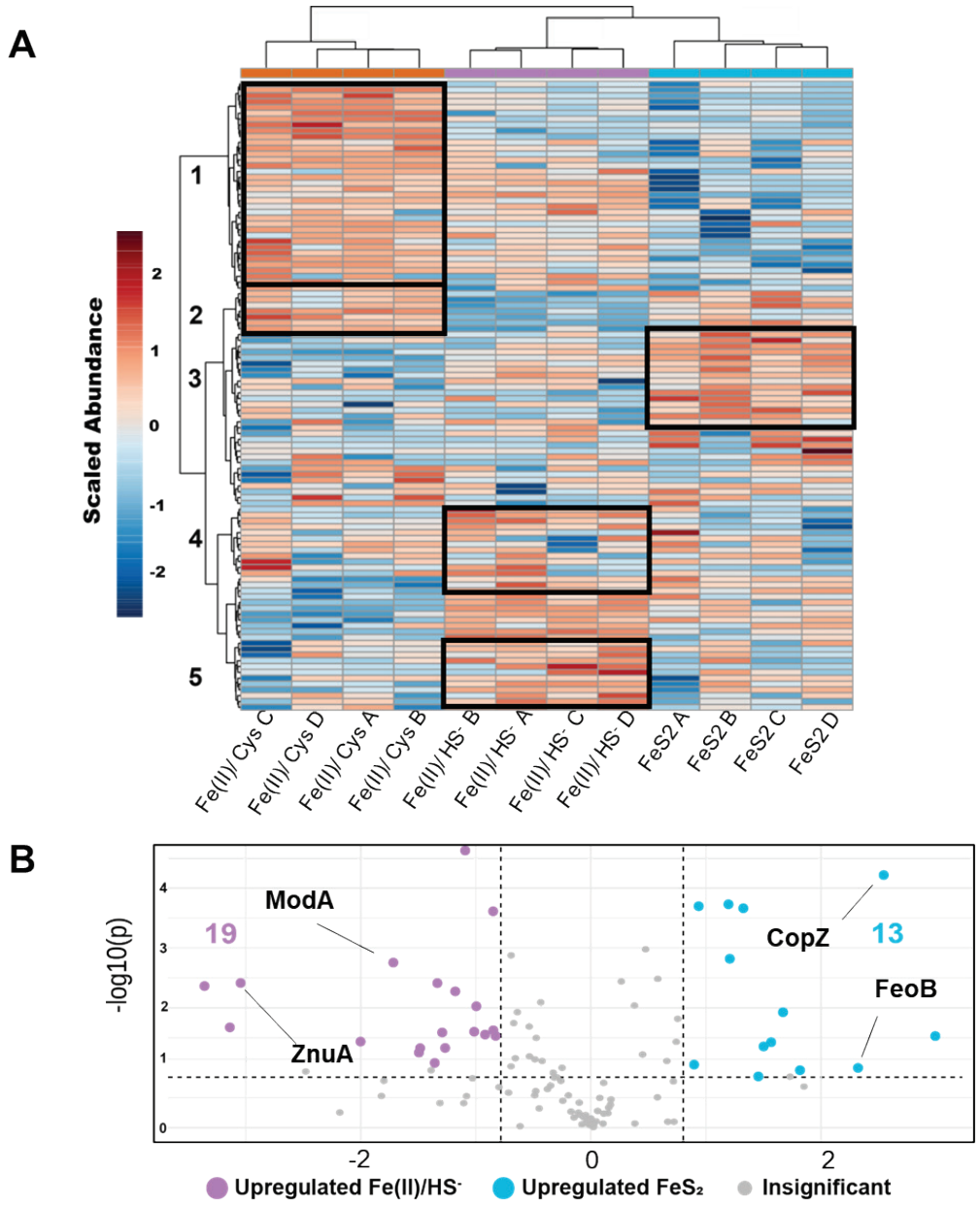


Figure 6: Metal binding protein differential analysis. A: Hierarchical clustering analysis, showing clustering of metal binding archetypes based on their regulatory patterns. Cluster 1 contains a high proportion of Fe-S cluster binding and carbon metabolism proteins, cluster 2 is comprised of all metal ion binding motifs, 3 has a low Fe-S cluster binding representation and heavy metal binding motifs, cluster 4 is comprised mainly of methane/carbon metabolism proteins, and cluster 5 is mainly Fe-S cluster binding protiens. B: Volcano plot of the FeS2 vs HS comparison. FC is plotted against statistical significance (-log10(p)). 19 proteins are upregulated in HS while 13 are upregulated in FeS2. Representative metal binding proteins are highlighted (Copper Chaperone CopZ, Ferrous iron transporter FeoB, Zinc ABC transporter ZnuA, Molybdenum ABC transporter ModA).

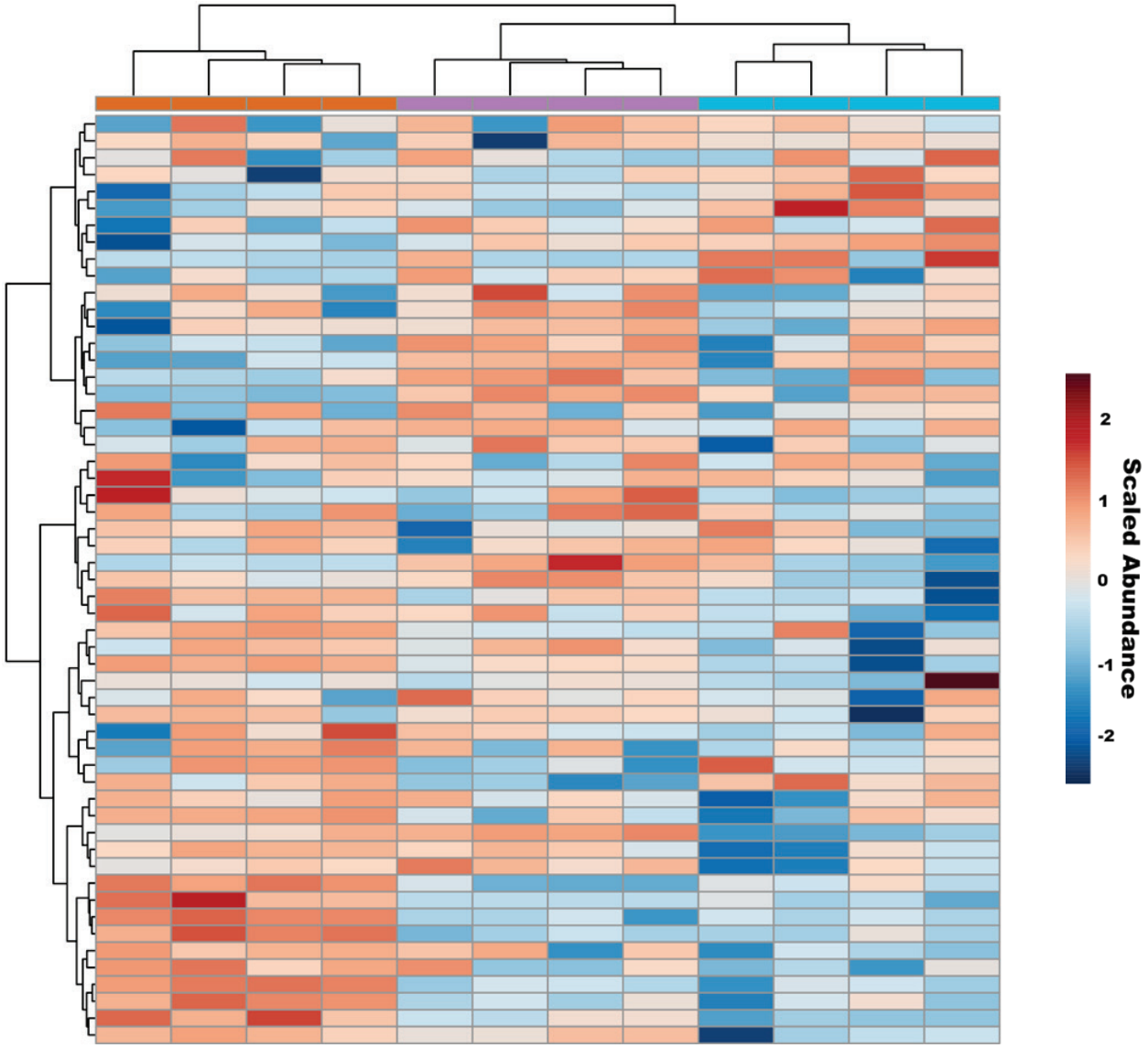


Figure S6: Hierarchical clustering heatmap of all iron-sulfur binding proteins. Dendrograms on each axis employ Euclidian distance and a ward agglomeration. Y axis dendrogram clusters by protein, while the x axis clusters by samples. Each block displays the relative (to the overall average) abundance in each sample (red> average, blue < average)

Figure S7

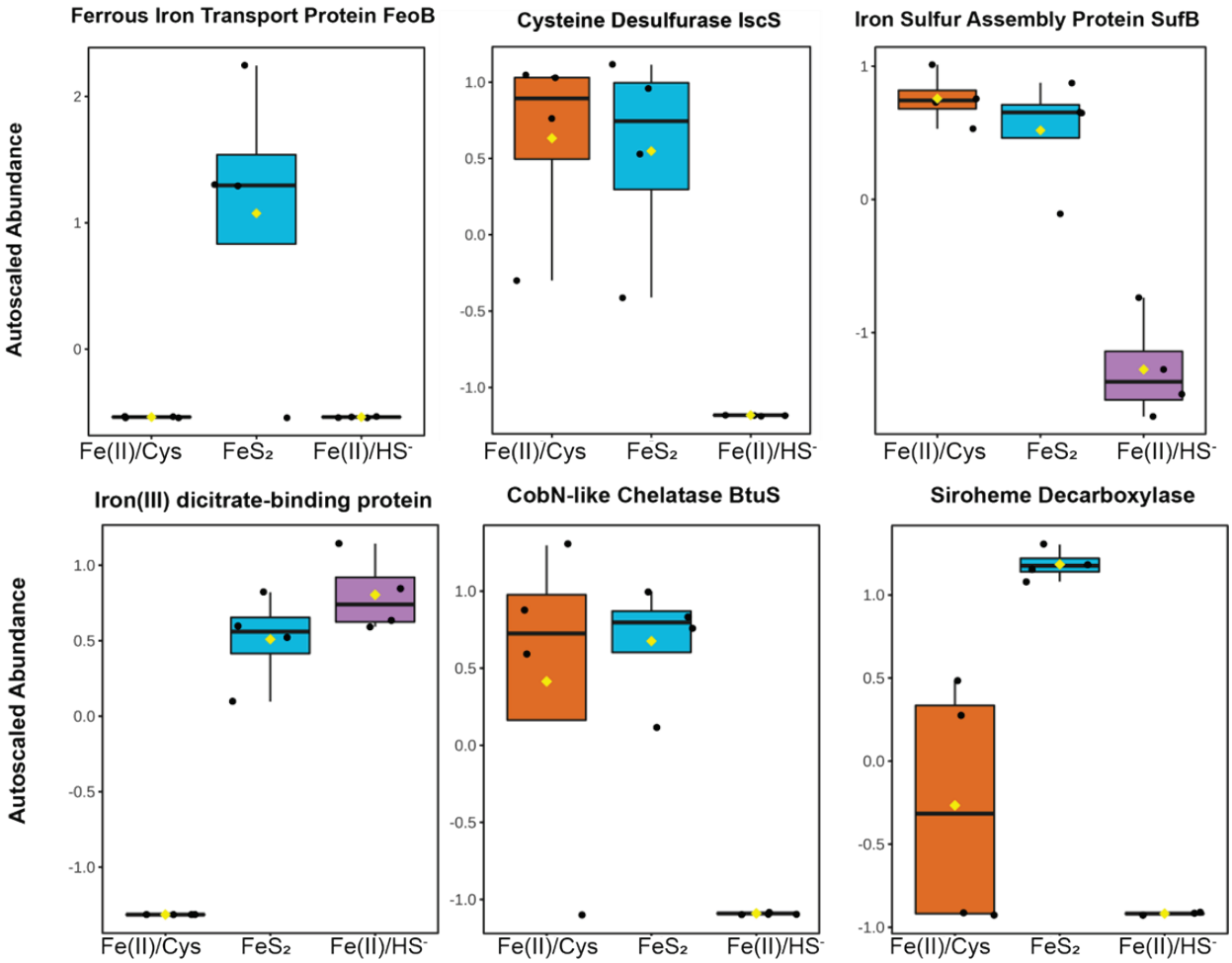


Figure S7: Regulation of key metal binding proteins. Box and whisker plots show auto-scaled abundances of each protein in each growth condition.

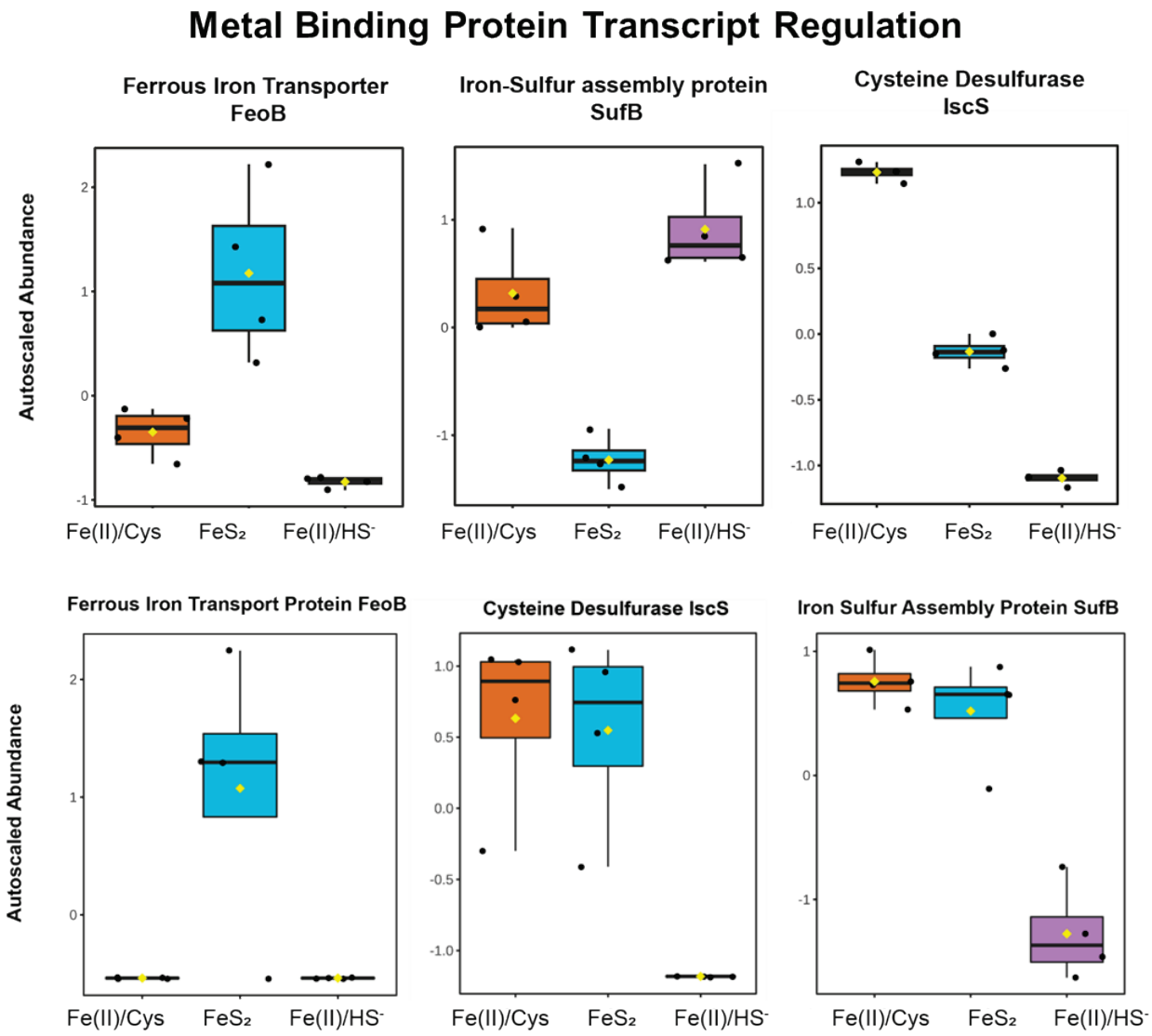


Figure S8: Regulatory patterns of the mRNA transcripts (top row) and proteins (bottom row) of 3 key metal binding proteins. Box and whisker plots show scaled abundance in each condition.

References:

1. Archibald F. 1983. Lactobacillus plantarum, an organism not requiring iron. FEMS Microbiol Lett 19:29–32.

2. Beinert H, Holm RH, Münck E. 1997. Iron-sulfur clusters: nature's modular, multipurpose structures. Science 277:653–659.

3. Johnson DC, Dean DR, Smith AD, Johnson MK. 2005. Structure, function, and formation of biological iron-sulfur clusters. Annu Rev Biochem 74:247–281.

4. Imlay JA. 2006. Iron-sulphur clusters and the problem with oxygen. Mol Microbiol 59:1073–1082.

5. Ayala-Castro C, Saini A, Outten FW. 2008. Fe-S cluster assembly pathways in bacteria. Microbiol Mol Biol Rev 72:110–25, table of contents.

6. Brzóska K, Meczyńska S, Kruszewski M. 2006. Iron-sulfur cluster proteins: electron transfer and beyond. Acta Biochim Pol 53:685–691.

7. Pattus F, Abdallah MA. 2000. Siderophores and Iron-Transport in Microorganisms. Journal of the Chinese Chemical Society 47:1–20.

8. Lau CKY, Krewulak KD, Vogel HJ. 2016. Bacterial ferrous iron transport: the Feo system. FEMS Microbiol Rev 40:273–298.

9. Wu B, Liu F, Fang W, Yang T, Chen G-H, He Z, Wang S. 2021. Microbial sulfur metabolism and environmental implications. Science of The Total Environment 778:146085.

10. Liu Y, Sieprawska-Lupa M, Whitman WB, White RH. 2010. Cysteine is not the sulfur source for iron-sulfur cluster and methionine biosynthesis in the methanogenic archaeon Methanococcus maripaludis. J Biol Chem 285:31923–31929.

11. Major TA, Burd H, Whitman WB. 2004. Abundance of 4Fe–4S motifs in the genomes of methanogens and other prokaryotes. FEMS Microbiol Lett 239:117–123.

12. Christina J, Alexis E, Mason M-E, R. CD, L. DJ, S. BE, W. MW. 2022. Pathways of Iron and Sulfur Acquisition, Cofactor Assembly, Destination, and Storage in Diverse Archaeal Methanogens and Alkanotrophs. J Bacteriol 203:e00117-21.

13. Spietz RL, Payne D, Szilagyi R, Boyd ES. 2022. Reductive biomining of pyrite by methanogens. Trends Microbiol https://doi.org/https://doi.org/10.1016/j.tim.2022.05.005.

14. Truche L, Berger G, Destrigneville C, Guillaume D, Giffaut E. 2010. Kinetics of pyrite to pyrrhotite reduction by hydrogen in calcite buffered solutions between 90 and 180 C: Implications for nuclear waste disposal. Geochim Cosmochim Acta 74:2894–2914.

15. Boyd E, Payne D, Shepard E, Spietz R, Guo Y, Broderick W, Broderick J. 2022. Biomining metals from pyritic ores: physiological, biochemical, and biophysical considerations. Biophys J 121:296a.

16. Devon P, M. SE, L. SR, Katherine S, Sue B, Mark Y, Brian B, E. BW, B. BJ, S. BE, Y. GM. 2022. Examining Pathways of Iron and Sulfur Acquisition, Trafficking, Deployment, and Storage in Mineral-Grown Methanogen Cells. J Bacteriol 203:e00146-21.

17. Spietz RL, Payne D, Kulkarni G, Metcalf WW, Roden EE, Boyd ES. 2022. Investigating Abiotic and Biotic Mechanisms of Pyrite Reduction. Front Microbiol. Lausanne : https://doi.org/10.3389/fmicb.2022.878387.

18. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3567.

19. Pang Z, Zhou G, Ewald J, Chang L, Hacariz O, Basu N, Xia J. 2022. Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. Nat Protoc https://doi.org/10.1038/s41596-022-00710-w.

20. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques P-É, Li S, Xia J. 2021. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49:W388–W396.

21. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. 2022. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res gkac194.

22. Chaudhuri S, Vyas K, Kapasi P, Komar AA, Dinman JD, Barik S, Mazumder B. 2007. Human ribosomal protein L13a is dispensable for canonical ribosome function but indispensable for efficient rRNA methylation. RNA 13:2224–2237.

23. Qi Y, Li X, Chang C, Xu F, He Q, Zhao Y, Wu L. 2017. Ribosomal protein L23 negatively regulates cellular apoptosis via the RPL23/Miz-1/c-Myc circuit in higher-risk myelodysplastic syndrome. Sci Rep 7:2323.

24. Hidese R, Inoue T, Imanaka T, Fujiwara S. 2014. Cysteine desulphurase plays an important role in environmental adaptation of the hyperthermophilic archaeon Thermococcus kodakarensis. Mol Microbiol 93:331–345.

25. Mihara H, Esaki N. 2002. Bacterial cysteine desulfurases: their function and mechanisms. Appl Microbiol Biotechnol 60:12–23.

26. Welz D, Braun V. 1998. Ferric citrate transport of Escherichia coli: functional regions of the FecR transmembrane regulatory protein. J Bacteriol 180:2387–2394.

27. Rocha ER, Bergonia HA, Gerdes S, Jeffrey Smith C. 2019. Bacteroides fragilis requires the ferrous-iron transporter FeoAB and the CobN-like proteins BtuS1 and BtuS2 for assimilation of iron released from heme. Microbiologyopen 8:e00669.

28. M. B, M. G, I. F, G. P, G. C, M. R, G. P. 2011. Structural and Functional Insights into Aeropyrum pernix OppA, a Member of a Novel Archaeal OppA Subfamily . J Bacteriol 193:620–630.

29. Payne D, Spietz RL, Boyd ES. 2021. Reductive dissolution of pyrite by methanogenic archaea. ISME J 15:3498–3507.

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

31. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. 2020. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res 49:D545–D551.

32. Steward KF, Payne D, Kincannon W, Johnson C, Lensing M, Fausset H, Németh B, Shepard EM, Broderick WE, Broderick JB, Dubois J, Bothner B. 2022. Proteomic Analysis of Methanococcus voltae Grown in the Presence of Mineral and Nonmineral Sources of Iron and Sulfur. Microbiol Spectr 10:e01893-22.

33. Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57.

34. Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. Protein Science 28:1947–1951.

METABOLIC AND MEMBRANE ASSOCIATED CHANGES IN METHANOSARCINA BARKERI FACILITATE GROWTH ON PYRITE AS A SOLE SOURCE OF IRON AND SULFUR

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Hunter Fausset

Contributions: Main author. Experimental design, sample prep, data acquisition, statistical analysis, interpretation, writing, submission, editing.

Co-Author: Brett Sather

Contributions: Task delegation, sample preparation, data processing, statistical analysis, writing, editing

Co-Author: Savannah Cox

Contributions: Bioinformatic data processing writing, editing

Co-Author: Gwendolyn Cooper

Contributions: Bioinformatic data processing writing, editing

Co-Author: Scott Spurzem

Contributions: Sample preparation, bioinformatic data processing writing,

Co-Author: Monika Tokmina-Lukaszewska

Contributions: Data interpretation. writing, editing

Co-Author: Rachel L. Spietz

Contributions: Cell culturing and harvesting, intellectual assistance, writing, editing

Co-Author: Eric S. Boyd

Contributions: Grant writing, cell culturing, data interpretation, writing, editing

Co-Author: Brian Bothner

Contributions: Corresponding author, grant writing, data interpretation, writing, editing

Manuscript Information

Hunter Fausset, Brett Sather, Savannah Cox, Gwendolyn Cooper, Scott Spurzem, Monika Tokmina-Lukaszewska, Rachel L. Spietz, Eric S. Boyd, Brian Bothner

Microbiology Spectrum

Status of Manuscript:
__**x**__ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Submission in Q2 2023

Abstract

From microbes to mammals, access to iron (Fe) and sulfur (S) dictates a wide range of cellular physiology. For example, Fe and S have critical roles in protein structure, enzyme catalysis and biosynthetic pathways. To meet this demand, cells have evolved sophisticated assimilatory and processing mechanisms so they may exploit a variety of Fe and S sources. Pyrite (FeS_2) and other iron sulfide minerals, have generally been considered to be Fe/S sinks with respect to biology. Here, we characterize the metabolic and membrane lipid compositional changes the archaeon, *Methanosarcina barkei,* undergoes when grown on solid phase FeS_2. This analysis, supported by a parallel proteomics investigation, shows widespread metabolic and membrane compositional changes in response to Fe and S source. Sulfur equilibria, including levels of L-cysteine and its precursors, are heavily perturbed depending on growth condition. Our analysis indicates that the organism may be actively shuttling S away from cysteine to H_2S, as observed in the accumulation of its precursors and methionine. There are also changes in metabolites involved in gluconeogenesis, reductive TCA cycle, and amino acid biosynthesis. The membrane lipids of *M.barkeri* change according to the presence of iron sulfide minerals, with marked downregulation of glycerophospholipids based on head group, and mixed regulation of archaeol and related phytanyl compounds. This study provides insight into the nuances allowing *M.barkeri* to bio-mine Fe and S from mineral surfaces and provides the foundational information for future studies.

Introduction

Fe and sulfur S are cornerstone elements in the dynamic chemical equilibria of biological systems. Molecules containing Fe and/or S are crucial in every facet of living systems- from DNA replication and transcriptional regulation to the complex electron transfer reactions that constitute the core of energetic metabolism(1, 2). Due to their dynamic oxidation state, Iron-sulfur clusters ([Fe-S]) are particularly important in an organism's ability to generate usable chemical energy from its environment(3, 4). The evolution of life, therefore, has often revolved around the acquisition, processing, and regulation of these critical elements.

Iron is often growth-limiting, so organisms have evolved sophisticated uptake mechanisms(1). Transmembrane transporters, such as the Feo system, have evolved to efficiently shuttle ferrous (Fe(II)) iron across the membranes of bacteria and archaea(5). More difficult to transport forms of Fe, such as Fe(III), require a siderophore transport system to be assimilated intracellularly(6). The form of Fe largely depends on the extracellular environment- Fe(III) predominates in the presence of oxygen and Fe(II) in its absence(5). The form and availability of S is also dependent on the environment. Inorganic sulfur, in the form of sulfate and hydrogen sulfide, alongside organic forms like cysteine (Cys) constitute the majority of biologically available sources(7–9). Across most taxa, Cys typically represents the key metabolic intermediate between exogenous S and the other biologically relevant S containing compounds (biotin, methionine, [Fe-S] clusters, etc.).

Methanogenic archaea are key constituents of the biogeochemical circulation of Fe and S(8, 10). These organisms require more iron and sulfur per cell than most other forms of life(11, 12). Both diverse and ancient, methanogens inhabit a wide range of ecological niches where they

generate methane as a metabolic biproduct(10, 13, 14). Depending on lineage, they utilize hydrogen and carbon dioxide, formate, methanol, methylamines, and/or acetate as their ultimate carbon source(15–17). In laboratory settings, methanogens have been traditionally grown in the presence of soluble Fe(II) and HS^- or cysteine as their Fe and S source.(11, 17–22) While representative of some environments, concentrated Fe(II) and soluble sulfides readily react forming insoluble iron sulfides such as pyrite (FeS_2)(23–26). The mineralization of sulfides and Fe(II) has been occurring for millions of years in marine environments, locking valuable Fe and S away in what was previously assumed to be biologically inaccessible pools (25–27). The methanogens already inhabiting anoxic marine sediments where this process was occurring were thus presented with a large evolutionary incentive to evolve a mechanism to extract Fe and S from iron sulfide minerals.

Recently, it has been discovered that certain methanogenic archaea can live on FeS_2 as their only source of Fe and S(18, 19, 21). Two *Methanococcus* species, *maripaludis* and *voltae*, alongside their cousin *Methanosarcina barkeri* can actively reduce FeS_2 minerals and assimilate the soluble products. The mechanism involves active electron pumping into the mineral surface, whereby pyrrhotite ($Fe_{1-x}S$) forms following the release of S as hydrogen sulfide (HS^-). Fe then solubilizes from the mineral surface and reacts with HS^- to form aqueous, Fe-S clusters of unknown size ($nFeS_{aq}$)(28). $nFeS_{aq}$ is the ultimate form of Fe and S for cells only given access to mineral FeS_2. While the extracellular bio-geochemical processes involved have been suggested, the electron transfer mechanism at the mineral surface and import of FeS remains unclear. Likewise, downstream biochemical manipulation of $nFeS_{aq}$ to usable forms (Fe-S clusters, vitamins, enzyme cofactors, etc.) has not been elucidated. The biological changes surrounding

the cellular-mineral interface are becoming clearer, but much remains unknown about the metabolic compensations that allow for growth on the mineral surface.

Previous proteomics and transcriptomics analysis indicated that these methanogens undergo significant phenotypic changes in response to the source of Fe and S. However, the biology driving many of the changes at the level of transcription and translation remain unexplained. To better understand the cellular adjustments for growth on pyrite on a global level, we performed untargeted metabolomic and lipidomics of *M.barkeri* on different Fe and S sources. Metabolomics was employed to provide insight into S metabolism as well as general respiration of the system. Membrane compositional lipidomics was used to explain previously observed changes in cellular volume (~30% decrease in mineral Fe/S conditions) and the presence of extracellular appendages (image reference) as they relate to membrane lipids.

Results

Global Untargeted Metabolomics and Lipidomics

To better describe the metabolic and physical changes in *M.* barkeri that depend on Fe and S, we performed LCMS-based untargeted metabolomic and lipidomics in parallel (on the same cell pellets) as our previous shotgun proteomics analysis. Cells were grown on mineral FeS_2 or soluble Fe(II) supplemented with Cys or HS^- for S requirements, generating three different Fe/S environments for comparison (Fe(II)/Cys, Fe(II)/HS^-, and FeS_2) . *Methanosarcina barkeri* strain MS (ATCC-43569) were grown following previously described methods (18). Harvested biomass was used for proteomics, metabolomics, and membrane lipidomics in parallel. All groups shared growth conditions including carbon source, temperature, pH etc, only differing in their source for Fe and S. In brief, cell culture media differences were as follows:

Fe(II)/Cys; 2 mM cysteine and 20 μM $FeCl_2$, Fe(II)/HS^-; 2 mM Na_2S and 20 μM $FeCl_2$, and FeS_2; 2 mM synthetic nanoparticulate FeS_2 prepared as previously described (19–21) Filtered cellular biomass underwent a multi-omics extraction protocol yielding appropriate samples for proteomics, metabolomics, and membrane composition lipidomics. Metabolomics samples were analyzed using HILIC chromatography in positive mode LC-MS parameters, and lipidomics utilized RP C18 chromatography, a longer gradient, and negative mode.

After LCMS acquisition and processing, our untargeted metabolomics analysis yielded 10,438 features. Using missing value estimation/removal and interquartile range filtering (IQR), the analysis was carried forward using the 2500 highest quality (least likely to be noise or irrelevant signals) features. Approximately 40% of these features (981) change in abundance significantly in at least one conditional comparison (p-value<0.05) (Figure 1). A principal component analysis (PCA) revealed little overlap between the metabolomes of each group- with Fe(II)/HS^- clustering away from the other two. The metabolism of FeS_2 exposed cells looks drastically different than that of Fe(II)/HS^- (Figure 1). A parallel hierarchical clustering heatmap (PCHA) of the top 200 features mirrors the observations in the PCA, with the groups clustering perfectly. When filtered to the most important (based on p value) features, FeS_2 clusters away from the other two groups. A large band of downregulated features is observed in the FeS_2 condition relative to the two others, with smaller bands of upregulated features near the bottom of the heatmap (Figure 1). The number of dysregulated features, separation of groups in the PCA, and distinct banding patterns in the PHCA suggest that small molecule metabolism is at least moderately implicated in *M.barkeri*'s ability to grow on mineral FeS_2. To expand on these findings, a series of pairwise comparisons were performed to examine which metabolic changes

are associated with specific environmental conditions as well as rule out any obvious statistical bias. For this analysis, since FeS_2 is the phenotype of interest, the Fe(II)/Cys vs FeS_2 and Fe(II)/HS^- vs FeS_2 comparisons were prioritized. As before, the analysis was limited to the 2500 most relevant features based on IQR.

The Fe(II)/Cys vs FeS_2 comparison was designed to highlight the metabolic impact of ; i) soluble, cationic Fe(II) compared to mineral-sourced Fe, ii) additional sources of organic carbon via cys supplementation, and iii) acquisition of S through amino acids vs mineral uptake. This comparison yielded a 34% dysregulation rate (855/2500 features @ p-value < 0.1). Approximately half (460) were higher in abundance in FeS_2 relative to Fe(II)/Cys. PCA shows clear separation between the groups (Figure S1). The PHCA shows a similar clustering pattern to the all-group comparison, in this case perfectly clustering the groups when all features are included (Figure S1). The PCHA of the top 40 features displays features with remarkable differences in their abundance profiles based on their Fe and S environment.

In contrast to the previous comparison, there is no difference in the source of Fe and S when comparing Fe(II)/HS^- to FeS_2 grown cells. The major difference between these conditions lies in the abiotic mechanism of nFeS cluster formation in Fe(II)/HS^-; cells grown on mineral FeS_2 must perform this step using biological processes. There is a greater degree of abundance differences in this comparison, with 42% of the features changing in response to the Fe and S environmental shift (1057/2500) (Figure S2) . This is clearly echoed in the PCA and PHCA plots. Interestingly, the replicates of the Fe(II)/HS^- group together much closer than those of the FeS_2 group (Figure S2) . This is usually indicative of a stress-phenotype, or a situation where cells are otherwise receiving strong biochemical signals.

Identifications and Metabolic Interpretation

To assign biological significance to observed statistical variance, we utilized established databases (KEGG and Biocyc) as a reference for potential identifications using *in silico* fragmentation. Structure files are pulled from each database and subjected to theoretical fragmentation based on the Progenesis fragmentation algorithm (nonlinear source). The resulting fragmentation patterns, paired with mass error and isotopic distributions, are used to assign identifications to experimental features. Identified features are grouped in tiers based on the quality of their fragmentation pattern match.

Sulfur equilibria are important for biological function and understanding the biomining process. We identified key cysteine biosynthesis intermediates including L-cysteine, O-acetylserine, and L-serine (Figure 2). As expected, the feature identified as cysteine accumulated in the Fe(II)/Cys experimental group. This feature is completely absent, by contrast, in both FeS_2 and Fe(II)/HS^-. Cysteine represents an intersection of critical sulfur metabolism, with most other sulfur containing molecules (Coenzyme A, Coenzyme M, [Fe-S] clusters, molypbterin cofactor, and methionine) using cysteine as a precursor at some point in their synthesis. The two upstream metabolites of cysteine synthesis, O-acetylserine and L-serine, were also found. Like cysteine, these molecules accumulated in the Fe(II)/Cys condition and were found in low abundance in the other two groups (Figure 2). Serine and O-acetylserine differ from cysteine in that they are present in measurable amounts (especially for O-acetylserine) in FeS_2 and Fe(II)/HS^-. Therefore, the precursors to cysteine exist as free molecules, but the equilibrium appears to shift away from cysteine considering it is completely absent in Fe(II)HS^- and FeS_2.

Our previous proteomics investigation, performed on the same cells, identified some of the proteins crucial in sulfur equilibrium. Cysteine synthase, the key enzyme in sulfur

assimilation and anabolic processes, shows a peculiar regulatory pattern. It shows distinct downregulation in the Fe(II)/HS^- group with similar abundances in FeS_2 and Fe(II)/Cys (Figure 2). A cysteine desulfurase enzyme shows a similar regulatory pattern. The cysteine—tRNA ligase does not change in abundance based on its one-way ANOVA pvalue (Figure 2). The abundance patterns of these proteins do not explain the absence of cysteine in FeS_2, and can only loosely describe its absence in the Fe(II)/HS^- conditions. The abundance difference pattern of cysteine synthase is surprising considering the level of accumulation of its product in the Fe(II)/Cys condition. This enzyme is annotated as having a catalytic bias in the direction of transferring S from H_2S to cysteine. Our data suggest the pathway is running in both directions, which would shuttle organic acid intermediates towards other metabolic needs while liberating the S in cysteine. This is supported by the similarity of regulation pattern (accumulation in Cys, much lower abundance in Fe(II)/HS^- and FeS_2) for L-serine, O-acetylserine, and L-cysteine. The synthesis of cysteine, the cysteine-tRNA, and homocysteine are the only presently annotated entry points for H_2S, and this data could indicate the importance of shuttling S away from the amino acid. (biocyc, probably another citiation).

The relative abundances of Methionine are nearly identical to those of cysteine. Free methionine was not detected in FeS_2 or Fe(II)/HS grown cells, however, it accumulated in Fe(II)/Cys (Figure 3). It is possible that methionine is accumulating due to the buildup of sulfur (As H_2S) from the reversed cysteine synthase reaction, where the newfound H_2S reacts with O-Actyl-L-Homoserine or O-succinyl-L-homoserine to form homocysteine, which can then convert directly to methionine. The lack of methionine, in addition to cysteine, in FeS_2 and Fe(II)/HS^- conditions may suggest a sulfur limitation- where there is little H_2S available to form the

necessary precursors. The proteomics investigation tracked the abundance of two proteins that would affect methionine levels, homocysteine methyltransferase and methionine tRNA ligase (Figure 3).These enzymes catalyze the synthesis of methionine from homocysteine and the incorporation of Met into its tRNA, respectively. These two enzymes were not found to change in abundance in any of the growth conditions.

Another crucial sulfur-containing metabolite investigated was sulfoethylcysteiene. It is peculiar in that it contains two sulfur atoms, most other sulfur metabolites have just one. It is also the only identified sulfur metabolite that is in lower abundance in Fe(II)/Cys relative to the other experimental conditions (Figure 3). It is essentially absent in this group and has moderate-high abundance in FeS_2 and Fe(II)/HS^-, essentially the inverse of Cys and Met levels. This molecule is the direct precursor to coenzyme M(CoM), the carbon donor for methane production in methanogenesis. No other metabolites in the CoM biosynthetic pathway were identified in this investigation, but two relevant proteins, cysteate synthase and sulfopyruvate decarboxylase, show regulation patterns similar to that of sulfoethylcysteine **(Figure 3)**. Cysteate synthase catalzyes the conversion of O-phospho-L-serine to L-cysteate and is in high abundance in Fe(II)/HS^- and FeS_2 relative to Fe(II)/Cys. Sulfopyruvate decarboxylase catalyzes the conversion of 3-sulfopyruvate (one additional step downstream from cysteate) to sulfoacetaldehyde, and mirrors the abundance pattern of cysteate synthase. Two further enzymatic reactions generate sulfoethylcysteine which is the direct precursor to coenzyme M. The enzymes that carry out these latter steps remain unknown in *M.barkeri* (based on biocyc database)- but isotopic labeling studies done in 1985 have shown this is likely the pathway of de-novo CoM synthesis in methanobaceterium (different class).

Our analysis also identified molecules with mixed abundance patterns involved in general metabolism. Acetate is a carbon source available to each condition, and shows slight accumulation in the FeS_2 group and is in lowest abundance in the Fe(II)/Cys condition (Figure S4). Pyruvate, a main metabolic intermediate, shares this regulatory pattern. Malate is downstream from pyruvate in the incomplete, reductive TCA cycle utilized by methanogens to synthesize anabolic intermediates (source). Malate is regulated differently than its precursor pyruvate, showing highest abundance in Fe(II)/HS^- and is in comparable abundance in FeS_2 and Fe(II)/Cys. 2-oxoglutarate, a molecule used in the synthesis of numerous amino acids (Glutamate, glutamine, arginine, ornithine, isoleucine, leucine, serine, valine, aspartate, histitine, phenylalanine, and tyrosine) as well as CoM, shares the abundance pattern observed in malate (Figure S4). Oxoaloacetate, like pyruvate, is used in methanogens as an intermediate in the incomplete TCA cycle to generate biosynthetic intermediates. This molecule is in high abundance in the Fe(II)/Cys group and is similarly downregulated in FeS_2 and Fe(II)/HS^-. 2-oxobutanoate, an isoleucine biosynthesis intermediate, is upregulated in FeS_2 and has a lower abundance profile in the other two groups. The mixed regulatory pattern of these molecules and others, which are often adjacent in metabolic pathways, complicate analysis and biological interpretation. This is further complicated by the fact that many related proteins show either dissimilar regulation or remain unchanged in the environmental conditions tested in this analysis. For these reasons, a further analysis combining the datasets in a concerted statistical analysis is underway and will present a more comprehensive metabolic model in the future. Despite the nuance, it is clear that central metabolic pathways such as amino acid biosynthesis, gluconeogenesis, the incomplete reductive TCA cycle and cofactor biosynthesis, as well as sulfur

metabolism discussed above, are involved in the biological differentiation *M.barkeri* undergo to grow on iron-sulfide minerals.

Membrane composition and Lipidomics

To assess changes in membrane composition associated with the Fe and S environment of *M.barkeri*, we performed an untargeted lipidomics analysis in parallel with the metabolomics described in the previous section. The insoluble lysate containing nucleic acids, membranes, cytoskeletal elements, and other insoluble molecules was extracted to enrich for membrane lipids and other hydrophobic compounds associated with the cellular membrane. Our investigation yielded nearly 12,000 MS features in the raw data over the course of the 28 minute LC method. Missing value estimation/removal and IQR filtering were again used to filter out features from the noise, leaving 2500 for statistical analysis and interpretation. The ANOVA analysis showed that ~36% (895/2500) change significantly between experimental groups, most of which were less abundent in FeS_2 relative to the other two groups (Figure 4). The PCA shows a distinct but subtle statistical difference between the groups – it is noticeably more overlapping than the metabolomics PCA and much more condensed than that found in our proteomics investigation (Figure 4). Notably, FeS_2 groups away from the other two, indicating possible correlation with the presence of bulk mineral and membrane composition. The PHCA heatmap of the top 200 features (based on p-value) expands on this, showing that the bulk of the features are in lower abundance in FeS_2 (Figure 4). There is, however, a moderate number of features in much higher abundance in the FeS_2 condition as seen in the dark red cluster at the top of the heatmap. To more thoroughly characterize the statistical differences between growth conditions and investigate potential statistical bias relating to the low abundance of many features in the FeS_2

condition, we separated the analysis into two-group comparisons. Due to the nature of the project and to simplify these results, the Fe(II)/Cys vs Fe(II)/HS^- comparison was not included and will be analyzed in a future publication.

When compared directly, the lipidomes of the membranes from cell cultures grown on mineral FeS_2 look moderately different than the Fe(II)/HS^- condition. The volcano plot shows that 337 features are higher in abundance and 343 are lower in the FeS_2 condition relative to Fe(II)/HS^-(FC>2, p-value<0.1)(Figure S5). This changes to 207 and 237 at a p-value threshold of 0.05. Most importantly, the volcano plot shows an exceptionally equal distribution of up and downregulation, with 1243 up and 1254 downregulated with all filters removed. The majority of the features cluster close to the center, which is indicative of a rigorous, unbiased statistical approach; most lipids and nonpolar components of the membrane should not be changing drastically in a proper experimental comparison. This eliminates any concerns of statistical bias that may have come up during the ANOVA analysis, where most of the significant features are low in abundance in FeS_2. The PCA plot, all feature PCHA and top 200 feature PCHA similarly mirror the results from the all-group comparison. Additionally, some of the features that are changing in this comparison are likely directly related to the bio-mediated reductive dissolution of the mineral surface due to the specific environments being compared.

As above, the FeS_2 and Fe(II)/Cys conditions were compared to evaluate any statistical artifacts and specific MS feature trends associated with these two groups. This volcano plot analysis yields 439 features in higher abundance in FeS_2 and 347 in lower abundance relative to the Fe(II)/Cys group (FC>2, p-value<0.1) (Figure S6). These numbers contract to 333 and 238 at the traditional statistical significance threshold (p-value<0.05). With filters removed, the

numbers are fairly equal with 1197 up and 1300 downregulated, with most clustering towards the center of the volcano plot. Again this lends credence to the analysis and subsequent inferences. The PCA and PCHA heatmaps show distinct separation between the groups, especially when narrowed down to the top 200 features (filtered by p-value).

Archaeol lipids, true to their name, are unique to the kingdom archaea and possess unique chemistry surrounding their acyl-linkages. They are attached to glycerol as phytanyl chains which utilize an ether linkage as opposed to the traditional ester bonds. Using accurate mass and isotopic distribution, as well as fragmentation patterns in some features, we assigned 7 features as being some type of archaeol (Figure 5). Other than Benzodithiopheno-quinone (BDTQ) 1Gly-hydroxyl—archaeol, and a cyclized crenarchaeol, the lipids that change significantly have the opposite pattern as observed in most annotated lipids from the previous section (Figure 5). These lipids are in high abundance in FeS_2 grown cells and are downregulated in the Fe(II)/Cys and Fe(II)/HS^- groups. Interestingly, BDTQ is the only annotated archeaol that also contains S, and is in very low abundance (near the LOD) in FeS_2. The only lipid of this class found not changing with statistical significance has a PG headgroup (Figure 5).

Two proteins involved in the synthetic process of archaeols were characterized in the proteomics investigation- geranylgeranylglyceryl phosphate synthase (GGPS) and digeranylgeranylglycerophospholipid reducatse (DGPR) (Figure S7). The former catalyzes the conversion of gerangeranyl diphosphate and glycerol-1-phosphate to 3(O-gernanylgeranyl-sn-glycerol 1-phosphate (GGGP), the precursor to both saturated and unsaturated archaeol lipids. The latter enzyme catalyzes a downstream reaction converting 3(O-gernanylgeranyl-sn-glycerol 1-phosphate to 2,3-bis-O-phytanyl-sn-glycerol 1-phosphate, from which only saturated archaeols

can be synthesized. The abundance patterns in each protein are dissimilar; GGPS is in very low abundance in Fe(II)/Cys and FeS_2 and upregulated in Fe(II)/HS^- while DGPR is abundant in each group, with its abundance highest in Fe(II)/Cys followed by Fe(II)/HS– and lowest in FeS_2 (Figure S7). Changes in GGPS indicate overall differences in overall archaeol abundance between the groups, particularly in Fe(II)/HS^-, relative to the other groups. This is particularly interesting when taking the lipid data into consideration – none of the annotated archaeol lipids are highest in abundance in the Fe(II)/HS^- condition. This could represent a concerted upregulation of the enzyme in order to synthesize more archaeols for cells in the Fe(II)/HS^- condition. It is also peculiar how low in abundance this enzyme is for the other groups – it averages around 2e07 intensity for Fe(II)/HS^-, but is completely absent in FeS_2 and Fe(II)/Cys and was replaced by an approximated LOD measurement in the missing value estimation step. This could be due to multiple causes, such as another gene product that carries out this reaction, thereby recovering the necessary activity for these groups, that was not annotated in our proteomics investigation. This will be explored in the coming complete omics-model/ bioinformatic investigation of the system.

In addition to the archeaol- specific database search performed above, we also searched against LipidMAPS and metacyc to identify other lipids and nonpolar molecules in the membrane of *M.barkeri.* A major class that is observed to be altered in response to changes in Fe/S source are glycerophospholipids (GPLs). Ether linked forms of these lipids are major constituents of the cellular membrane. In these lipids, acyl chains attached to a glycerol-phosphate, that may or may not have an additional headgroup such as a choline (PC) serine (PS) additional glycerol (PG), inositol (PI), or other functionalization. Our analysis annotated three

classes of GP that appear to be involved in the biological processes at hand; PA, PG and PI (Figure 6). Note: the annotations for these lipids contain ester linkages, whose synthesis are not characterized to exist in *M.barkeri*. Either this is evidence for the presence of ester linked lipids, which has been reported previously, or they are misannotated ether linked lipids that are not contained in the databases used in our search. In any case, the listed head groups have distinct fragmentation patterns, so the remainder of this analysis will focus on the regulatory patterns based on head group.

Three PC lipids were assigned identifications but none exhibit measurable change between groups. An interesting pattern emerges with GPL abundance relative to the FeS_2 group. For the most part, the GPs characterized in this investigation are in low abundance in FeS_2, with levels in Fe(II)/Cys and Fe(II)/HS^- being similar (Figure 6). This is true for 10 of the 12 annotated PI, all 10 of the 12 annotated PG, and x of the 31 PA features. 15 of the 31 PA were statistically significant, 5 PI were significant, and only 2 PG met a pvalue threshold of 0.05. An expanded pvalue threshold of 0.1, justified by the high variance and subsequent inclusion of biologically important molecules (a PI(13:0/18:2) has a pvalue of 0.058, for example), increases the number of significant features annotated as PA to 18, PI to 7, and PG to 5. A table of all annotated lipids in this categorization is seen in Table 1

As stated, most GPLs are in lower abundance in cells grown on FeS_2, with an intriguing exception – those with an odd chain acyl group. The pattern is noticeably different for PI(19:1/22:2) and PG(13:0/18:2), but is maintained in PA(17:2/22:6), suggesting the observed odd-chain acyl pattern is headgroup dependent (Figure 6). Little is understood about the non-archaeol biosynthesis of membrane lipids in this archaeon, but recent studies have shown an

abundance of non-archaeol lipids changing in response to environmental stressors. The fourth sub-class of GPLs tracked in this investigation, PC, remains unperturbed between experimental groups. This can be seen in the lack of abundance differences of PC(10:0/24:0), and two PC(15:0/18:2) species (Figure S8). These represent both even and odd chain acyl group GPLs and lend further evidence towards a headgroup specific regulation event.

Due to the distinct isotopic distribution of S in organic molecules, our structural approximation/identification pipeline accurately annotates the presence of S within detected features. This, paired with consistent fragmentation patterns of a cysteinyl-moiety, allows us to annotate lipids as containing cys or cys containing peptides (Figure 8). These lipids are annotated as leukotrienes from our search, which are mammalian inflammatory signaling molecules and precursors. Their structure closely mimics arachidonic acid and many of the well characterized archaeal carotenoids, which are well characterized in archaea. With this taken into consideration, we are likely characterizing cysteine-moiety containing carotenoids or other unsaturated fatty acid derivatives.

These lipids share the regulatory pattern of many of the other lipids discussed above with a similar abundance in Fe(II)/Cys and Fe(II)/HS^- and are low in FeS_2 (Figure 8). The feature annotated as leukotriene E4 is entirely absent in FeS_2 and in high abundance in the other environmental conditions. LTF4 and LTD4 are present in each group, but are very similarly downregulated in FeS_2 grown cells (Figure 7). 14,15-HXA3-D(11S) is another feature sharing the eicosanoid acid structural determination that contains a S as a cysteinyl peptide functional group with a slightly altered abundance pattern that exhibits high variability in the Fe(II)/HS^- group (Figure 8).

Other lipids containing non-cysteinyl S were also annotated using this approach (Figure S9). Of particular interest is a 13-chloro-docosane-1,14-disulfate, which shares S chemistry with previously discussed molecules involved in the synthesis of CoM. The chlorine group here is annotated as part of the structure, but is likely an artifact of ESI ionization and not part of the original molecule. Its relative abundance pattern goes against that of most other lipids, and is in lowest abundance in Fe(II)/Cys, similar to the patterns observed in sulfoethylcysteine (Figure S8**)** . Molecules containing a sulfate group are of interest because they require sulfur-oxygen substrates (SO_3 is annotated as the substrate in the synthesis of L-cysteate). The equilibria surrounding this type of S chemistry remains largely unannotated, and this evidence suggests that a major difference between these experimental groups lies therein. A feature annotated as a sulfur-containing purine (6-(methylthio)purine) also changes in abundance in an interesting pattern. While this is a low-strength structural identification, the isotopic distribution, mass error, and retention time suggest that this is a highly nitrogenated, methyl-sulfur compound. Taken together, this data provides evidence towards the presence of S in the lipid membrane of *M.barkeri*, specifically related to cysteine, sulfoxy functionalization, and nitrogenated S compounds. Further characterization is needed to strengthen these identifications and fully elucidate their structure, but the presence of S containing lipids expands the fields knowledge of the complex S equilibria in this organism.

Discussion:

Our study provides key insights into the metabolism and lipid compositional adaptations allowing for *Methanosarcina barkeri* to subside with minerals as their sole source of Fe and S. Unsupervised, global statistical investigations of the metabolome and lipidome reveal moderate

to severe overall changes in metabolic and membrane profiles when comparing cells grown on minerals to those in soluble Fe/S conditions. Approximately 40% of the filtered metabolomics dataset change in response to Fe/S environment when considering all groups. Pairwise investigations, focusing on FeS_2 compared to the other two groups, support this magnitude of disruption and also alleviate concerns with statistical bias. Slightly less of the MS features from the membrane compositional analysis changed in abundance, 36%, and the multivariate comparative analyses showed a greater degree of overlap. Pairwise analyses analogous to those in the metabolomics investigation corroborate these findings.

To interrogate the biological significance of these statistics, we searched the LC-MS data for identifications using metacyc and KEGG as structural reference databases. Tentative identifications were assigned using mass error, isotope distribution, relative retention time, and fragmentation patterns. Key sulfur metabolism intermediates were identified as changing between conditions. Most notably, cysteine appears completely absent, or otherwise below detectible levels in the FeS_2 and Fe(II)/HS^- groups. This molecule accumulates in the Fe(II)/Cys group, as would be expected. Likewise, molecules upstream in the synthesis of cysteine accumulate in the Fe(II)/Cys group. Using data from our previous proteomics investigation, the regulatory patterns of proteins in this biosynthetic pathway cannot account for the absence of the amino acid in the FeS_2 and Fe(II)/HS groups. Methionine abundance mirrors the patterns observed in cysteine, as do the proteins that might influence its abundance. Sulfoethylcysteine, the other key sulfur containing metabolite and a key intermediate in the synthesis of CoM, is absent in Fe(I)/Cys but abundant in the other two groups. Metabolomics is well suited to track

the changes associated with sulfur equilibria, and it is evident that *M.barkeri* cells look very different with respect to their S metabolism in the conditions tested here.

Several lipids and other nonpolar molecules containing sulfur were identified in our analysis. Numerous cystinyl-carotenoid like compounds appear to be downregulated, or entirely absent, in the FeS_2 condition. Similarly, a methyl-thiopurine and sulfate-group containing lipids were identified, but displaying the opposite regulatory pattern. Our identification matching algorithm does not contain many archaeol lipids, so we curated a database from a previously published review on the subject. We matched several archaeols, based on mass error and isotopic distribution, displaying mixed regulation. There appears to be head group and acyl chain length-based dysregulation, when looking at the identification matches to non-traditional archaeal lipids. Follow up investigations are underway to confirm these identifications and strengthen any subsequent inferences.

Our previous, parallel proteomics investigation corroborates many of the findings observed here. That analysis showed 30% of the identified proteins as changing significantly from the changes in Fe/S environment. The multivariate comparisons (PCA and PHCA) of the proteomics investigation showed a greater difference between groups than those from either the metabolomics or membrane lipidomics analyses presented here. This makes sense, given that many of the proteins observed changing in the previous study were not catalytically active (ribosomal proteins, structural components, S-layer proteins), meaning they only indirectly affect the small molecule and lipid profiles of the organism. Thus, one would expect the relative contraction of PCA plots resulting from the analysis of the small molecules and lipids we observed here. This degree of change is echoed in another previous proteomics investigation on,

Methanococcus voltae, a close relative of *M.barkeri*(30). The findings observed here largely support our previous inference that *M.barkeri* undergo phenotypic change as a result of changes in their immediate Fe/S environment. Other inferences from our collaborators, such as the decrease in overall cellular volume and appearance of appendages in mineral grown cells, also corroborate this finding(18, 21).

Key takeaways from the presented studies surround the observed perturbations in sulfur equilibrium, potential presence of novel sulfur containing molecules, changes in lipids based on head group, and the importance of sulfite (SO_3)/sulfate (SO_4) in these organisms. Due to the regulatory patterns observed in cysteine, O-acetylserine, and L- serine, it is clear that sulfur equilibrium is disrupted in FeS_2 grown cells. The proteins that consume cysteine- such as cysteine desulfurase and the tRNA ligase are not higher in abundance in FeS_2, providing no evidence of increased consumption of the molecule. This provides evidence that the cells in the Fe(II)/HS^- and FeS_2 cells could be sulfur limited, with respect to sulfur in H_2S. Instead, these groups may accumulate sulfur as sulfoethylcysteine, which utilizes SO_3^- as a precursor rather than H_2S. This is supported by its accumulation in these groups, while it is completely absent in Fe(II)/Cys.

Another interpretation of these findings as they relate to sulfur metabolism surrounds cysteine synthase. This enzyme is annotated as going primarily in one direction from O-acetylserine + H_2S → L- Cysteine. Our data shows the accumulation of the upstream metabolites from cysteine, as well as methionine, in the Fe(II)/Cys condition. This is also supported by cysteine synthase's abundance profile; it is notably abundant in the Fe(II)/Cys condition. The lack of downregulation of this enzyme despite product accumulation, paired with the buildup of

O-acetylserine, L-serine, and methionine leads us to believe this may be a bidirectional enzyme used to liberate H_2S from cysteine to be shuttled elsewhere. It would be advantageous for this to be the case, considering methionine can only be synthesized from homocysteine, which in turn gets its S from H_2S.

As discussed earlier, presented here is only the top 2500 features from each analysis. These two datasets are very large and likely contain additional inferences and findings that will be explored in future, multi-omic investigations. There are clear patterns to be found as it is, but the full incorporation of these two datasets with the previously acquired transcriptomics and proteomics data will provide the clearest look into the system. Other limitations surround the method used for identifications, which is currently being strengthened through the use of chemical standards to identify key sulfur intermediates we missed (CoM, CoA) as well as archaeol lipids. Likewise, this analysis honed in on sulfur metabolism -there are also key central carbon metabolites and likely others changing in response to the experimental conditions tested here that will require further identification confirmation. Despite these limitations, the inferences contained here will propel our knowledge and future studies into the system. For instance, we now know that SO_3/SO_4 could be key sulfur carriers for *M.*barkeri cells in mineral environments, sulfur could be contained in carotenoid-like unsaturated fatty acids, and the relative metabolic burden associated with growth on pyrite. These, and a host of other new hypotheses generated from this analysis, can now be more directly interrogated using the information contained herein.

Methods:

Cell cultivation

M. barkeri strain MS was obtained from the American Type Culture Collection (ATCC-43569). Cells were grown following previously described methods (Spietz et al 2022 Frontiers) in anoxic, low salinity medium and without added Fe or S. Cultures for proteomics analysis were grown with three different Fe and S sources: 1) 2 mM cysteine and 20 μM FeCl2, 2) 2 mM Na2S and 20 μM FeCl2, or 3) 2 mM synthetic nanoparticulate FeS2 prepared as previously described (Payne et al 2021, Spietz et al 2022). When cultures reached mid-log phase, as determined by CH4 and DNA production monitored using previously described methods (Spietz et al 2022), cells were harvested under anoxic headspace by centrifugation at 4,600 x g for 30 minutes at 4 °C, then supernatant was decanted in an anaerobic chamber before submerging the cell pellet in liquid N2 to flash freeze. Frozen cell pellets were stored at -80 °C until further processing for proteomics, metabolomics, and lipidomics.

Cell lysis

Cell pellets were resuspended in 5mL 50mM ammonium bicarbonate (NH4HCO3) with protease inhibitor mix (Complete Mini EDTA Free Protease Inhibitor Cocktail, Roche) ph 8. Samples were lysed using an ultrasonic homogenizer at 40% power output for 15 minutes. Resulting lysate was centrifuged at 4,000G for 15 minutes. The supernatant continued on to the metabolomics sample preparation and the insoluble pellet was extracted for lipidomics.

Metabolomics sample preparation

Cold acetone (-20°C) was added (5:1 v/v Acetone) to the soluble lysate and incubated overnight at -80°C. Samples were then centrifuged and supernatant transferred to fresh micro-centrifuge tubes, dried in a vacuum centrifuge at room temperature until dry (approx. 4 hours). Dried metabolite samples were stored at -80°C until analysis.

Lipid extract

Insoluble pellet was resuspended (mixed by pipetting) in 1.3 ml of HPLC grade Methanol and transferred to a 7ml glass culture tube held within a 15 ml falcon tube. 1 ml of HPLC grade MTBE and 1 ml of HPLC grade CHCl3 (1:1, v/v/v) are then added and vortexed. Samples were incubated on a shaker at 225 rpm for 1hour at 22°C followed by a centrifuge step at 17,500g for 10 min. 1.75 ml of the supernatant was extracted and dried in a vacuum centrifuge for 1 h at 30 °C and stored for analysis at -80°C.

LCMS acquisition, data processing

Metabolomics. LCMS acquisition was performed on a Waters Synapt XS paired with a Waters Acquity I class liquid separation system. The day of analysis, metabolomic preps were resuspended in 5:1 MeoH:H_2O and sonicated to fully solubilized. 5uL were injected into the autosampler system. LC separation was performed on a Waters Acquity BEH HILIC column. LC gradient separation used H_2O + 0.1% Formic Acid (FA) as solvent A and Acetonitrile +0.1% FA as solvent B. LC separation gradient: 0-1min 1%A, 1-11min 1%A to 99%A, 11-12.5min 99%A, 12.5-14min 99%A-1%A, 14-15min 1%A. Metabolomics analysis was performed in positive mode, 0.5s scan length, 50-1200Da m/z range, at 2.00kV capillary voltage, and 80°C source temperature.

Lipidomics. LCMS acquisition for the lipidomics fraction were performed as above with some changes: LC gradient; 0-1min 99%A, 1-22min 99%A-1%A, 22-25min 1%A, 25-25.1min 1%A-99%A, 25.1min-28min 99%A. Lipidomics was performed in negative mode, with all other parameters the same as metabolomics other than a 1.8kV capillary voltage.

Data processing

Raw data was imported into Nonlinear Dynamics (Waters corp) Progenesis QI for processing and assignment of identifications. The default peak picking algorithm was employed, retention times aligned to pooled samples, and all default adducts were allowed for analysis. After peak picking and alignment, CSV files were exported from Progenesis QI for statistical analysis in Metaboanalyst. Filtering and normalization took place in Metaboanalyst as follows; filtered to the top 2500 features based on interquartile range, quantile normalized, log transformed, and auto-scaled.

IDs

Tentative identifications were assigned through the use of Progenesis Metascope, an in-silico structural database, theoretical fragmentation based searching algorithm. Chemical structures are acquired from metacyc, lipidmaps, and lipidblast online databases. Measured M/Z features are first assessed for all potential chemical formulae using M/z value and observed isotopic distribution, within 20ppm. Measured mz formulae are then matched to formulae from the database, and any matches undergo theoretical fragmentation to match observed MS/MS peaks with predicted fragmentation patterns. Identifications are assigned a score based on the cumulation of all of these parameters. All identifications reported in this document have a high

score based on isotopic distribution and PPM error, and many also include matches to MS/MS fragments.

Figures and Tables

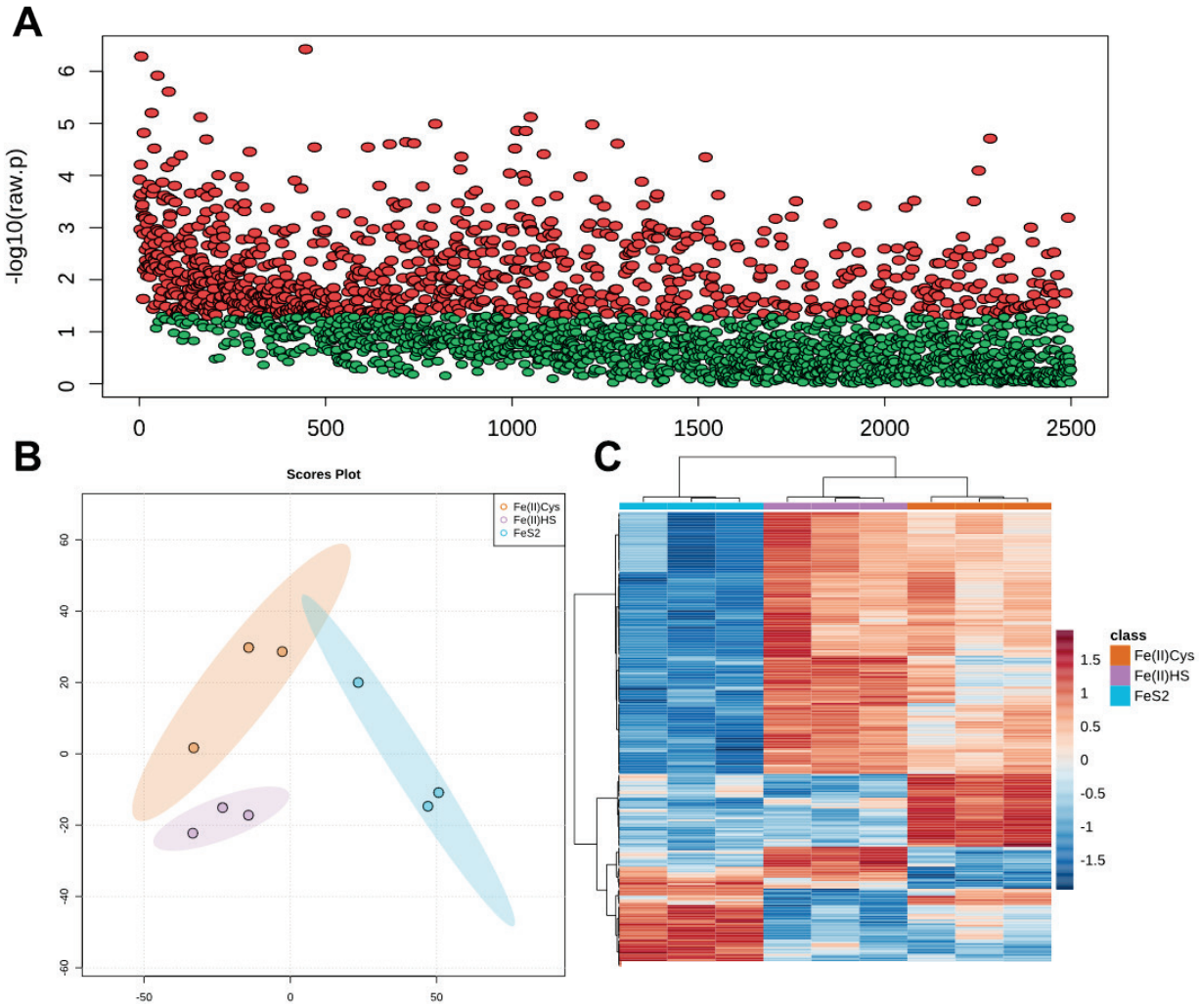


Figure 1: Global Multivariate statistics summary. A: one way ANOVA of 2500 filtered metabolomics features, 981 change significantly (Pvalue<0.05, red) and 1519 remain unchanged (green). B: Principal component scores plot of processed metabolomics data (Purple = $Fe(II)/HS^-$, orange = Fe(II)/Cys, blue = FeS_2). Ovals indicate 95% condifence interval. Plot shows PC1 (42.% of total variance) and PC2 (17.3% of total variance) coordinate pairs from each sample. C: Parallel hiercical clustering anaylsis (PCHA) heatmap of the top 200 features, filttered by lowest p-value. Samples are clustered in columns, features are clustered in rows, and their intersection represents the abundance of each feature in each sample relative to the average (Red=high abundance, blue= low abundance.)

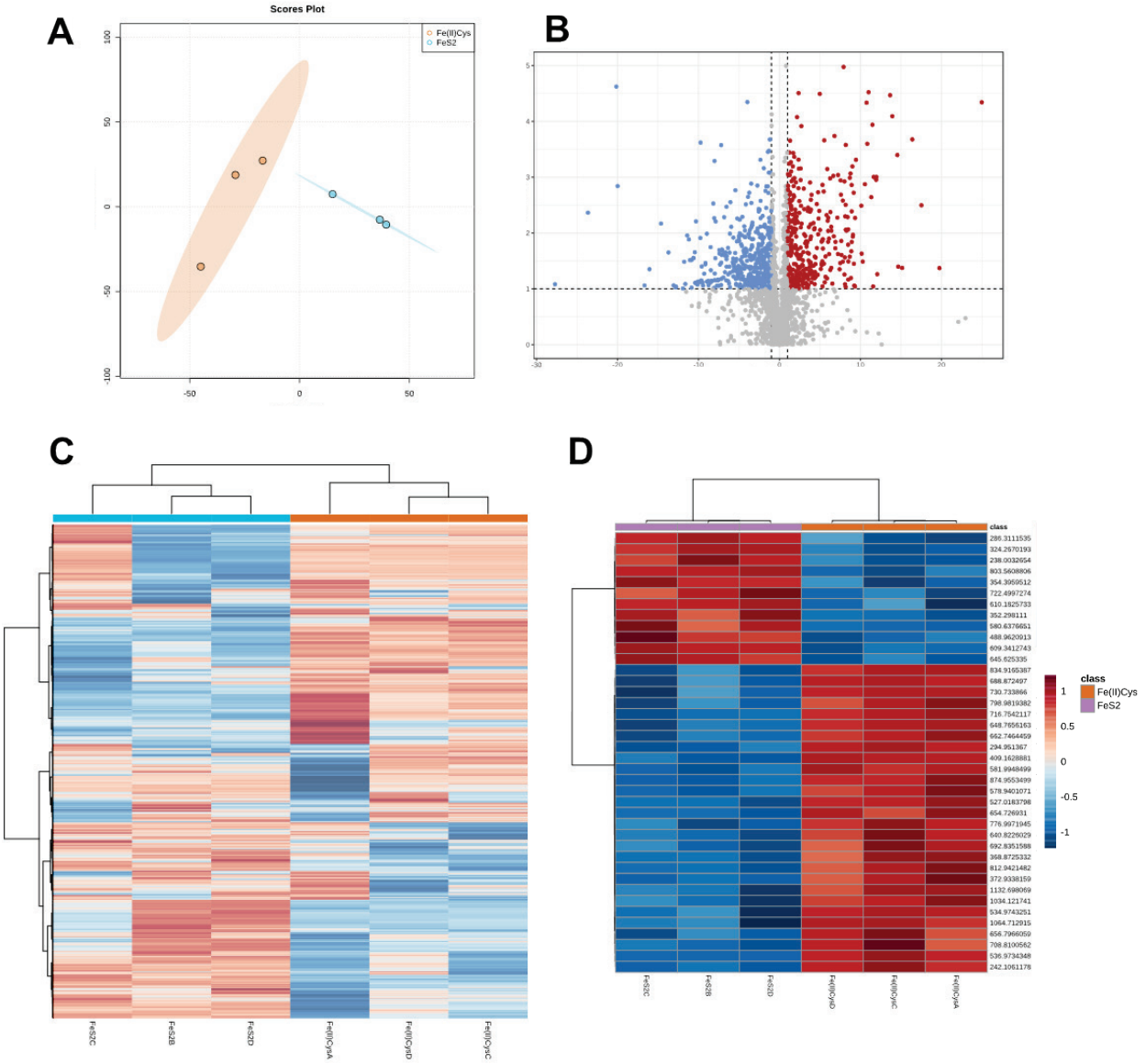


Figure S1: Two-group global multivariate statistics, FeS_2 vs Fe(II)/Cys. A: PCA scores plot. PC1 coordinate plane covers 50.2% of the total variation, PC2 covers 20.5%. Fe(II)/Cys PCA coordinates represented in orange, FeS_2 in blue. Ovals represent 95% confidence intervals. B: Volcano plot of the two-group comparison. Log transformed fold change values (Log2(FC)) on the x axis are compared to statistical significance (-log10(p)) on the y axis. Directionality of the comparison is FeS_2 vs Fe(II)/Cys, blue dots represent features lower in abundance in FeS_2 (395), red dots represent features higher in abundance in FeS_2 (460). Grey dots indicate features below the threshold of statistical significance and fold change (1645)(Pvalue<0.1, FC>|2|). C: PHCA heatmap of all features. Samples are in columns, features in rows. D: PHCA heatmap of top 40 features.

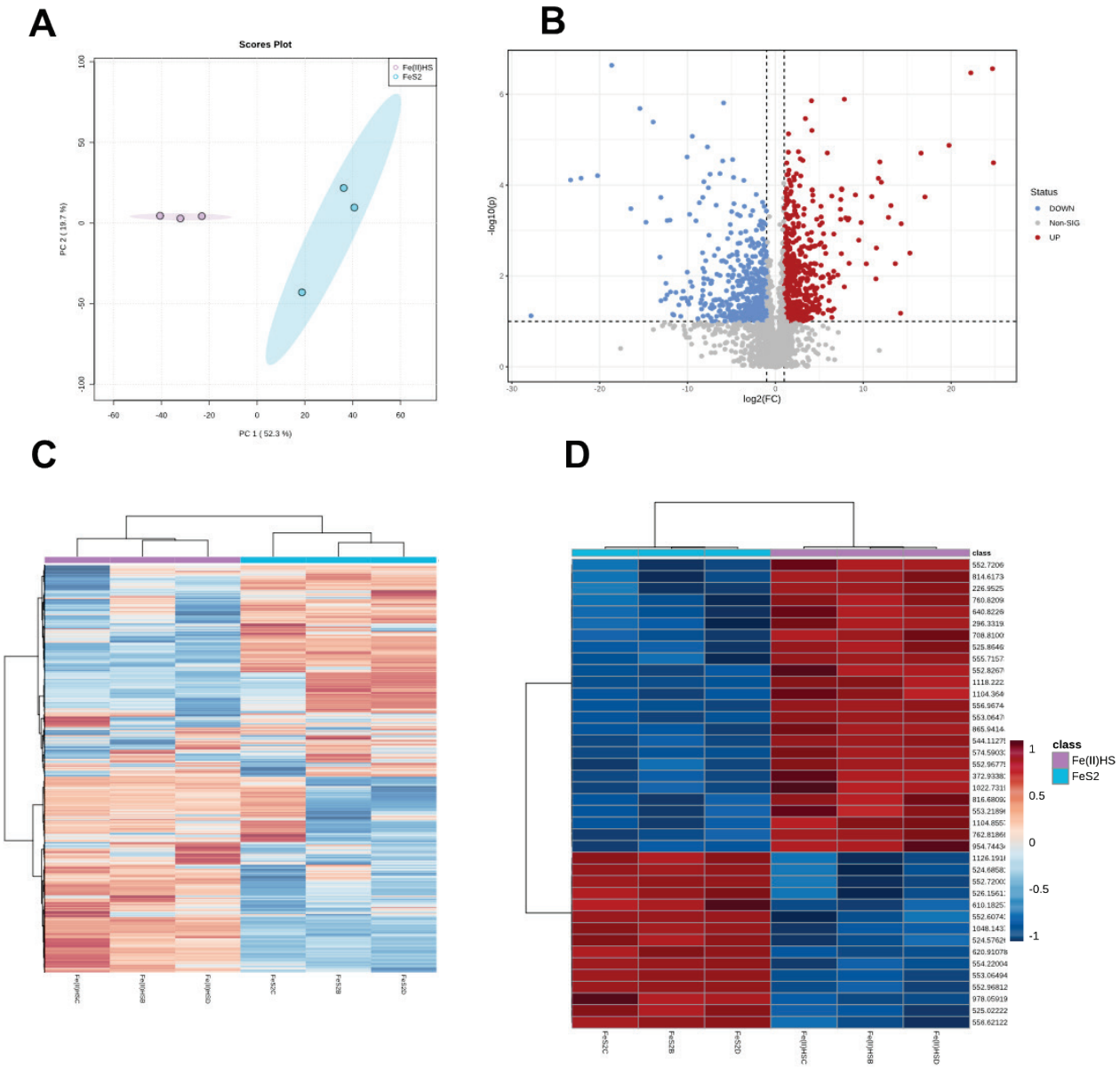


Figure S2: Two-group global multivariate statistics, FeS_2 vs $Fe(II)/HS^-$. A: PCA scores plot. PC1 coordinate plane covers 52.3% of the total variation, PC2 covers 19.7%. $Fe(II)/HS^-$ PCA coordinates represented in purple, FeS_2 in blue. Ovals represent 95% confidence intervals. B: Volcano plot of the two-group comparison. Log transformed fold change values (Log2(FC)) on the x axis are compared to statistical significance (-log10(p)) on the y axis. Directionality of the comparison is FeS_2 vs $Fe(II)/HS^-$, blue dots represent features lower in abundance in FeS_2 (564), red dots represent features higher in abundance in FeS_2 (493). Grey dots indicate features below the threshold of statistical significance and fold change (1443)(Pvalue<0.1, FC>|2|). C: PHCA heatmap of all features. Samples are in columns, features in rows. D: PHCA heatmap of top 40 features.

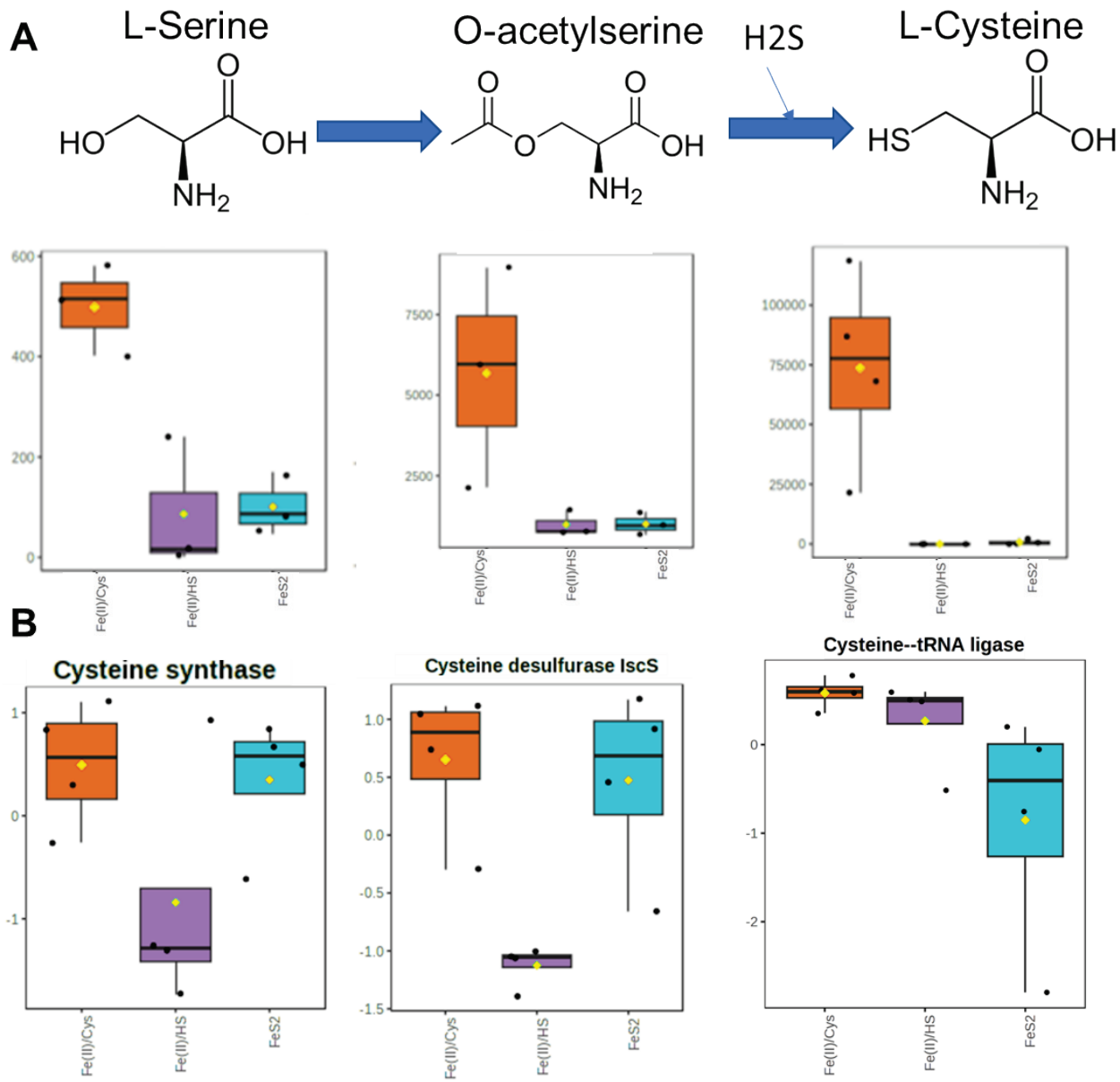


Figure 2: Dysregulation of cysteine biosynthesis metabolites and related proteins. A: Pathway diagram of the 3 step synthesis of L-cysteine. L-Serine is converted to O-acetylserine by serine acetyltransferase (not identified) and subsequently combined with hydrogen sulfide (H2S) to form L-cysteiene by cysteine synthase. Relative abundances (absolute MS feature intensity) of each metabolite are shown as box-and-whisker plots below its structure. B: Box-and-whisker plots of proteins that can influence L-cysteine abundance. Cysteine synthase generates L-cysteine from O-acetylserine, cysteiene desulfurase converts it into L-alanine and transfers its S to a sulfur carrier (usually a protein), and cysteine –tRNA ligase charges the cysteinyl tRNA for protein translation. Proteins abundance is shown as autoscaled (mean centered and scaled to 0) values.

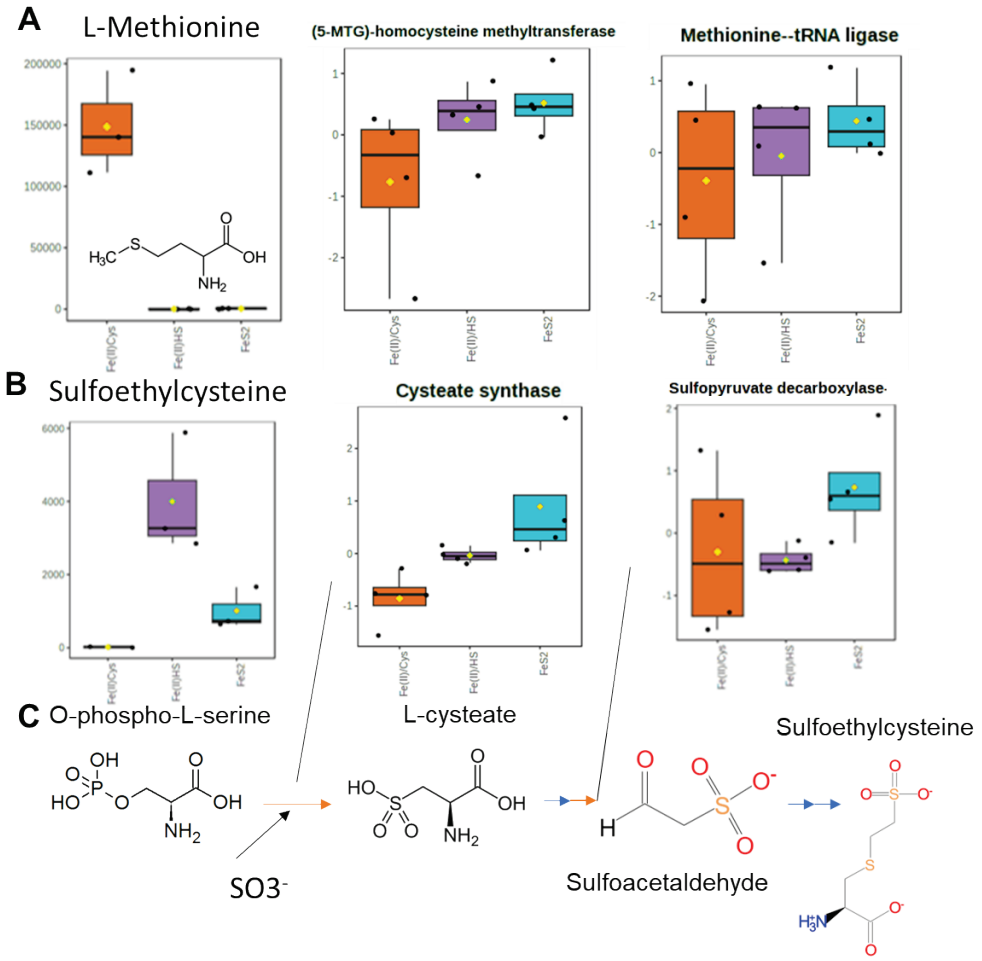


Figure 3: Methionine and sulfoethylcysteine dysregulation. A: Box-and-whisker plots of methionine and related protein relative abundances. Methionine abundance is shown as absolute MS feature intensity, 5-methyltetrahydropteryl tri-L-glutamate (5-MTG)-homocysteine methyltransferase and methionine tRNA ligase abundance are displayed as autoscaled values. B: Box-and-whisker plots of sulfoethylcysteine and related CoM biosynthesis proteins. Sulfoethylcysteiene is displayed as absolute MS feature intensity, cysteate synthase and sulfopyruvate decarboxylase are autoscaled values. C: Representative diagram of coenzyme M biosynthesis precursors. O-phospho-L-serine is combined with free sulfite to L-cysteate via cysteate synthase(identified). L Cysteate goes through two steps to become sulfoacetaldehyde, the second step is carried out by sulfopyruvate decarboxylase (identified). Sulfoacetaldehyde undergoes a two step conversion to sulfoethylcysteiene (identified). Proteins identified and quantified in the proteomics investigation are labeled via orange arrows.

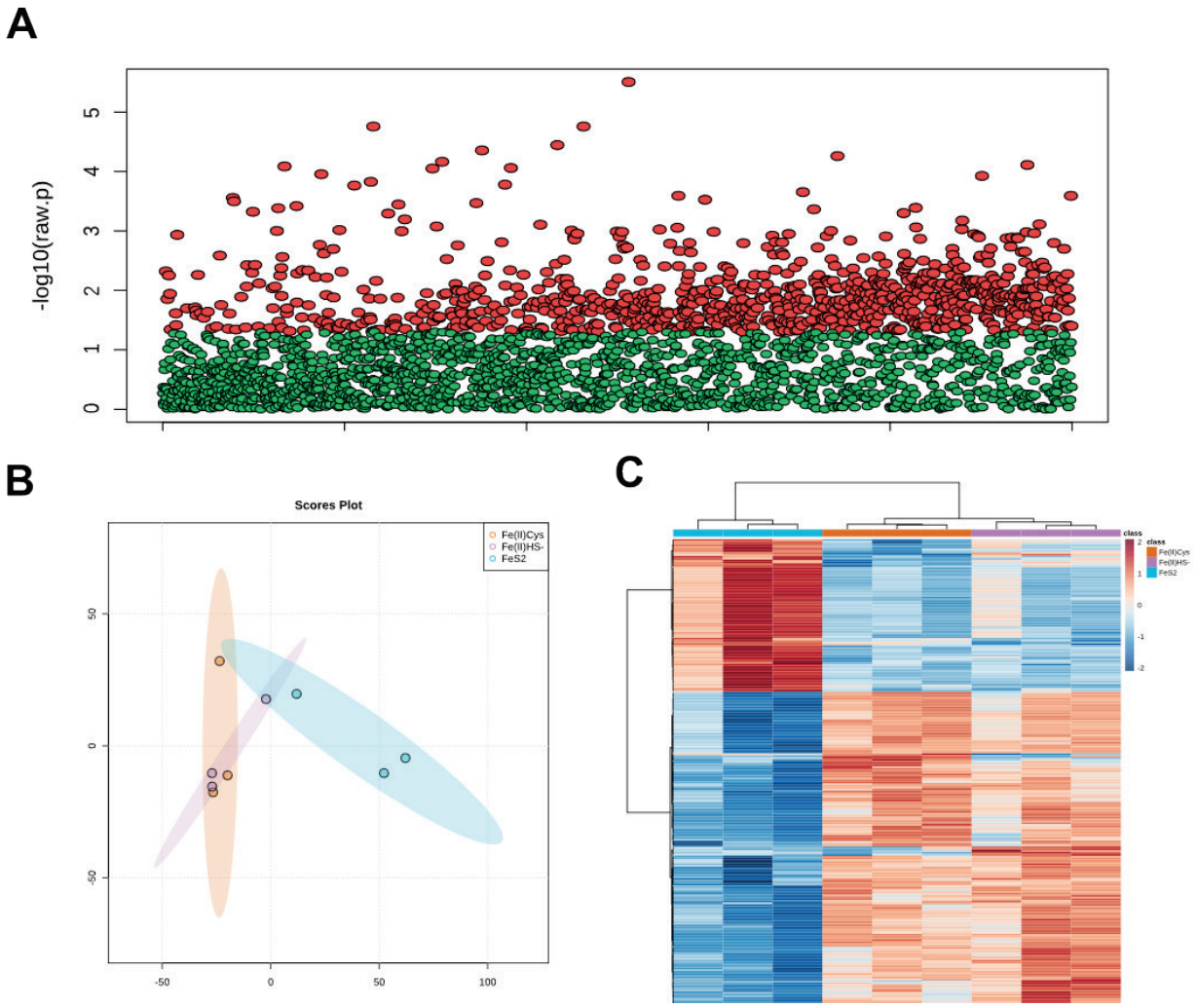


Figure 4: Global multivariate statistics for membrane composition lipidomics. A: One way anova of the 250 filtered membrane lipidomics features. 895 of these 2500 met the threshold for statistical significance (pvalue<0.05)(red). B: PCA 2D scores plot of all groups. Blue dots represent FeS_2 grown cells, orange Fe(II)/Cys, and purple Fe(II)/HS^-. Colored ovals represent 95% confidence intervals of the assigned principal component coordinates. C: PCHA heatmap of the top 200 features filtered by p-value.

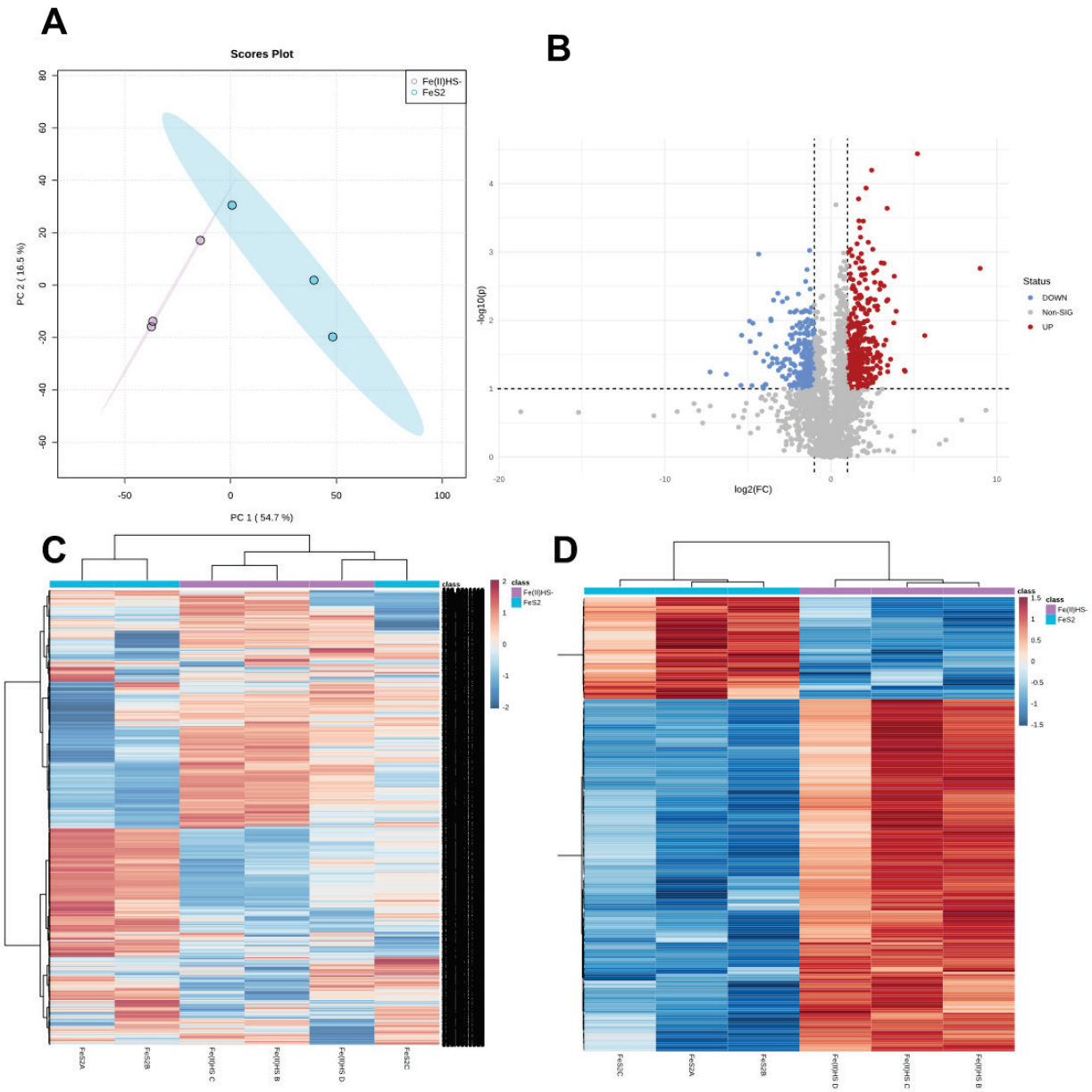


Figure S4: Membrane lipidomics two group comparison, FeS_2 vs $Fe(II)HS^-$. A: PCA 2D scores plot. B. Volcano plot of the two group comparison. Dashed lines represent pvalue and FC thresholds (0.1 and 2, respectively. Blue dots are features downregulated in the FeS_2 group(343), red dots are upregulated in FeS_2 (343), and grey features are below the stated thresholds for significance (1820). C: All feature PHCA heatmap. D: top 200 feature PHCA heatmap, filtered by pvalue

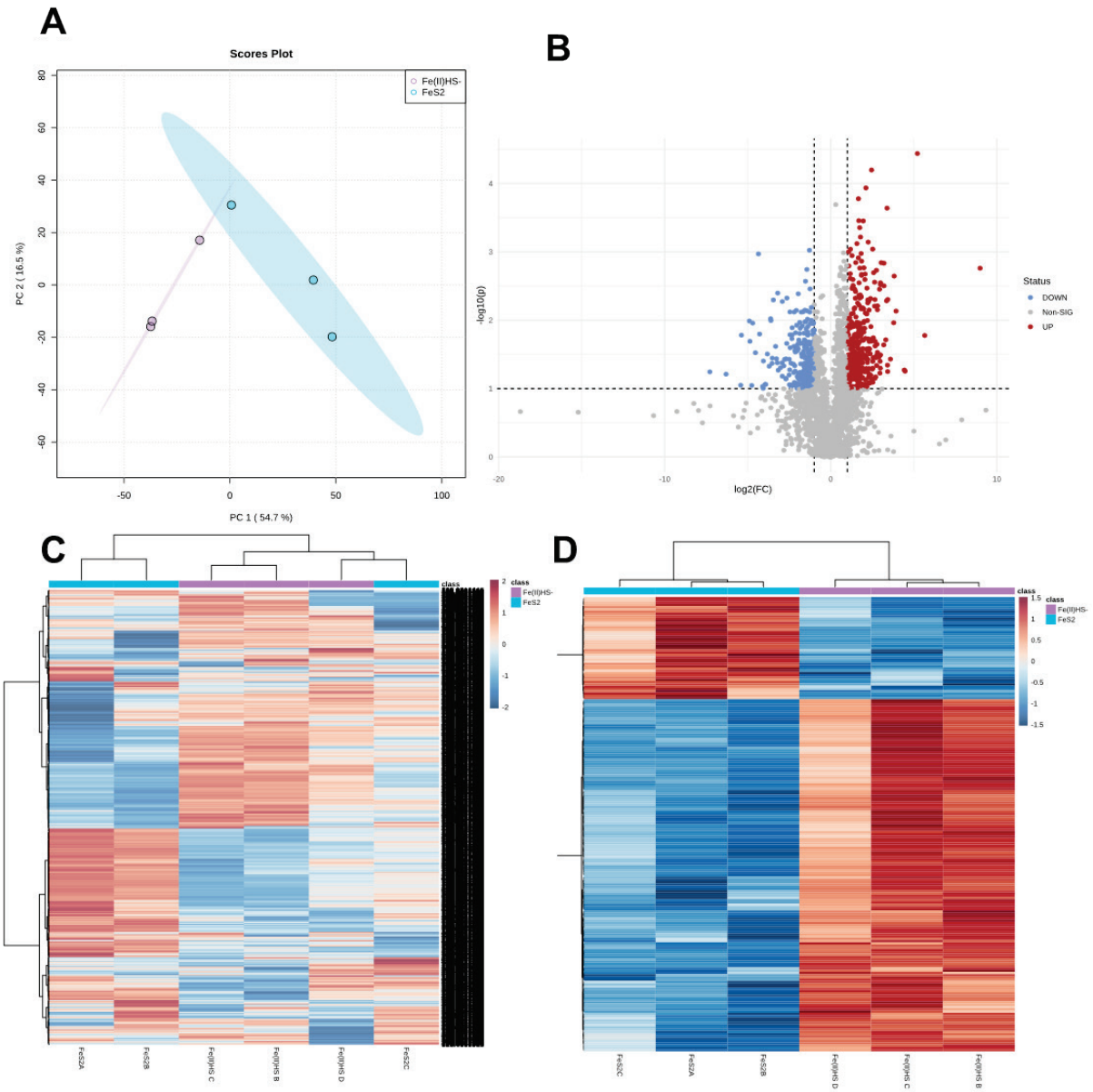


Figure S5: Membrane lipidomics two group comparison, FeS_2 vs $Fe(II)HS^-$. A: PCA 2D scores plot. B. Volcano plot of the two group comparison. Dashed lines represent pvalue and FC thresholds (0.1 and 2, respectively. Blue dots are features downregulated in the FeS_2 group(343), red dots are upregulated in FeS_2 (343), and grey features are below the stated thresholds for significance (1820). C: All feature PHCA heatmap. D: top 200 feature PHCA heatmap, filtered by pvalue

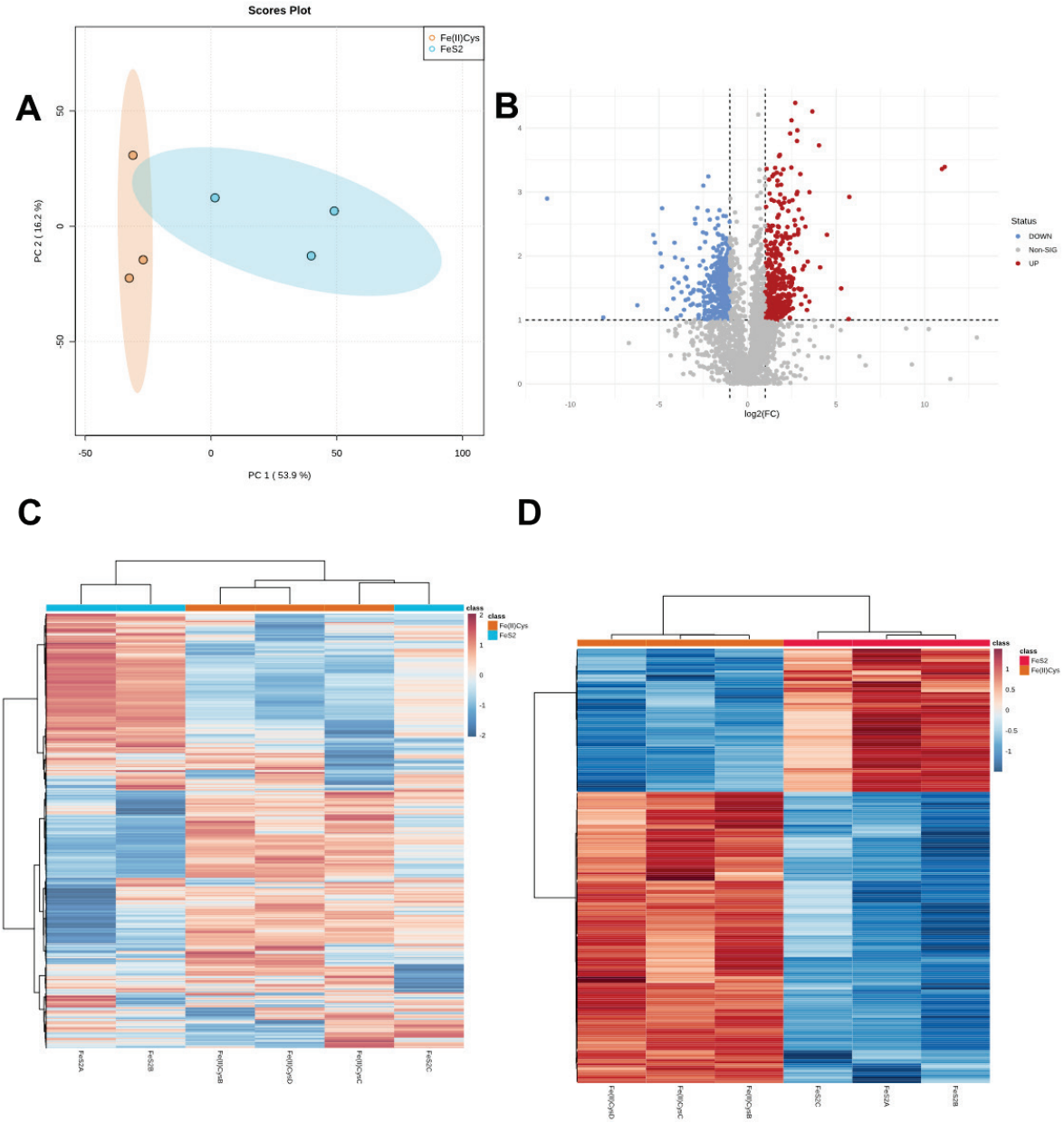


Figure S6: Membrane lipidomics two group comparison, FeS_2 vs Fe(II)/Cys. A: PCA 2D scores plot. B. Volcano plot of the two group comparison. Dashed lines represent pvalue and FC thresholds (0.1 and 2, respectively. Blue dots are features downregulated in the FeS_2 group(347), red dots are upregulated in FeS_2 (439), and grey features are below the stated thresholds for significance (1714). C: All feature PHCA heatmap. D: top 200 feature PHCA heatmap

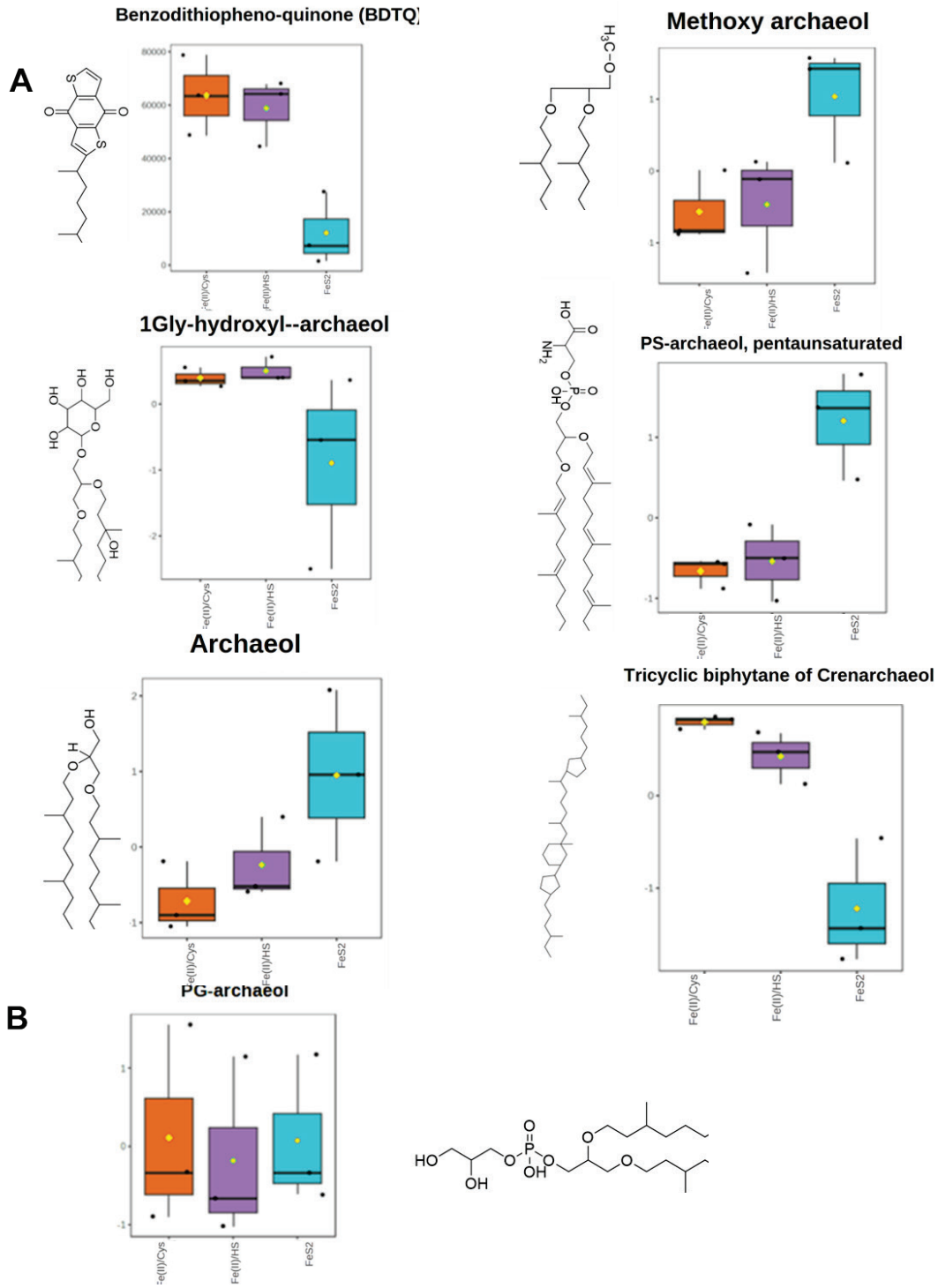


Figure 5: Dysregulation of identified Archaeol lipids. A: Strucutres and box-and-whisker plots of potential archaeol lipid identifications. Abundances in box-and-whisker plots are shown as absolute MS feature intensity for BDTQ due to low abundance in FeS_2 group, and are auto scaled for all other molecules. B: Structure and box-and-whisker plot for PG archaeol, showing no change in abundance between groups.

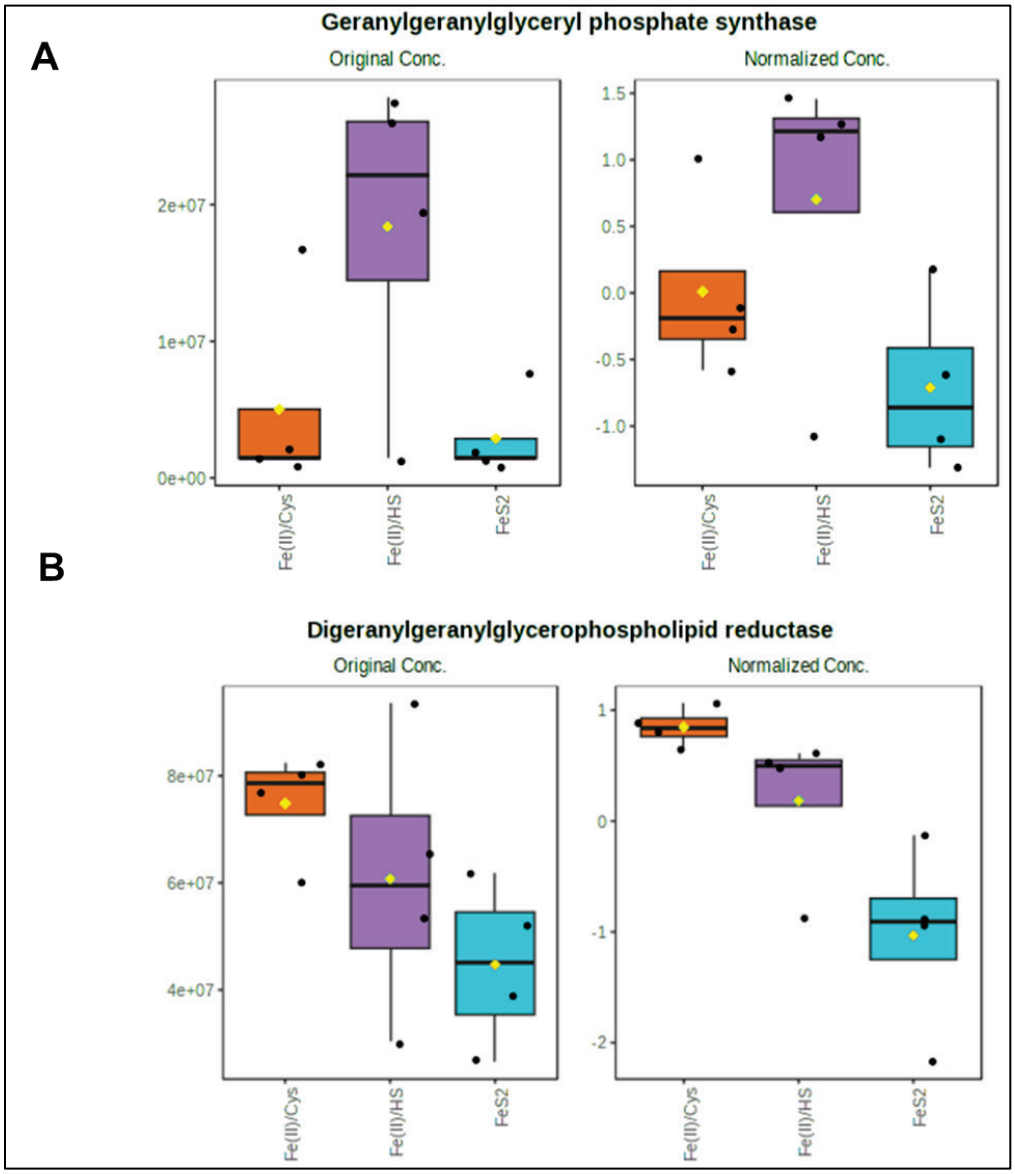


Figure S9: Regulatory patterns of Archaeol synthesis proteins. A: Absolute abundance and autoscaled box-and-whisker plot for GGPS. B: Absolute abundance and autoscaled box-and-whisker plot for DGPR

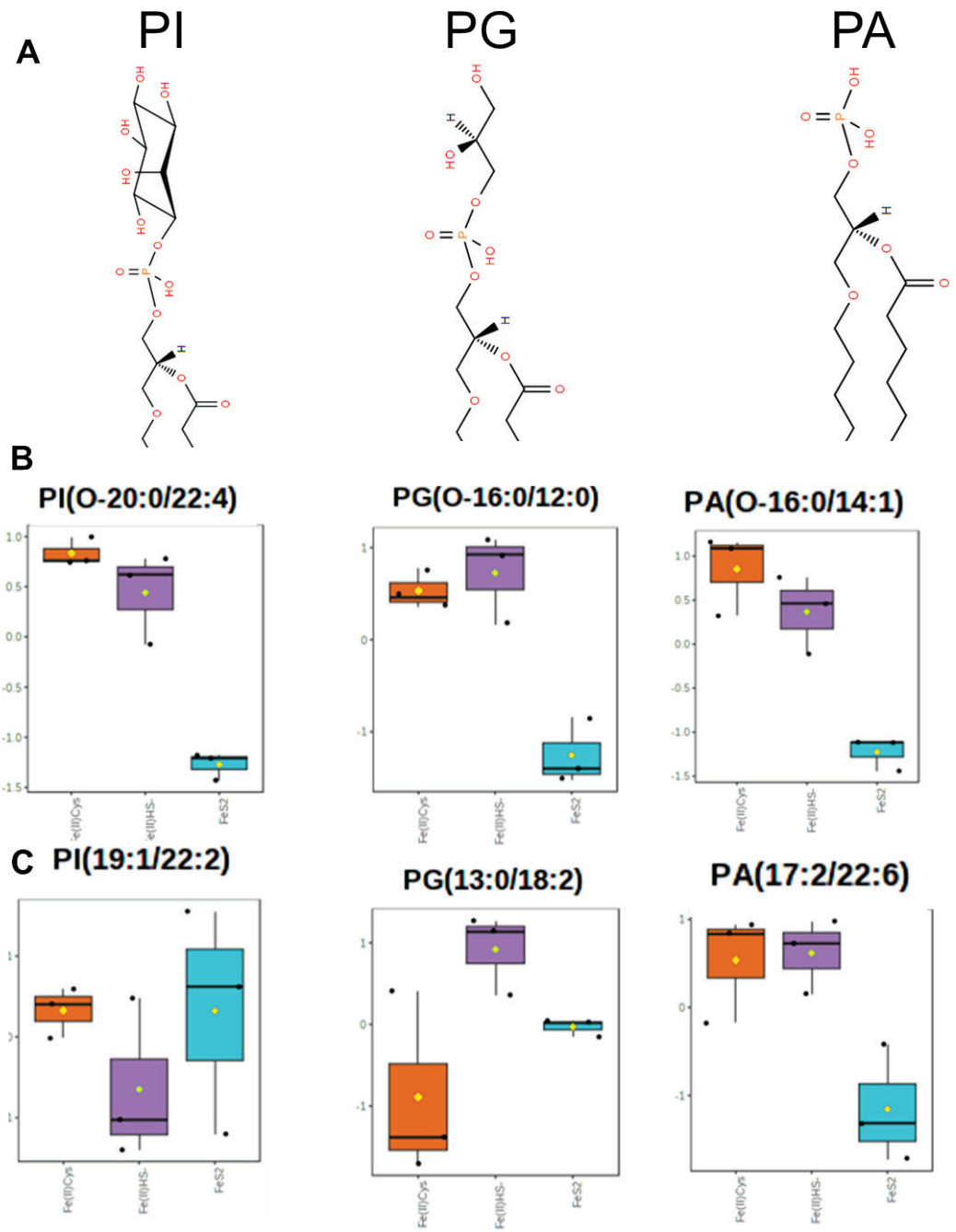


Figure 6: Phosphoglycerolipid (PGL) dysregulation. A: Structures of the three classes of PGL observed changing between groups. Acyl chains are shortened to illustrate differences in headgroup chemistry. B: Box-and-whisker plots of identified even-chain PGLs in each class. C: Box-and-whisker plots of identified PGLs containing one odd chain acyl group, from each class. Abundaces are autoscaled

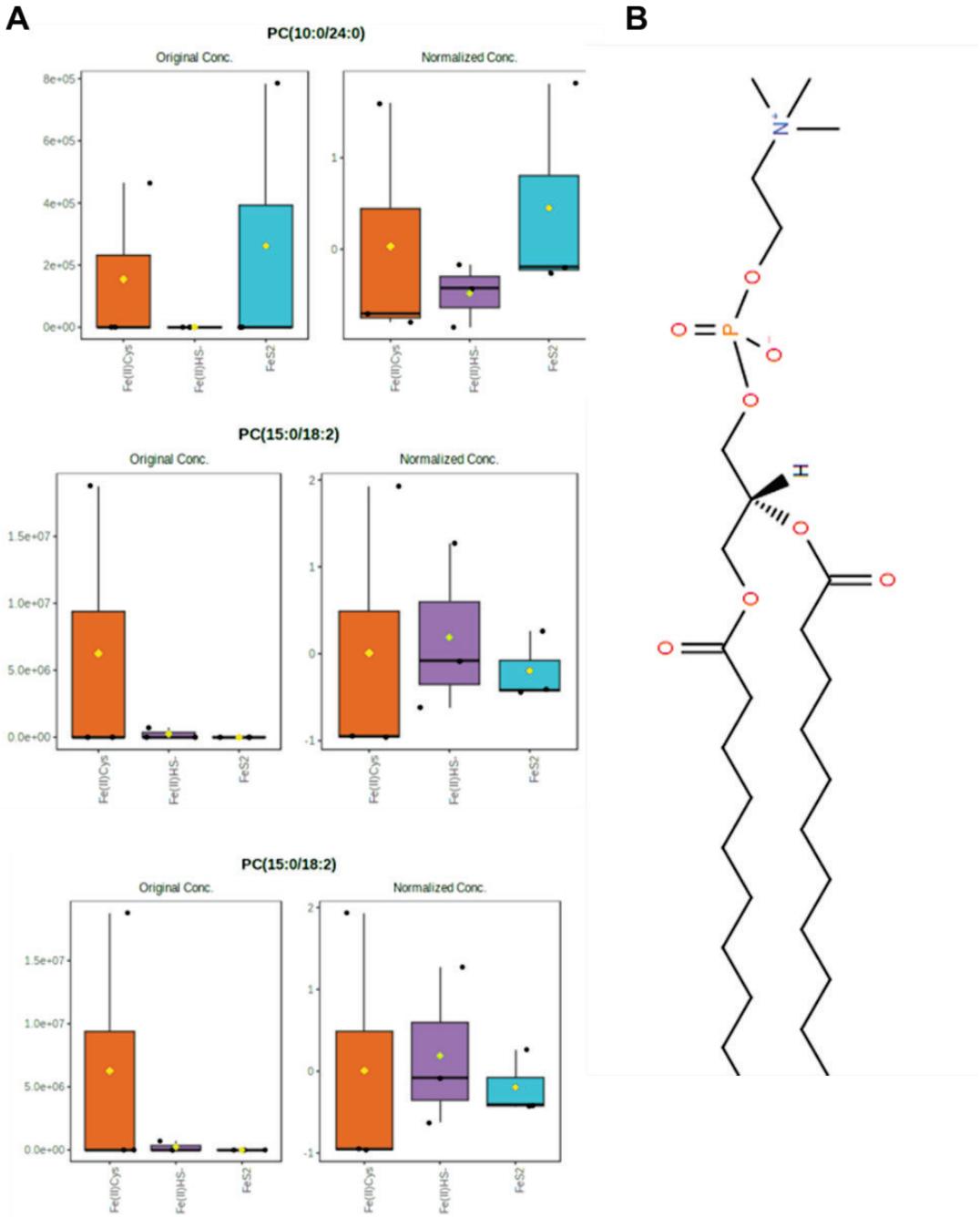


Figure S8: Abundance patterns of PC GPLs. A: Box-and whisker plots of all identified PC GPLs. Abundances are shown as absolute MS feature abundance as well as auto scaled abundance to highlight the low abundance in both PC(15:0/18:2) species in the Fe(II)/HS^- and FeS_2 conditions.

Phosphoglyercolipid	ANOVA p-value	Phosphoglyercolipid	ANOVA p-value
PA(17:2/22:6)	0.015418	PI(17:2/0:0)	0.078
PA(17:2/22:6)	0.021582	PI(18:1/22:2)	0.54642
PA(18:0/18:1)	0.26204	PI(18:4/22:4)	0.071687
PA(18:2/0:0)	0.00016633	PI(18:4/22:4)	0.0082998
PA(18:3/20:5)	0.11842	PI(19:1/22:2)	0.44416
PA(18:3/20:5)	0.075194	PI(22:2/22:6	0.27454
PA(18:3/20:5)	0.0033694	PI(22:2/22:6)	0.37818
PA(18:3/20:5)	0.00087265	PI(O-20:0/22:4)	0.16401
PA(18:4/0:0)	0.6811	PI(O-20:0/22:4)	0.00022254
PA(18:4/0:0)	0.36521	PI(O-20:0/22:4)	0.0019673
PA(18:4/0:0)	0.043837	PI(O-20:0/22:4)	0.010817
PA(18:4/0:0)	0.90445	PI(P-20:0/22:6)	0.006427
PA(18:4/22:6)	0.15802	PG(14:0/0:0)	0.081685
PA(18:4/22:6)	0.069607	PG(14:0/14:0)	0.32478
PA(18:4/22:6)	0.020849	PG(14:0/14:0)	0.27238
PA(22:2/22:6)	0.33968	PG(17:2/22:6)	0.2112
PA(6:0/6:0)	0.92522	PG(O-16:0/12:0)	0.31715
PA(6:0/6:0)	0.21194	PG(O-16:0/12:0)	0.0012051
PA(8:0/8:0)	0.5757	PG(O-16:0/12:0)	0.10472
PA(O-16:0/12:0)	0.029395	PG(P-16:0/16:1)	0.060374
PA(O-16:0/14:1)	0.40949	PG(P-16:0/16:1)	0.17079
PA(O-16:0/14:1)	0.0013882	PG(P-16:0/16:1)	0.19944
PA(O-16:0/14:1)	0.023269	PC(10:0/24:0)	0.58379
PA(O-16:0/14:1)	0.0049987	PC(15:0/18:2)	0.91893
PA(O-16:0/14:1)	0.0081526	PC(15:0/18:2)	0.9519
PA(O-16:0/14:1)	0.0042164		
PA(O-16:0/14:1)	0.0037486		
PA(O-16:0/14:1)	0.0085045		
PA(O-16:0/22:6)	0.7945		
PA(O-18:0/14:0)	0.25708		
PA(P-20:0/0:0)	0.057878		

Table 1: All annotated PA, PI, PG, and PC. Yellow cells are p-value<0.05, red cells are 0.05< pvalue <0.1

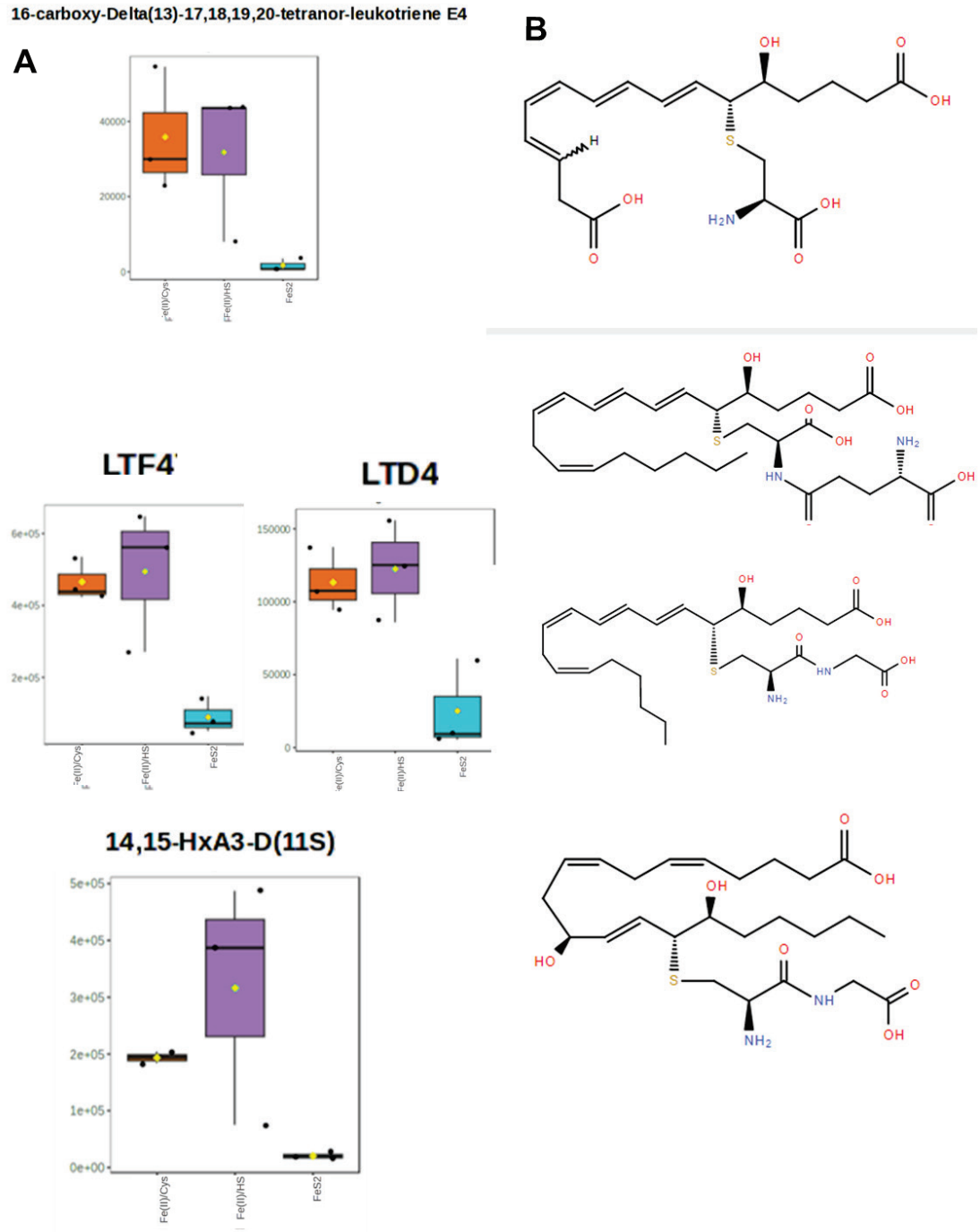


Figure 7: Cysteinyl- carotenoid like lipids and their dysregulatory patterns. A: Box-and-whisker plots of the abundances of cysteine containing lipid IDs. Abundances are shown as absolute MS feature intensity to highlight low abundance in the FeS_2 condition. B: Structures of identification hits.

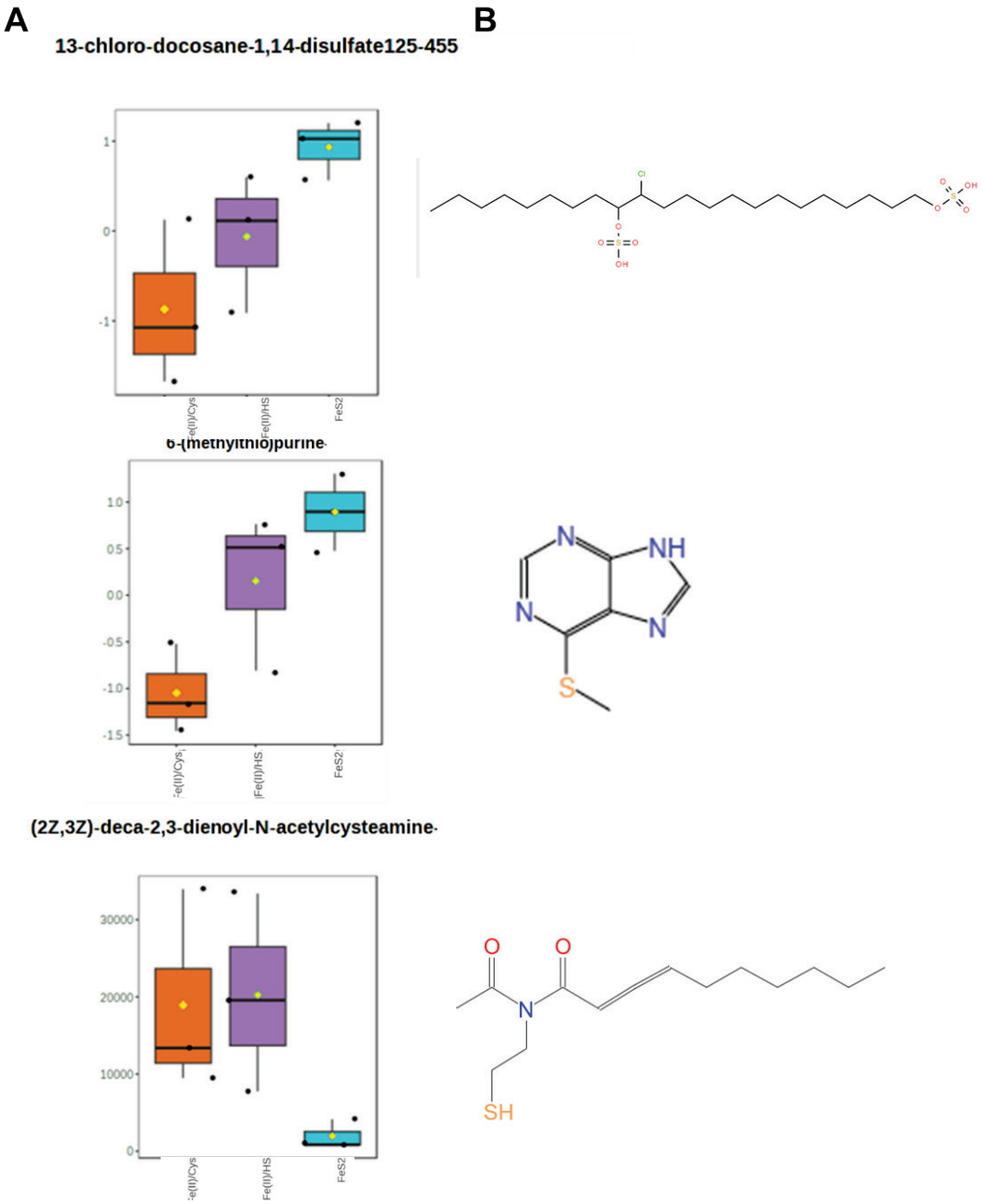


Figure S9: Non-cysteine containing lipid//nonpolar molecule ID hits and their abundances. A: box-and-whisker plots of the abundance of three S containing molecules. 13-chloro-docosane-1,14-disulfate and 6-(methylthio)purine are shown as autoscaled abundances to better showcase the differences between conditions, (2Z,3Z)-deca-2,3-dienoyl-N-acetylcysteamine is displayed in absolute MS abundance due to low abudnace in FeS_2 group. B: structures of the ID hits corresponding with the box-and-whisker plots

References:

1. Iron | Linus Pauling Institute | Oregon State University. https://lpi.oregonstate.edu/mic/minerals/iron. Retrieved 10 April 2023.

2. Dlouhy AC, Outten CE. 2013. The iron metallome in Eukaryotic organisms. Met Ions Life Sci 12:241–278.

3. Beinert H, Holm RH, Münck E. 1997. Iron-sulfur clusters: nature's modular, multipurpose structures. Science 277:653–659.

4. Johnson DC, Dean DR, Smith AD, Johnson MK. 2005. Structure, function, and formation of biological iron-sulfur clusters. Annu Rev Biochem 74:247–281.

5. Lau CKY, Krewulak KD, Vogel HJ. 2016. Bacterial ferrous iron transport: the Feo system. FEMS Microbiol Rev 40:273–298.

6. Pattus F, Abdallah MA. 2000. Siderophores and Iron-Transport in Microorganisms. Journal of the Chinese Chemical Society 47:1–20.

7. Rickard D, Luther GW. 2007. Chemistry of iron sulfidesChemical Reviews.

8. Wu B, Liu F, Fang W, Yang T, Chen G-H, He Z, Wang S. 2021. Microbial sulfur metabolism and environmental implications. Science of The Total Environment 778:146085.

9. Liu Y, Sieprawska-Lupa M, Whitman WB, White RH. 2010. Cysteine is not the sulfur source for iron-sulfur cluster and methionine biosynthesis in the methanogenic archaeon Methanococcus maripaludis. J Biol Chem 285:31923–31929.

10. Liu Y, Beer LL, Whitman WB. 2012. Methanogens: a window into ancient sulfur metabolism. Trends Microbiol 20:251–258.

11. Christina J, Alexis E, Mason M-E, R. CD, L. DJ, S. BE, W. MW. 2022. Pathways of Iron and Sulfur Acquisition, Cofactor Assembly, Destination, and Storage in Diverse Archaeal Methanogens and Alkanotrophs. J Bacteriol 203:e00117-21.

12. Major TA, Burd H, Whitman WB. 2004. Abundance of 4Fe–4S motifs in the genomes of methanogens and other prokaryotes. FEMS Microbiol Lett 239:117–123.

13. De Vrieze J, Hennebel T, Boon N, Verstraete W. 2012. Methanosarcina: The rediscovered methanogen for heavy duty biomethanation. Bioresour Technol 112:1–9.

14. Hook SE, Wright A-DG, McBride BW. 2010. Methanogens: Methane Producers of the Rumen and Mitigation Strategies. Archaea 2010:945785.

15. Thauer RK, Kaster AK, Seedorf H, Buckel W, Hedderich R. 2008. Methanogenic archaea: Ecologically relevant differences in energy conservation. Nat Rev Microbiol 6:579–591.

16. C. TH, E. BN, B. PP. 2005. Microbial origin of excess methane in glacial ice and implications for life on Mars. Proceedings of the National Academy of Sciences 102:18292–18296.

17. Devon P, M. SE, L. SR, Katherine S, Sue B, Mark Y, Brian B, E. BW, B. BJ, S. BE, Y. GM. 2022. Examining Pathways of Iron and Sulfur Acquisition, Trafficking, Deployment, and Storage in Mineral-Grown Methanogen Cells. J Bacteriol 203:e00146-21.

18. Spietz RL, Payne D, Kulkarni G, Metcalf WW, Roden EE, Boyd ES. 2022. Investigating Abiotic and Biotic Mechanisms of Pyrite Reduction. Front Microbiol. Lausanne : https://doi.org/10.3389/fmicb.2022.878387.

19. Boyd E, Payne D, Shepard E, Spietz R, Guo Y, Broderick W, Broderick J. 2022. Biomining metals from pyritic ores: physiological, biochemical, and biophysical considerations. Biophys J 121:296a.

20. Spietz RL, Payne D, Szilagyi R, Boyd ES. 2022. Reductive biomining of pyrite by methanogens. Trends Microbiol https://doi.org/https://doi.org/10.1016/j.tim.2022.05.005.

21. Payne D, Spietz RL, Boyd ES. 2021. Reductive dissolution of pyrite by methanogenic archaea. ISME J 15:3498–3507.

22. Amenabar MJ, Boyd ES. 2019. A review of the mechanisms of mineral-based metabolism in early Earth analog rock-hosted hydrothermal ecosystems. World J Microbiol Biotechnol 35:29.

23. Truche L, Berger G, Destrigneville C, Guillaume D, Giffaut E. 2010. Kinetics of pyrite to pyrrhotite reduction by hydrogen in calcite buffered solutions between 90 and 180 C: Implications for nuclear waste disposal. Geochim Cosmochim Acta 74:2894–2914.

24. Lyons TW, Severmann S. 2006. A critical look at iron paleoredox proxies: New insights from modern euxinic marine basins. Geochim Cosmochim Acta 70:5698–5722.

25. Schoonen MAA. 2004. Mechanisms of sedimentary pyrite formation. Geological Society of America Special Papers 379:117–134.

26. Berner RA. 1984. Sedimentary pyrite formation: An update. Geochim Cosmochim Acta 48:605–615.

27. Klein C. 2005. Some Precambrian banded iron-formations (BIFs) from around the world: Their age, geologic setting, mineralogy, metamorphism, geochemistry, and origin. American Mineralogist 90:1473–1499.

28. Hall AJ. 1986. Pyrite-pyrrhotine redox reactions in nature. Mineral Mag 50:223–229.

29. Law KP, He W, Tao J, Zhang C. 2021. A Novel Approach to Characterize the Lipidome of Marine Archaeon Nitrosopumilus maritimus by Ion Mobility Mass Spectrometry. Front Microbiol 12:3416.

30. Steward KF, Payne D, Kincannon W, Johnson C, Lensing M, Fausset H, Németh B, Shepard EM, Broderick WE, Broderick JB, Dubois J, Bothner B. 2022. Proteomic Analysis of Methanococcus voltae Grown in the Presence of Mineral and Nonmineral Sources of Iron and Sulfur. Microbiol Spectr 10:e01893-22.

THE EFFECTS OF ARSENIC EXPOSURE ON THE LIPIDOME OF E.COLI: DETERMINING THE ULTIMATE FATE OF ARSENIC IN A MODEL MEMBER OF THE GUT MICROBIOME

Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: Hunter Fausset

Contributions: Main author. Experimental design, sample prep, data acquisition, statistical analysis, interpretation, writing, submission, editing.

Co-Author: Brett Sather

Contributions: Cell culturing, sample preparation, data processing, statistical analysis, writing, editing

Co-Author: Scott Spurzem

Contributions: Cell culturing, method optimization, sample preparation, bioinformatic data

Co-Author: Rylee Close

Contributions: Method optimization, cell culturing, sample preparation

Co-Author: Anya Knowlton

Contributions: Method optimization, cell culturing, sample preparation

processing writing,

Co-Author: Cole Kayser

Contributions: Method optimization, cell culturing, sample preparation

Co-Author: Gerogia Frances Eastham

Contributions: Method optimization, cell culturing, sample preparation

Co-Author: Sydney Peterson

Contributions: Method optimization, cell culturing, sample preparation

Co-Author: Brynne Miller

Contributions: Method optimization, cell culturing, sample preparation

Co-Author: Isabelle Johnson

Contributions: Method optimization, cell culturing, sample preparation

Co-Author: Brian Bothner

Contributions: Corresponding author, grant writing, data interpretation, writing, editing

Manuscript Information

Hunter Fausset, Brett Sather, Scott Spurzem, Rylee Close, Anya Knowlton, Cole Kayser,

Gerogia Frances Eastham, Sydney Peterson, Brynne Miller, Isabelle Johnson, Brian Bothner

Status of Manuscript:
__**x**__ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Submission in Q2 2023

Abstract

Arsenic poisoning via contaminated drinking water is an unfortunately common occurrence across the globe, especially in developing countries. The metal affects human health through multiple mechanisms, causing cancer, metabolic dysfunction, and cardiovascular disease with chronic exposure. Studies have shown that specific members of the human gut microbiome, including *Escherichia coli*, play a part in its overall toxicity. Microbial detoxification mechanisms are relatively well characterized, however, less clear is the ultimate fate of arsenic in microbial biomass. This study sought to characterize the fate of As in populations of *E.coli*, as well as its overall effects on the bacterial lipidome. Two strains, K12 and AW3110, were assessed for global lipidomic changes, arsenical distribution among biological macromolecules, and for the presence of uncharacterized bioarsenicals when grown in the presence or absence of arsenic. LC-MS based global lipidomic profiling was paired with ICP-MS to achieve these aspirations. The overall bacterial lipidome is altered by the presence of arsenic. Cultures exposed to arsenite or arsenate separate from controls in unsupervised statistical analyses. Lipid profiles are also dependent on bacterial strain, with AW3110 and K12 cultures having significant differences in the presense of arsenic. Key, metabolites like choline and dethiobiotin as well as glycerol-phosoethanolamines and other membrane lipids contribute to the lipidomic discrepancies. Overall, this study provides a basis for the full characterization of the interaction between *E.coli* membranes/nonpolar metabolites and As, and will inform future microbiome scale investigations that could lead to novel therapeutic treatments.

Introduction

Millions of people worldwide are exposed to harmful concentrations of arsenic (As)(1, 2). This typically occurs through ingestion of water and food (2–5). As a metalloid, As has grave effects on human health and is known cause cancer, cardiovascular disease, and various metabolic disorders(6–9). Save for inhaled or cutaneous methods of exposure, As interacts with the gut microbiome (GM) prior to making it to many of the affected tissues(3, 10–12). This has direct effects on the GM, but also provides an opportunity for detoxification prior to As reaching the intestines. Certain members of the GM, including *Escherichia coli,* have been shown to be enriched in populations chronically exposed As(13, 14). This organism, and others in the GM, inherently contain As detoxification mechanisms such as the Ars operon(15–19).

The Ars operon is a set of genes allowing for the transformation and subsequent efflux of inorganic arsenicals (iAs). iAs represent the most toxic forms of the metal due to their reactivity and relative solubility. Two iAs species, arsenate (As^{V}) and arsenite (As^{III}), are the most common environmental forms resulting in biological exposure(20). As^{V} mimics pentavalent phosphorous species such as phosphate (PO_4), and has been shown to interact with enzymes that utilize PO_4 as a substrate. As^{III} is more chemically reactive, and readily binds with any free thiols in the cell, including those found in the active sites of many key metabolic enzymes(21). The Ars operon confers the ability to reduce As^{V} to As^{III} and efflux (ArsC) and efflux the trivalent species from the cell (ArsB)(22). ArsR is a transcriptional regulator influencing genes related to As detoxification as well as other key cellular processes. This and other As detoxification mechanisms are relatively well characterized, yet the ultimate fate of As in many prokaryotic systems remains unknown. The degree to which As reacts with intracellular molecules prior to

being exported out of the cell is of particular value in characterizing the overall effect of As on the GM as a system.

A range of bioarsenicals have been characterized, with the majority of work coming from marine systems, namely fish and algae(23, 24). Fungi have also been shown to produce specific bioarsenicals, but prokaryotes of the GM are understudied for their As containing molecules. Additionally, knowledge of the effects of arsenic exposure on the membranes and lipid metabolism of *E.coli* is lacking, as previous studies have typically focused on polar, soluble compounds or were performed in other organisms(21, 25). Thus, several questions surrounding As containing lipids and other nonpolar molecules remain. It is important to characterize the arsenicals, specifically those that are nonpolar or in the membrane, found in both Ars operon containing organisms and those that lack it. Likewise, understanding the impact of As exposure on lipid metabolism as a whole will provide insight into the both the formation of bioarsenicals as well as the associated stress response.

E.coli is an ideal model organism to study these effects in a controlled environment. It has been shown to contribute to the GM-based detoxification of arsenic in human cohort, is well characterized and easy to grow, and strains of the organism exist that both contain (K12) and lack (AW3110) the Ars operon(13, 26). Understanding the effects of As on the *E.coli* lipidome and identifying resulting arsenicals resulting from exposure will expand the field's knowledge on the interaction with this microorganism and As. To evaluate these outstanding scientific questions, we performed LC-MS based untargeted lipidomics paired with ICP-MS based, direct arsenic quantification on K12 and AW3110 cultures exposed to either As^{III}, As^{V}, or a vehicle control. The LC-MS approach was used to evaluate the magnitude of biological change in

cellular lipids associated with each type of arsenic. This data was also used as a basis for the identification of yet-to-be characterized arsenic containing molecules. ICP-MS was used to compare the macromolecular distribution of As based on As species (As^{III} vs As^{V}) in the K12 strain. Overall, this study provides the foundation for further investigations into the GM /As interactome, specifically relating to nonpolar molecules like lipids. Understanding how the GM affects As, and vice versa, could allow for the development of future microbe-based treatments of chronic As (and other toxic metals/metalloids) exposure.

Results

ICP-MS based arsenic biotracking:

To determine the accumulation patterns of arsenic, we employed a liquid-liquid extraction-based protocol utilizing ICP-MS to quantify arsenic associated in different biochemical macromolecule classes. Specifically, insoluble lysate debris, cytosolic proteins, and cytosolic small molecules from cultures of K12 *E.coli* were assessed to compare the distribution of As resulting from treatment of As^{V} or As^{III}. The majority of the arsenic in the As^{V} treated group remained in the cytosol (65%), with less (27.4%) landing in the insoluble debris, and very little (7.5%) quantitated as associated with proteins (Figure 1). In contrast, As^{III} treated cultures saw the bulk of the arsenic (52.7%) bound to cytosolic proteins, with roughly equal distribution in the soluble and insoluble fractions (25.2%, 22,1%). Additionally, the data suggests roughly the same degree of arsenical bioaccumulation in both conditions; approximately 0.113% of the total arsenic available in the As^{V} treatment was quantified in the biomass. This is comparable to the 0.119% seen in As^{III}. Our analyses recovered 92% and 88% (K12 and AW3110, respectively) of the total arsenic associated with *E.coli* biomass, when comparing the sum of the three fractions

to the total quantified arsenic in whole cell digestions. These data suggest there to be quantifiable arsenic accumulation in the cellular membrane, cytoskeleton, or nucleic acid pools of *E.coli* and provide a foundation for the subsequent lipidomics investigation.

Global lipidomics results:

Cell cultures of either the AW3110or K12 (MG1655) strains of *E.co*li were grown in the presence (treatment) or absence (control) of As^{V} or As^{III} in order to explore the effect of arsenic on the lipidome of *E.co*li. Cells were harvested mid log phase and subjected to an optimized lipid extraction protocol prior to MS analysis. Our results yielded over 5000 MS features (5037) over the 28-minute C18 chromatography LC run. The data was normalized, run through a missing value estimation (KNN algorithm replaces missing values, mz values missing in >50% of samples were removed) and filtered based on mean down to the 2500 most relevant features. In our initial, all group comparison, 935 of these features met the threshold for statistical significance (raw p-value <0.05 from the one-way ANOVA), indicating they are significantly changing in at least one of the experimental groups (Figure 2). A PCA was performed to assess the magnitude of lipidomic change. The PCA plot shows a general cluster in the middle, where the AW3110 As^{V} (AWV) and control (AWC) groups significantly overlap with K12 As^{V} (K12V) and control (K12C) (Figure 2). Both groups treated with AsIII (AWIII and K12III) group away from this central cluster and on opposite ends of the plot from one another, suggesting their lipidomes to be markedly different than the other groups. A parallel hierarchical clustering heatmap (PCHA), filtered down to the top 500 features based on p-value, perfectly cluster each group (Figure 2). The high variability between samples, even within the same group, can be observed in this figure (e.g the band of upregulated features in AWV2). This is further evident in

the all-feature PCHA heatmap (Figure S1). Interestingly, the K12III group clustered away from all other groups, with the other two K12 cultures and all three AW3110 conditions grouped tightly together. There interesting group-wise trends present in the top 500 feature PCHA heatmap, as marked by the cluster boxes. To investigate the most biologically relevant features, a top 20 feature (filtered by p-value) heatmap was generated (Figure S2). Perfect clustering and clear patterns emerge from this heatmap, with the As^{III} groups clustering together due to two clusters of heavily upregulated (K12III) or downregulated (AWIII) features. Tentative identifications were assigned to the features from this heatmap (Table 1). The cluster of downregulated features in the AWIII group contains several long fatty acid chains (e.g octadecane, 9-methoxypentadecanoic acid) as well as dethiobiotin and a sterone-like molecule. The cluster of highly abundant features in K12III largely was not identified, but two arsenic containing lipids were annotated.

Intra-strain comparison:

In an attempt to look deeper into the data, 6-way comparison quickly becomes unwieldy to interpret, and can obscure observable intra-supergroup (in this case intra-strain) statistical differences, so the groups were separated by strain and reassessed for global statistical changes. When analyzed in this manner, clear differences arise between arsenic conditions within the AW3110 comparison (Figure 3). The PCA shows complete separation of the confidence intervals between groups, also exhibiting a relative contraction of variation for the AWV group, while the AWIII group appears more spread out, and the control group is somewhere between those. This is clearly echoed in the PCHA heatmap for this intra-strain comparison, with perfect clustering when considering all features (Figure 3). When the K12 strain is separated in this way,

key differences between within the groups become evident. The PCA shows K12C grouping away from the both K12 cultures treated with arsenic. The latter groups, K12III and K12V, overlap in this plot which is indicative of relative similarity between their lipidomes. This is not entirely mirrored in the all-feature PCHA, largely due to a distinctly upregulated cluster of lipids in the K12III group (highlighted in the cluster box) (Figure 3).

Further filtering of the data provides a closer look at the specific lipids changing in abundance. The top 40 feature heatmap (filtered by p-value) for AW3110 (Figure 4) and K12 (Figure 5) display distinct clusters of dysregulated features. 7 such clusters are present in the AW3110 groups and 4 patterns emerge from the K12 comparison (Figure 5). In both heatmaps, the As^{III} treatment clusters away from the others, mirroring the patterns observed in the intragroup all feature PHCA heatmaps. Regulation is varied in the AW3110 comparison, with many different regulatory patterns in the 6 clusters (Figure 4). K12, however, is more uniform in these top 40 features, with all but 1 being higher in abundance in the As^{III} treatment group relative to the average. A large band of highly abundant features in the K12III dominates clusters 3 and 4. Both heatmaps were subjected to the identification pipeline (Table 2). Over half of the features from the AW3110 heatmap were assigned tentative identifications (22/40), while only 7 IDs were made for the K12 comparison.

A closer looker at the top 40 features from the heatmap gave a list of tentative identifications (Figure 6). From the AW3110 comparison, three glycero-phosphoethanolamines (GPEtn(x:x)) are shown from clusters 4 and 5. These share a regulatory pattern where the abundance is high in the control cells, lower in the As^{III} treatment and very low or absent in the As^{V} treatment. 3 other AW3110 specific features representing the regulation patterns seen in

clusters 3, 6, and 7 are also shown (Figure 6). An identified diacylglycerol (DG18:0/18:10) has a stark, high abundance pattern in the As^V treatment, and was measured at lower levels in the other two groups. Choline is lower in abundance in both arsenic treatment groups relative to the control. Dethiobiotin also shows abundance pattern. This molecule is an intermediate in the synthesis of biotin and other enzyme cofactors, and is displayed as raw abundance to highlight its complete absence in the AWIII experimental condition. Figure 7 displays a similar analysis of features from the K12 comparison. 3 features from the heatmap representing the first 3 clusters are shown, including two long chain fatty acids (11-oxo-undeca-5,6,dienoic acid, pentadecylic acid) as well as a glycero-phospholipid (PA(18:4/0:0)). 2 other identifications not found in the heatmap include a galactoglycerol, and an additional glycerophospholipid (Figure 7). Cluster 4 remains largely unidentified despite is stark regulatory pattern in the K12III grouping. Efforts are underway to further characterize this system, though some tentative identifications have been made for novel, arsenic containing lipids (Figure 8). These molecules match to our curated bioarsenical database, and also display a regulatory pattern consistent with an arsenical molecule (absent in controls, high abundance in arsenic treatment). Two that fit this pattern are shown in Figure 8, with upwards of 400 other matches under further investigation.

Discussion

This study shows the effects of iAs treatment on the global lipidome of *E.coli*. We first observed As distribution, based on macromolecular class, to be dependent on valence state. As^V tends to remain soluble in the cytosol of K12 *E.coli*, though some remains measurable in the proteins and little is found in the insoluble fraction. The arsenic in the proteins of this condition could be from As^V mimicking PO_4 and associating with enzymes as a substrate or post

translational phosphorylation events. ArsC also converts As^{V} to As^{III}, which can covalently bind to proteins. Thus, the As found in the protein fraction of As^{V} treated cells is likely a combination of these two factors. As^{III} reacts with thiols, and therefore proteins with any exposed cysteine residues, so most cellular arsenic in As^{III} treated cells is found in the protein fraction. The remaining As is equally distributed between the soluble and insoluble lysate. Of the arsenic available to the K12 cells, approximately the same amount is captured in the biomass for both As^{V} and As^{III} treatments (0.11% and 0.12%, respectively). These data imply that there is a quantifiable amount of As present in the insoluble/nonpolar molecules from both treatments that we can identify using our LC-MS based lipidomics approach. This also could indicate that cells are undergoing some degree of perturbation to their lipid metabolism, further justifying an untargeted global lipidomics analysis.

Our lipidomics analysis suggested there to be a moderate to heavy dysregulation in the global lipidome, which depends both on species of arsenic as well as the presence of an Ars operon. Both As^{III} treated groups (K12 and AW3110) generally cluster away from the others, potentially indicating these cultures are more stressed than their counterparts. The As^{III} groups from both strains appear to be responding to the metal differently, showing disparate lipid profiles in their top feature lists (Figure S2). This prompted a separate, intra-strain statistical comparison where K12 group were separated from AW3110.

The K12 comparison clusters mainly on treatment vs non-treatment, meaning As^{III} exposed cells are fairly similar to the As^{V} treatment in this strain. Both iAs treated K12 groups look very dissimilar to the control cultures. This could be evidence for an overall stress phenotype. It is also likely due to the ArsB mediated reduction of As^{V} to As^{III} so both groups are

exposed to some level of As^{III}. Despite their similarity in the global analysis, the K12 As^{III} and As^{V} groups look very different with respect to their top features. There is a very strong high abundance trend for the top features in the K12 As^{III} group (Figure 5). Some of these features have been tentatively matched with arsenic containing compounds. Our identification pipeline matched several of the top 40 features to structures from public databases. 11-oxo-undeca-5,6-dienoic acid and pentadecylic acid were shown to have a degree of downregulation in the arsenic treated groups relative to the control. Two glycoophospholipids were shown to have a marked increase in abundance in the As^{III} group.

All three AW3110 groups look different from one another, in contrast to the pattern observed in K12 where the treatment groups display some statistical overlap. Interestingly, As^{V} exposed cells appear more dissimilar to the control than those in the As^{III} group. In general, there is stronger groupwise clustering in the AW3110 conditions. There are also more observable regulation patterns in this strain, shown in the 7 different major clusters in the top 40 heatmap (Figure 4). Many more identifications were assigned in this group, including important molecules such as choline and dethiobiotin, as well as a couple potential arsenical species. Three glycero-phosphoethanolamines share the same abundance profile in their downregulation as a result of arsenic treatment. Choline is noticeably lower in abundance in As^{III} and As^{V} treated cells relative to the control. There is some evidence for interplay between choline and arsenic, with arsenocholine being a known arsenical found in food products (10). It is also key in the synthesis of glycine betaine, key osmolyte and methylation intermediate(27). It also has an arsenic analogue, arsenobetaine, which is a less toxic arsenical found in fish(20, 28). It remains unclear why choline is showing dysregulation in arsenic treated cells, but it is obvious that it plays a role

in the associated metabolism. Another key intermediate, bethiobiotin, shows interesting regulation pattern in AW3110 cells. It is completely absent in As^{III} treated cells, low in As^{V}, and highly abundant in control cells. This is crucial, since biotin limitation would affect many cellular metabolisms, including key central carbon processes(29–31). More efforts are needed to identify other intermediates in the discussed pathways namely; biotin and the molecule used as an amine donor (8-amino-7-oxononoate) for its synthesis, glycine betaine and betaine aldehyde, and the host of potential membrane lipids matched in this analysis.

Several potential arsenicals were identified in our analysis, but more work is underway to confirm their identifications. Two prime candidates (Figure 8) show complete absence in the control cultures and high abundance in arsenic treatments. The 400+ potential arsenicals identified in this project will first be assessed for the presence of arsenic using ICP-MS, after which targeted LC-MS/MS analyses will be performed to determine their structure more acutely. Key findings from this data include the potential presence of methylated arsenicals. *E.coli* and other organisms with the Ars operon do not (theoretically) contain the methylases used in other organism's to transform As^{III} into methylated species(32). The presence of methylated arsenicals would indicate the contrary, or that there is an alternative pathway for methylation in *E.coli*. LC-ICPMS experiments are underway to determine the valence state and speciation of the intracellular arsenic, expanding on the analysis shown in Figure 1.

Presented here is a foundation for understanding the interaction between iAs and *E.coli* as a proxy for elucidating GM mediated arsenic detoxification. Moving forward, our efforts will expand to identifying additional features through the use of chemical standards, and running full-dataset metabolic modelling. The expanded analysis will also include orthogonal approaches,

such as targeted LC-MS/MS in parallel with ICP-MS and LC-ICP-MS to confirm the identity of arsenicals proposed here. The future is bright for populations chronically exposed to high levels of arsenic; understanding how microbes handle the stress from and biotransform arsenic will enable new treatment options.

Methods

Cell culturing and harvesting:

Cultures of *E.coli* K12 and AW3110 were inoculated 1:100 from overnight cultures, growing at 37°C at 125RPM in an orbital shaker-incubator. All flasks (4 per condition) were allowed to grow for 2 hours prior to the addition of arsenic. Arsenic was added to concentrations dependent on As species as well as *E.coli* strain; K12 cultures were dosed with 2mM As^{V} and 1mM As^{III}, AW3110 were exposed to 0.25mM As^{V} and 0.1mM As^{III}. Arsenic was supplied as sodium hydrate salts (Thermo-Fisher). Cells were harvested mid log phase, 4 hours post arsenic addition. Cells were harvested through centrifugation (4,000G 15 minutes), washed with 50mM EDTA phosphate buffer to remove extracellular arsenic 3x through centrifugation and removal of supernatant (for ICP-MS samples).

Macromolecular As distribution:

Cells underwent a liquid/liquid extraction protocol paired with an acetone precipitation to determine the distribution of arsenic in each fraction. Cells were lysed via ultrasonication and centrifuged to separate soluble from insoluble lysate. The supernatant containing cytosolic small molecules and proteins were separated from one another using a cold acetone precipitation. Cold acetone (-20°C) cell lysate at a 5:1 v/v ratio. The supernatant was analyzed for arsenic content

separately from the resulting pellet. All three fractions underwent a 10% nitric acid digest to fully solubilize metals and dissolve any remaining solids. 10% nitric acid (HNO3) was added to each sample prior to a autoclave sterilization cycle (30 minute fast pressure release cycle). Resulting samples were diluted to a total of 5mL in 2%HNO_3/0.5%HCl, the minimum volume for the SPS 4 autosampler system. *Whole cell digests*: The above protocol was performed on non-fractionated biomass. After harvest, cells were directly digested in 10% HNO_3 and subjected to the same autoclave cycle and diluted to 5mL total volume.

ICPMS acquisition:

Data was acquired on an Agilent 7800 ICP-MS paired with an Agilent SPS 4 autosampler system. Using an established EPA acquisition method, arsenic alongside 18 other metals were quantified in each sample. As 75 was integrated for 1 second per scan, used He in the collision filtering cell, and used default torch temperature and gas flow settings. As 75 counts were normalized to Germanium 72, which was flowed at constant 0.5mL per minute as an internal standard.

MMC Extraction for untargeted lipidomics*:*

Lipid extractions followed a modified version of the MTBE/Methanol/Chloroform (MMC) extraction. Cell pellets were resuspended (mix by pipetting) in 1.3 ml of HPLC grade Methanol and transferred to a 7ml glass culture tubes held within a 15 ml falcon tube. From this point on, only glass pipettes were used. 1 ml of HPLC grade MTBE and 1 ml of HPLC grade CHCl3 (1:1, v/v/v) were then added, followed by a brief vortex. The samples were then incubated on a shaker at 225 rpm for 1hour at 22°C after which the samples were centrifuged at 17,500g for 10 min. 2.0 ml of the resulting supernatant is collected in a second set of culture

tubes inserted into 15ml falcon tubes and dried in a vacuum centrifuge for 1 h at 30 °C or until all liquid was evaporated. Samples were then stored at -80°C

LCMS acquisition:

LCMS acquisition was performed on a Waters Synapt XS paired with a Waters Acquity I class liquid separation system. The day of analysis, metabolomic preps were resuspended in 5:1 MeoH:H_2O and sonicated to fully solubilized. 5uL were injected into the autosampler system. LC separation was performed on a Waters Acquity BEH HILIC column. LC gradient separation used H_2O + 0.1% Formic Acid (FA) as solvent A and Acetonitrile +0.1% FA as solvent B. LC gradient; 0-1min 99%A, 1-22min 99%A-1%A, 22-25min 1%A, 25-25.1min 1%A-99%A, 25.1min-28min 99%A. Lipidomics was performed in negative mode, , 0.5s scan length, 50-1200Da m/z range, at 2.00kV capillary voltage, and 80°C source temperature.

Data processing and statistics:

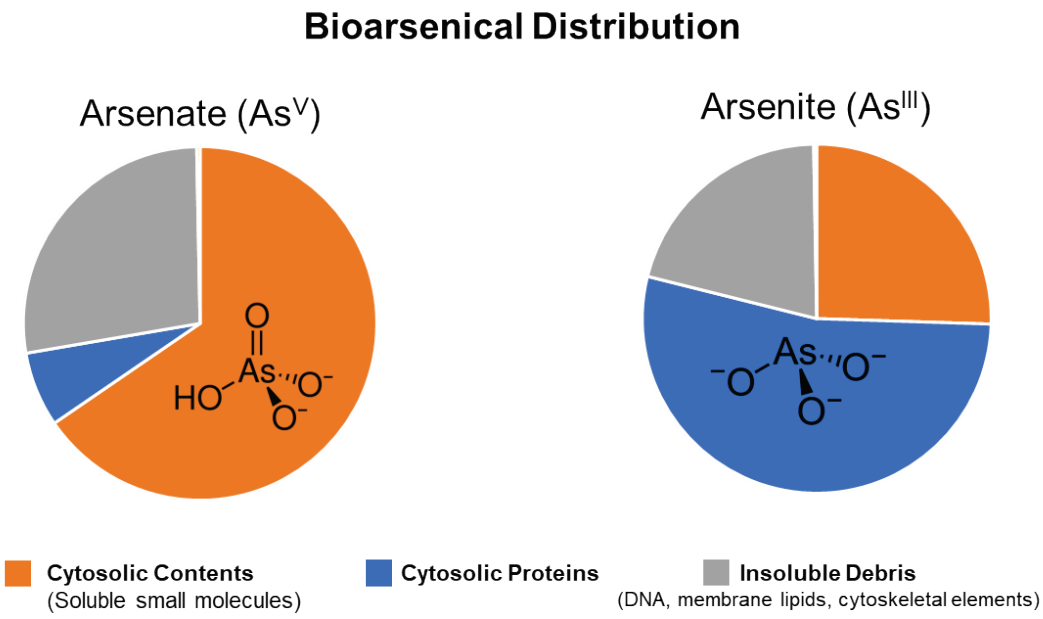
Raw data was imported into Nonlinear Dynamics (Waters corp) Progenesis QI for processing and assignment of identifications. The default peak picking algorithm was employed, retention times aligned to pooled samples, and all default adducts were allowed for analysis. After peak picking and alignment, CSV files were exported from Progenesis QI for statistical analysis in Metaboanalyst. Filtering and normalization took place in Metaboanalyst as follows; filtered to the top 2500 features based on interquartile range, quantile normalized, log transformed, and auto-scaled.

Identification pipeline:

Tentative identifications were assigned through the use of Progenesis Metascope, an in-silico structural database, theoretical fragmentation based searching algorithm. Chemical structures are acquired from metacyc, lipidmaps, and lipidblast online databases. Features are also compared to our curated database of theoretical bioarsenical structures. Measured M/Z features are first assessed for all potential chemical formulae using M/z value and observed isotopic distribution, within 20ppm. Measured mz formulae are then matched to formulae from the database, and any matches undergo theoretical fragmentation to match observed MS/MS peaks with predicted fragmentation patterns. Identifications are assigned a score based on the cumulation of all of these parameters. All identifications reported in this document have a high score based on isotopic distribution and PPM error, and many also include matches to MS/MS fragments.

Figures and Tables

A



B

Treatment	Cellular arsenic (ug)	% Recovery	Arsenic bioaccumulation
Arsenate (AsV)	24.68	92%	0.113%
Arsenite (AsIII)	17.45	88%	0.119%

Figure 1: ICP-MS quantification of As in K12 *E.coli* macromolecular fractions. A: Distribution of arsenic by fraction- cytosolic contents, proteincs, or insoluble debris. As^{V} (left) has 65% of quantified As in the cytosolic contents, 7.5% in proteins, and 27.4% in the insoluble debris. As^{III} (right) has 25.2% in its soluble cytosol, 52.7% in proteins, and 22.1% in the insoluble debris. B: Table of overall arsenical quantification from whole cell (unfractionated) analyses. Cellular arsenic is the quantified total in ug, % recovery represents the sum of the fractions from panel A compared to the cellular arsenic, and Arsenic bioaccumulation is the amount of arsenic quantified in cellular biomass compared to all available arsenic in the media, by mass.

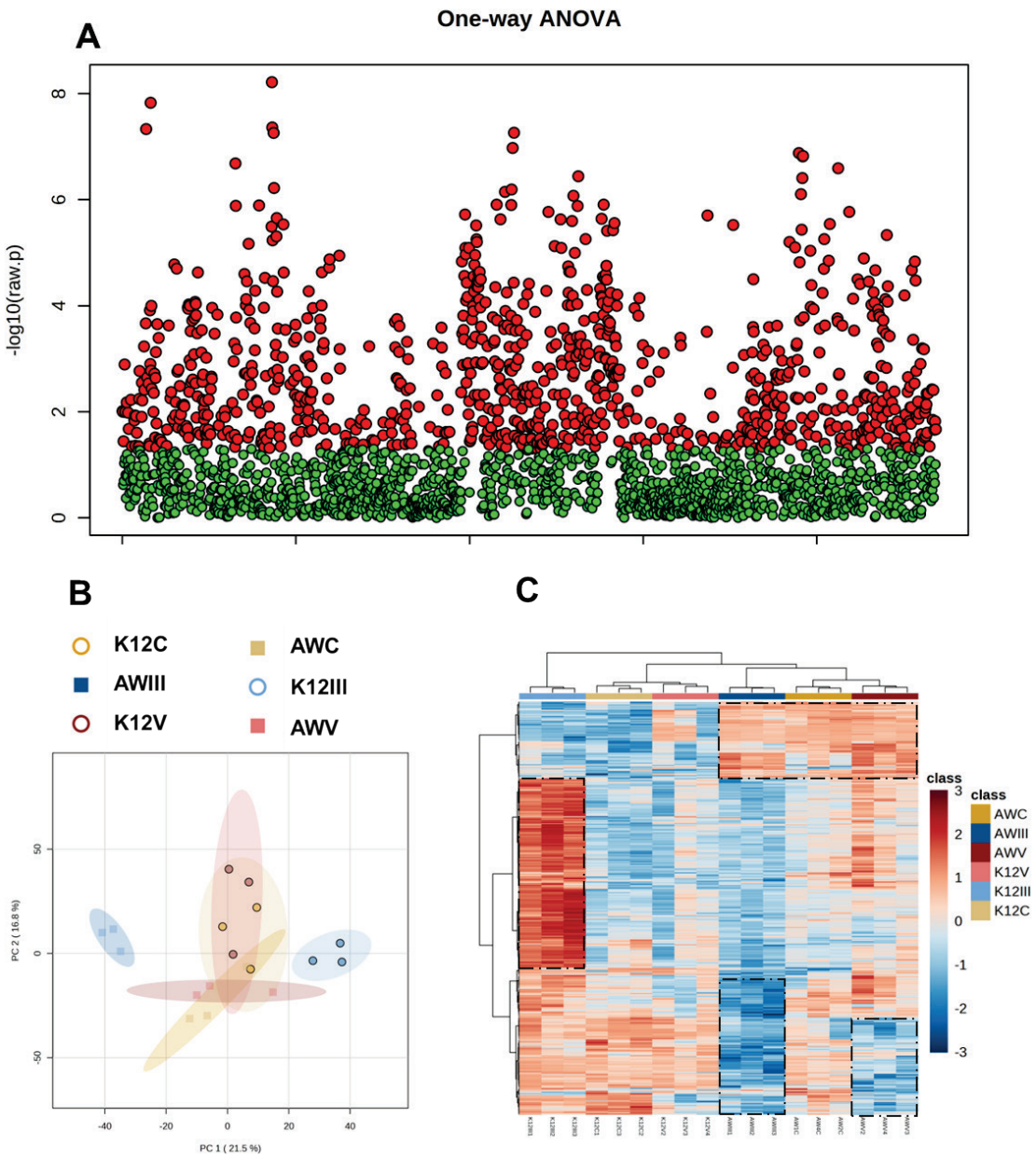


Figure 2: All group global lipidomics multivariate analyses. A: one-way ANOVA of the filtered dataset (2500 features). 935 show statistical significance (p-value <0.05)(Red). B: principal component analysis (PCA) including all groups. Each point represents a culture of *E.coli* .Ovals represent 95% confidence intervals. C: Parallel hierarchical clustering (PHCA) heatmap. Features are clustered on y axis, samples are clustered on x axis. Columns represent individual cultures of E.coli and are colored by group. Cluster boxes of interest are marked with dashed lines. AWC, AWIII, AWV correspond to Aw3110 strain exposed to vehicle, As^{III}, and As^{V} respectively.

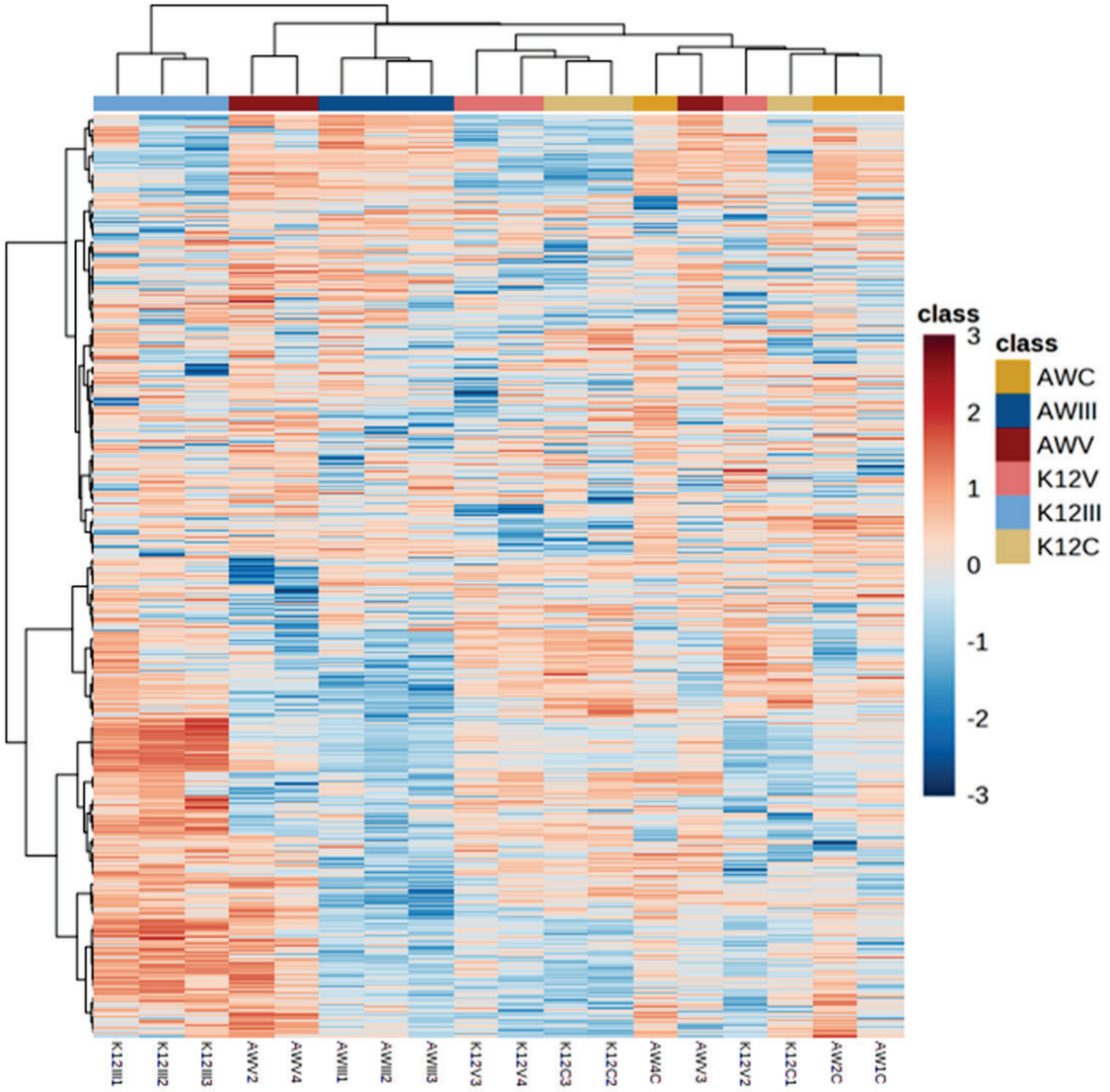


Figure S1: All feature PHCA heatmap. Features are clustered on y axis, samples are clustered on x axis. Columns represent individual cultures of E.coli and are colored by group. AWC, AWIII, AWV correspond to Aw3110 strain exposed to vehicle, As^{III}, and As^{V} respectively. K12C, K12III, K12V correspond to K12 strain exposed to vehicle, As^{III}, and As^{V} respectively.

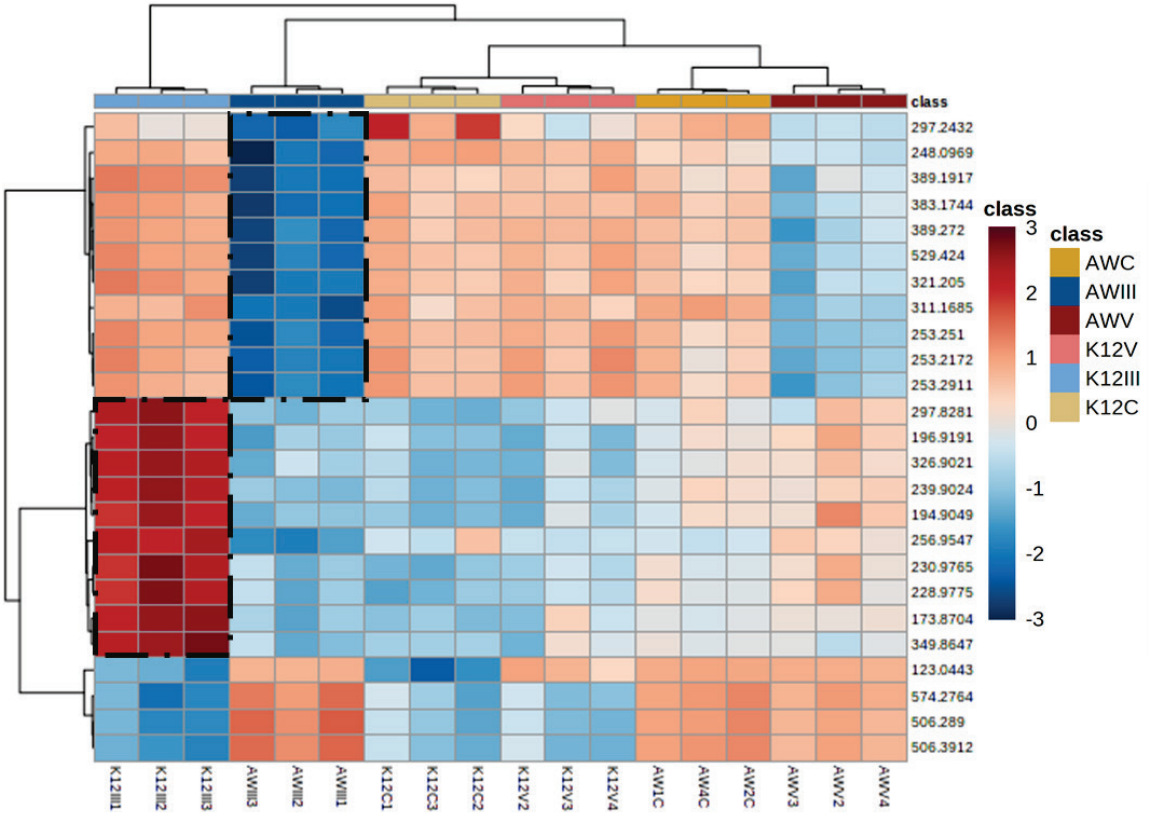


Figure S2: Top 20 (filtered by p-value) PCHA heatmap. Features are clustered on y axis, samples are clustered on x axis. Columns repreent individual cultures of E.coli and are colored by group. Clusters of interest are marked with dashed lines. AWC, AWIII, AWV correspond to Aw3110 strain exposed to vehicle, As^{III}, and As^{V} respectively. K12C, K12III, K12V correspond to K12 strain exposed to vehicle, As^{III}, and As^{V} respectively.

MZ Feature	ID Accession	Common name
297.2432	LMFA02000261	17-oxo-octadecanoic acid
248.0969	07M-71DS1g0pqQq	Dethiobiotin
389.1917	LMST02030192	11-Dehydrocorticosterone
383.1914	NID	
389.272	LMFA01140076	17-phenyl-13E-heptadecenoic acid
529.424	LMFA07090049	FAHFA(16:1/5-O-16:0)
321.205	NID	
311.1685	LMFA01100051	2,3-diamino-propionic acid
253.251	LMFA05000203	14-Methyl-8Z-hexadecen-1-ol
253.2172	LMFA01080013	9-methoxy-pentadecanoic acid
253.2911	LMFA11000581	Octadecane
297.8281	NID	
196.9191	NID	
326.9021	NID	
239.9024	NID	
194.9049	NID	
256.9547		5-(dimethylarsoryl)pent-2-enoic acid
230.9765	NID	
228.9775		butyldimethylarsine sulfide
173.8704	NID	
349.8647	NID	
123.0443	bVFUEdIHDrt9OKH	4-hydroxybenzyl alchohol
574.2764	LMGP20020033	PS-PE
506.289	GPEtn(17:1/2:0)	GPEtn(17:1/2:0)
506.3912	NID	

Table 1: Assigned identities to the features from the top 20 heatmap

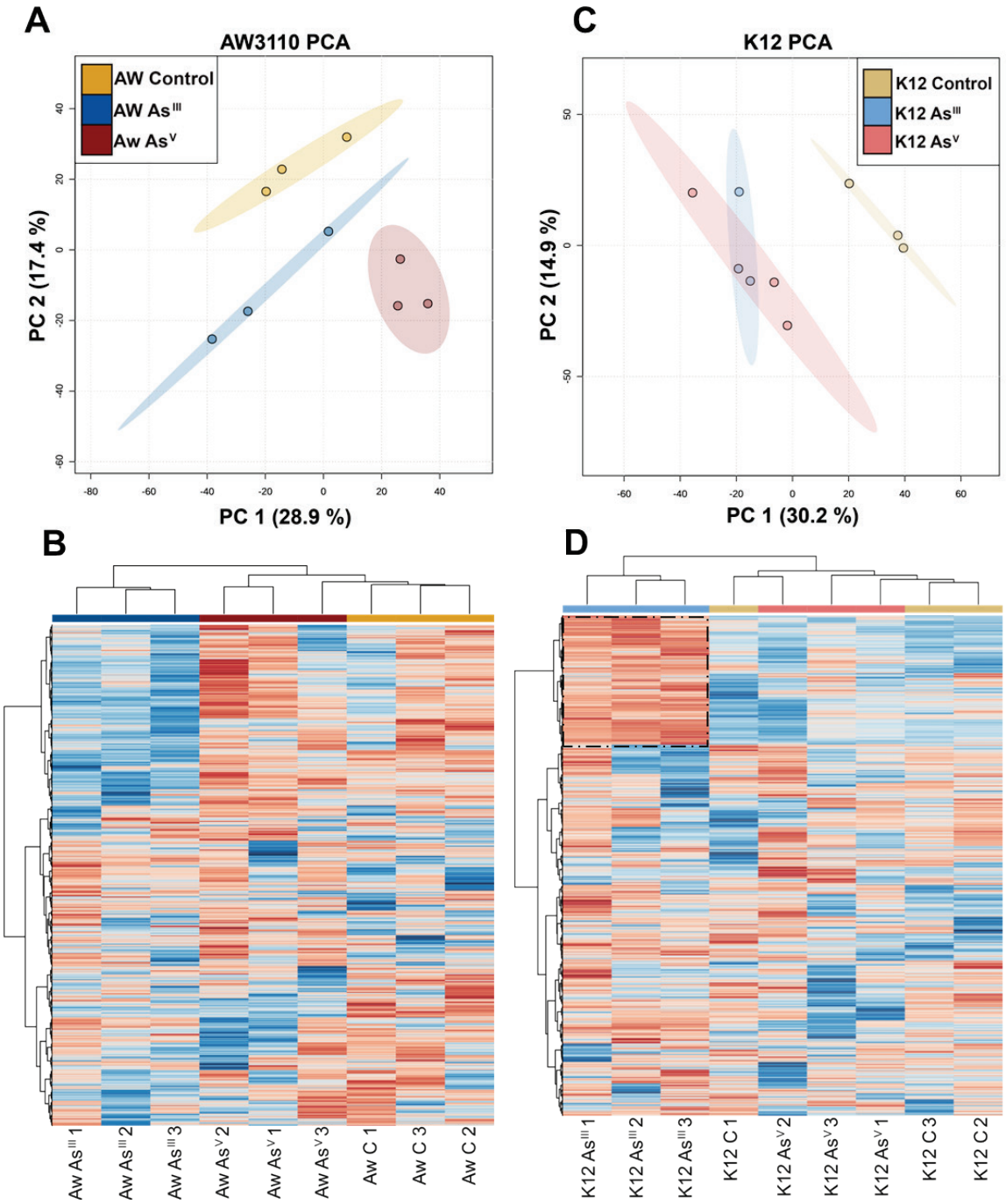


Figure 3: Intra-strain global lipidomic multivariate analyses. A: AW3110 strain PCA. Each point represents a culture of *E.*coli. Ovals represent 95% confidence intervals. B: PHCA heatmap of AW3110. C: K12 strain PCA. Each point represents a culture of *E.*coli. Ovals represent 95% confidence intervals. D: PHCA of K12 cultures. AWC, AWIII, AWV correspond to Aw3110 strain exposed to vehicle, As^{III}, and As^{V} respectively. K12C, K12III, K12V correspond to K12 strain exposed to vehicle, As^{III}, and As^{V} respectively

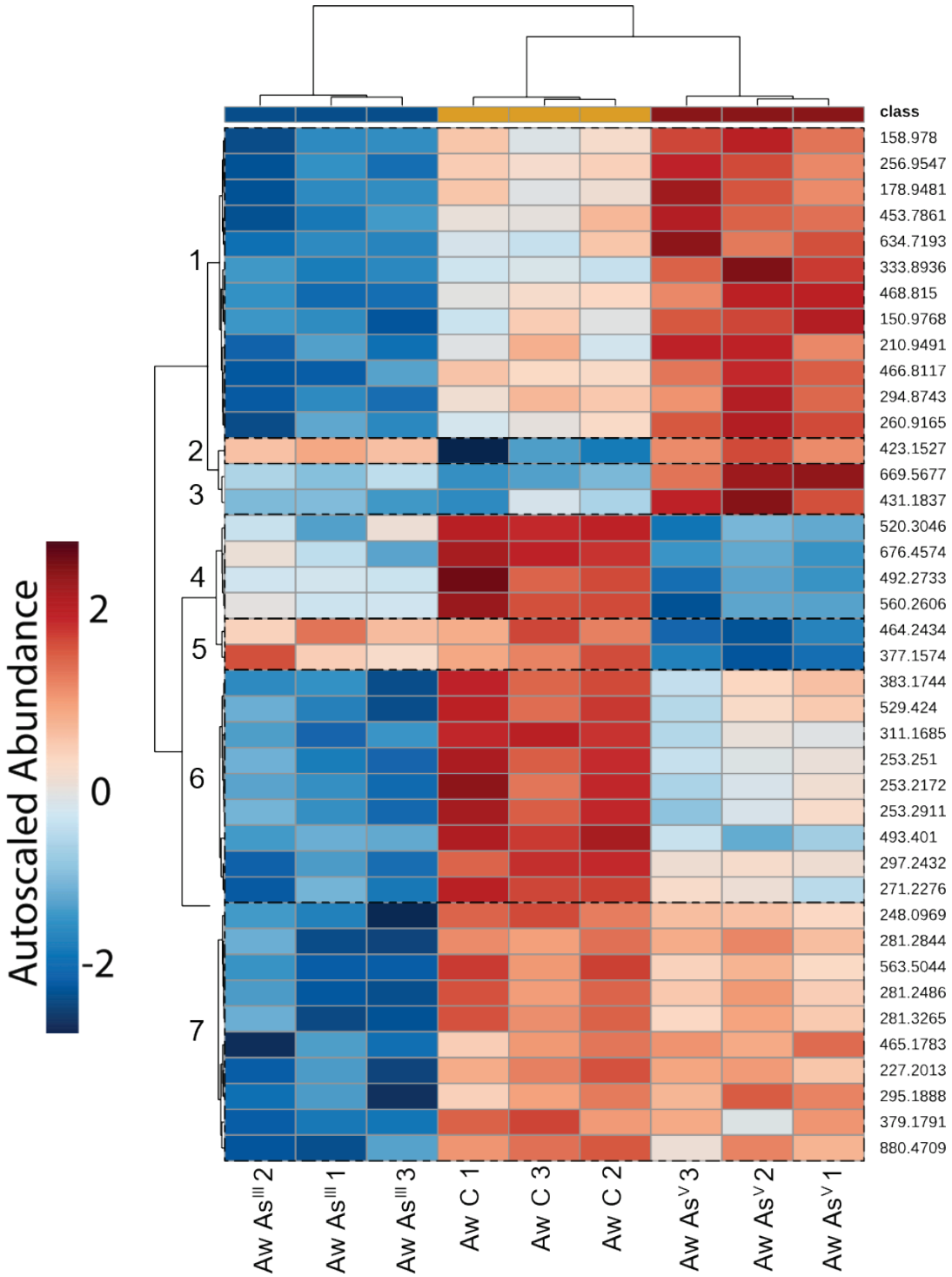


Figure 4: Top 40 (filtered by p-value) heatmap for the AW3110 intra-strain comparison. Clusters are broken up by regulation pattern and correspond to Table 2. AWC, AWIII, AWV correspond to Aw3110 strain exposed to vehicle, As^{III}, and As^{V} respectively.

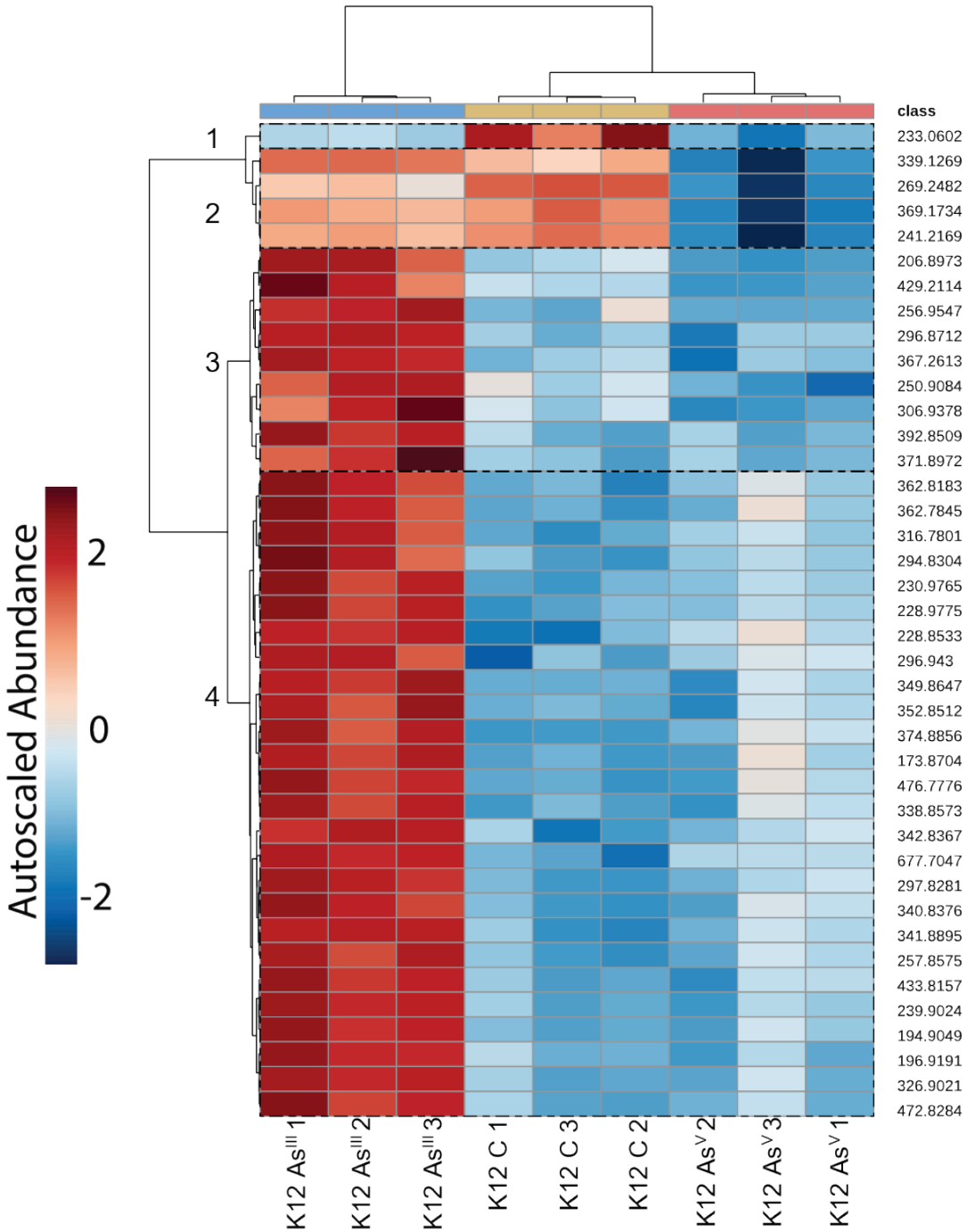


Figure 5: Top 40 (filtered by p-value) heatmap for the K12 intra-strain comparison. Clusters are broken up by regulation pattern and correspond to Table 2. K12C, K12III, K12V correspond to K12 strain exposed to vehicle, As^{III}, and As^{V} respectively.

	Aw3110 top 40	
	MZ	Identification
1	158.978	Not Identified
	256.9547	**5-(dimethylarsoryl)pent-2-enoic acid**
	178.9481	Not Identified
	453.7861	Not Identified
	634.7193	Not Identified
	333.8936	Not Identified
	468.815	Not Identified
	150.9768	Not Identified
	210.9491	**1buta-1,3-dienyldimethylarsine oxide**
	466.8117	Not Identified
	294.8743	Not Identified
	260.9165	Not Identified
2	423.1527	Not Identified
3	669.5677	**DG(18:0/18:0/0:0)**
	431.1837	delta5-7-NeuroF
4	520.3046	**GPEtn(17:1/3:0)**
	676.4574	Not Identified
	492.2733	**GPEtn(14:1/4:0)**
	560.2606	**GPEtn(18:4/2:0)**
5	464.2434	**GPEtn(14:1/2:0)**
	377.1574	Not Identified
6	383.1744	Not Identified
	529.424	**FAHFA(16:1/5-O-16:0)**
	311.1685	**2,3-diamino-propionic acid**
	253.251	**14-methyl-8-hexadecen-1-ol**
	253.2172	**Choline**
	253.2911	**7-Methylheptadecane**
	493.401	**32:04:00**
	297.2432	**8-hydroxyoctadeceneic acid**
	271.2276	**2-hydroxy palmitic acid**
7	248.0969	**Dethiobiotin**
	281.2844	**iso-1,2-nonadecanediol**
	563.5044	Not Identified
	281.2486	**2-methyl-16-heptadecenoic acid**
	281.3265	Not Identified
	465.1783	**6-bromo-tricosa-5E,9Z-dienoic acid**
	227.2013	**4,10-dimethyldodecanoic acid**
	295.1888	**2,15-dihydroxy-pentadecylic acid**
	379.1791	Not Identified
	880.4709	Not Identified

	K12 top 40	
	MZ	Identification
1	233.0602	**11-oxo-undeca-5,8-dienoic acid**
2	339.1269	**2,3-dihydroxy-3-methylpentanoate**
	269.2482	**2-methyl-hexadecanoic acid**
	369.1734	Not Identified
	241.2169	**Pentadecylic acid**
3	206.8973	Not Identified
	429.2114	**PA(18:4/0:0)**
	256.9547	**5-(dimethylarsoryl)pent-2-enoic acid**
	296.8712	Not Identified
	367.2613	Not Identified
	250.9084	Not Identified
	306.9378	Not Identified
	392.8509	Not Identified
	371.8972	Not Identified
4	362.8183	Not Identified
	362.7845	Not Identified
	316.7801	Not Identified
	294.8304	Not Identified
	230.9765	Not Identified
	228.9775	**butyldimethylarsine sulfide**
	228.8533	Not Identified
	296.943	Not Identified
	349.8647	Not Identified
	352.8512	Not Identified
	374.8856	Not Identified
	173.8704	Not Identified
	476.7776	Not Identified
	338.8573	Not Identified
	342.8367	Not Identified
	677.7047	Not Identified
	297.8281	Not Identified
	340.8376	Not Identified
	341.8895	Not Identified
	257.8575	Not Identified
	433.8157	Not Identified
	239.9024	Not Identified
	194.9049	Not Identified
	196.9191	Not Identified
	326.9021	Not Identified
	472.8284	Not Identified

Table 2: Assigned identities from top 40 heatmaps (Aw3110 & K12). Rows are colored corresponding to their cluster from the PHCA heatmap

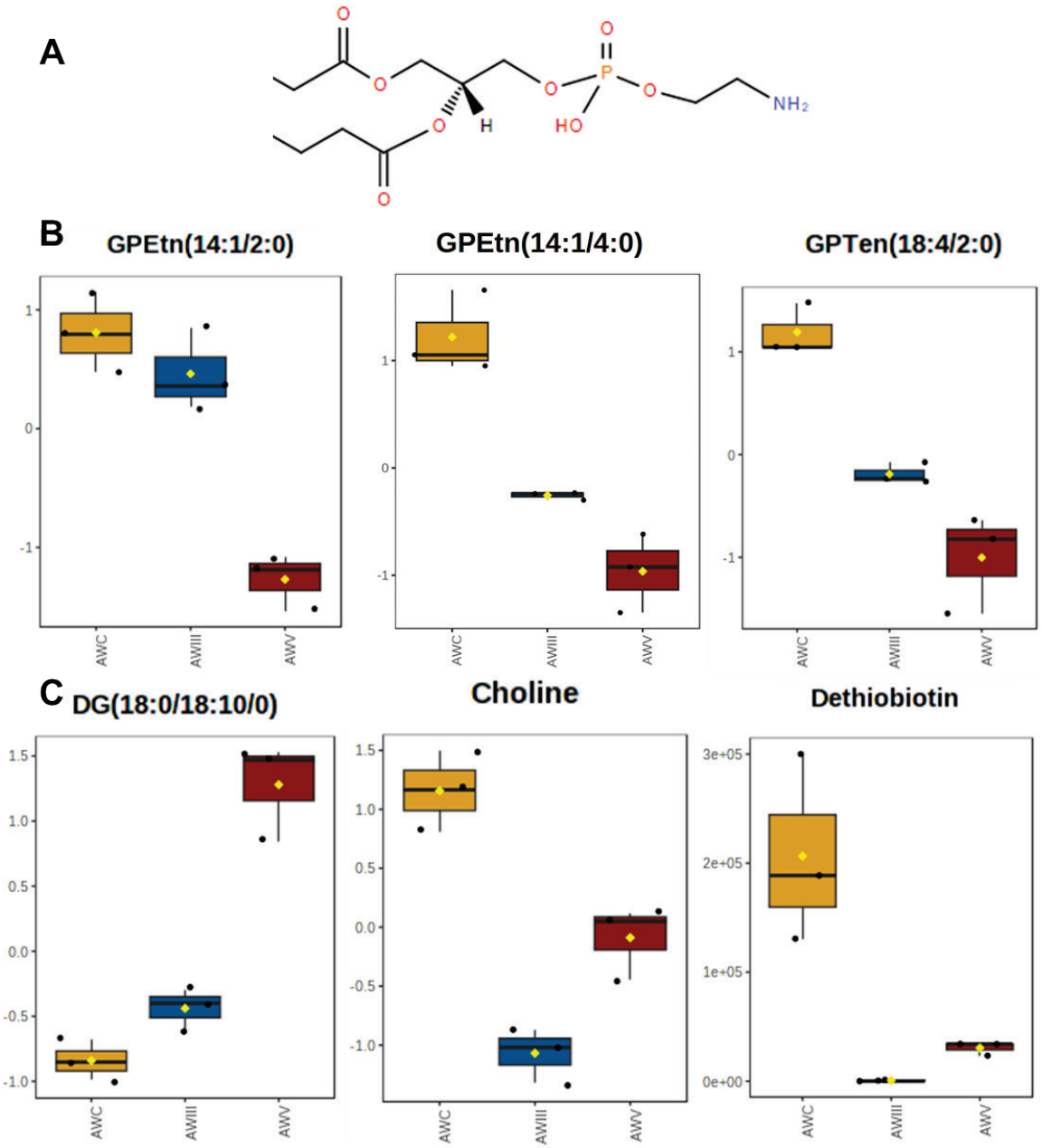


Figure 6: Dysregulated features in the AW3110 comparison. A: General structure of glycerophospoethanolamine lipids. Structure is truncated to highlight headgroup. B: Box and whisker plots of three GPEtn lpids. Y axis is scaled and normalized abundance. C: Box and whisker plots of three additional features of interest. Y axis is scaled and normalized abundance for DG(18:0/18:10/0) and choline, Dethiobiotin is shown as raw abundance to highlight its absence in AWIII. AWC, AWIII, AWV correspond to Aw3110 strain exposed to vehicle, As^{III}, and As^{V} respectively.

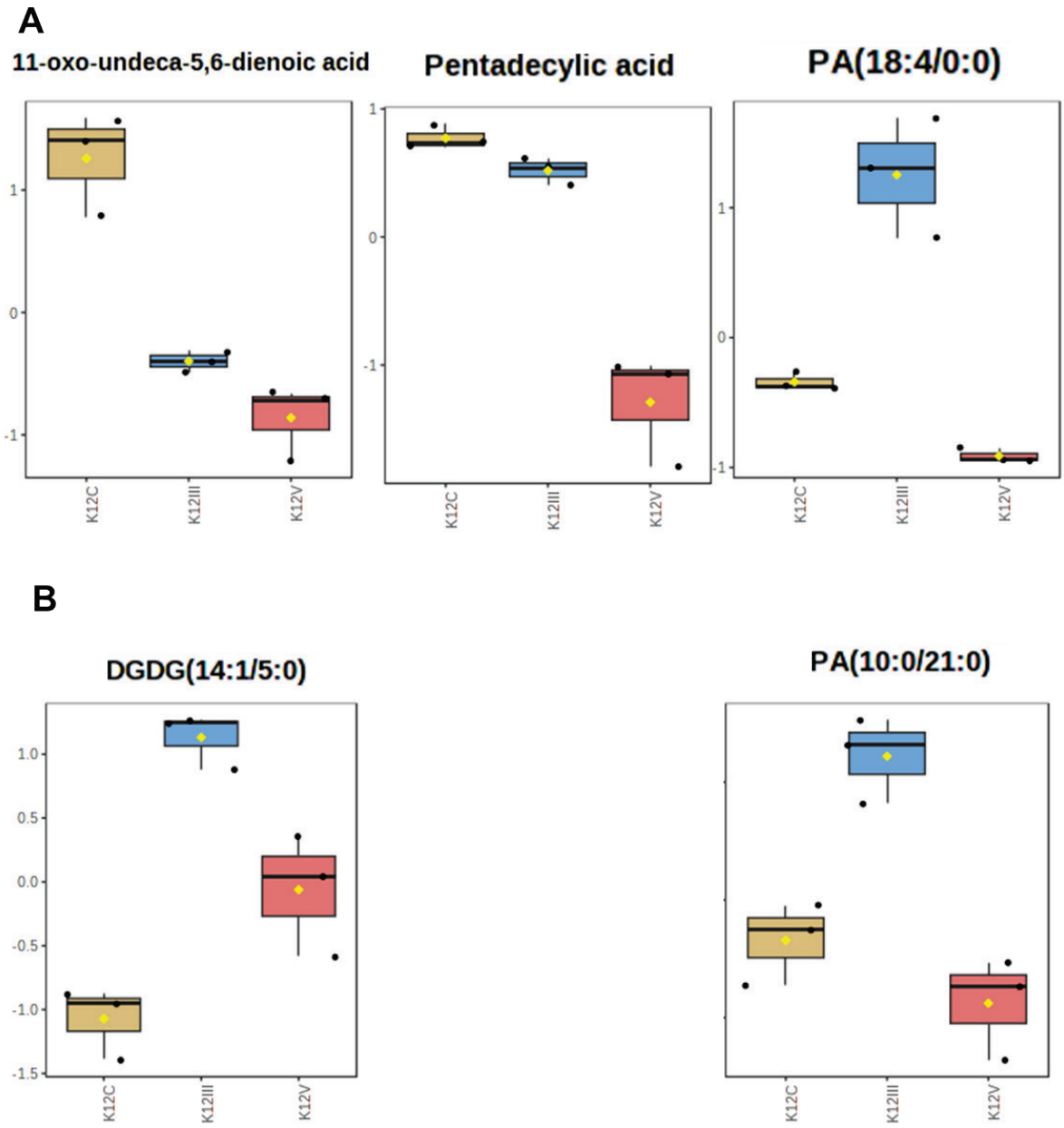


Figure 7: Dysregulated features in the K12 comparison. A. Structure is truncated to highlight headgroup. B: Box and whisker plots of three moderate length lipids from the PHCA heatmap. Y axis is scaled and normalized abundance. C: Box and whisker plots of two additional features of interest not found in the top 40 features (filtered by p-value). K12C, K12III, K12V correspond to K12 strain exposed to vehicle, As^{III}, and As^{V} respectively.

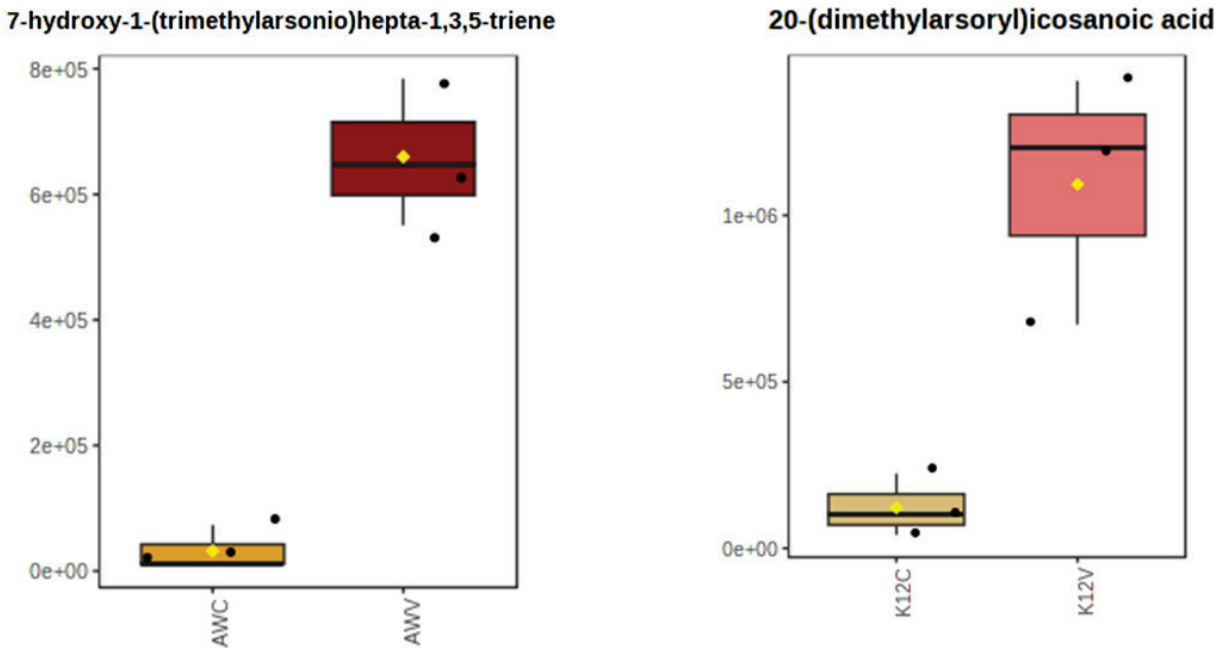


Figure 8: Potential bio-arsenical features identified. Box and whisker plots show raw abundance in each group. 7-hydroxy-1-(trimethylarsonio)hepta-1,3,5-triene was only found in the AW3110 comparison and 20-(dimethylarsoryl)icosanoic acid was only found in the K12 comparison. K12C, K12III, K12V correspond to K12 strain exposed to vehicle, As^{III}, and As^{V} respectively. AWC, AWIII, AWV correspond to Aw3110 strain exposed to vehicle, As^{III}, and As^{V} respectively.

References:

1. Science | AAAS. https://www.science.org/journal/science/no-first-release. Retrieved 14 April 2023.

2. Podgorski J, Berg M. 2020. Global threat of arsenic in groundwater. Science (1979) 368:845–850.

3. Garbinski LD, Rosen BP, Chen J. 2019. Pathways of arsenic uptake and efflux. Environ Int 126:585–597.

4. Palma-Lara I, Martínez-Castillo M, Quintana-Pérez JC, Arellano-Mendoza MG, Tamay-Cach F, Valenzuela-Limón OL, García-Montalvo EA, Hernández-Zavala A. 2020. Arsenic exposure: A public health problem leading to several cancers. Regulatory Toxicology and Pharmacology 110:104539.

5. Barrett JR. 2017. Rice versus drinking water: Estimating the primary source of arsenic in the U.S. Diet. Environ Health Perspect 125.

6. Argos M, Ahsan H, Graziano JH. 2012. Arsenic and human health: Epidemiologic progress and public health implications. Rev Environ Health 27:191–195.

7. Giles BH, Mann KK. 2022. Arsenic as an immunotoxicant. Toxicol Appl Pharmacol 454:116248.

8. Grau-Perez M, Kuo CC, Gribble MO, Balakrishnan P, Spratlen MJ, Vaidya D, Francesconi KA, Goessler W, Guallar E, Silbergeld EK, Umans JG, Best LG, Lee ET, Howard B V., Cole SA, Navas-Acien A. 2017. Association of low-moderate arsenic exposure and arsenic metabolism with incident diabetes and insulin resistance in the strong heart family study. Environ Health Perspect 125.

9. Cohen SM, Arnold LL, Eldan M, Lewis AS, Beck BD. 2008. Methylated Arsenicals: The Implications of Metabolism and Carcinogenicity Studies in Rodents to Human Risk Assessment. http://dx.doi.org/101080/10408440500534230 36:99–133.

10. Delaney SR, Nandakumar V, Jannetto PJ. 2020. Trust your gut: a case of persistent gastrointestinal disturbances. Toxicology Cases for the Clinical and Forensic Laboratory 433–435.

11. Luvonga C, Rimmer CA, Yu LL, Lee SB. 2020. Organoarsenicals in Seafood: Occurrence, Dietary Exposure, Toxicity, and Risk Assessment Considerations - A Review. J Agric Food Chem 68:943–960.

12. Ghaisas S, Maher J, Kanthasamy A. 2016. Gut microbiome in health and disease: Linking the microbiome–gut–brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. Pharmacol Ther 158:52–62.

13. Dong X, Shulzhenko N, Lemaitre J, Greer RL, Peremyslova K, Quamruzzaman Q, Rahman M, Ibn Hasan OS, Joya SA, Golam M, Christiani DC, Morgun A, Kile ML. 2017. Arsenic exposure and intestinal microbiota in children from Sirajdikhan, Bangladesh. PLoS One 12.

14. Fukuda S, Terasawa M, Shiomi K. 2011. Phosphatidylarsenocholine, one of the major arsenolipids in marine organisms: Synthesis and metabolism in mice. Food and Chemical Toxicology 49:1598–1603.

15. Rawle R, Saley TC, Kang YS, Wang Q, Walk S, Bothner B, McDermott TR. 2021. Introducing the ArsR-Regulated Arsenic Stimulon. Front Microbiol 12:162.

16. Tam LM, Price NE, Wang Y. 2020. Molecular Mechanisms of Arsenic-Induced Disruption of DNA Repair. Chem Res Toxicol 33:709–726.

17. Zhao FJ. 2016. A novel pathway of arsenate detoxification. Mol Microbiol 100:928–930.

18. Keren R, Méheust R, Santini JM, Thomas A, West-Roberts J, Banfield JF, Alvarez-Cohen L. 2022. Global genomic analysis of microbial biotransformation of arsenic highlights the importance of arsenic methylation in environmental and human microbiomes. Comput Struct Biotechnol J 20:559–572.

19. Coryell M, McAlpine M, Pinkham N V., McDermott TR, Walk ST. 2018. The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. Nat Commun 9.

20. Cullen WR, Reimer KJ. 1989. Arsenic Speciation in the Environment. Chem Rev 89:713–764.

21. Rawle RA, Tokmina-Lukaszewska M, Shi Z, Kang YS, Tripet BP, Dang F, Wang G, McDermott TR, Copie V, Bothner B. 2020. Metabolic Responses to Arsenite Exposure Regulated through Histidine Kinases PhoR and AioS in Agrobacterium tumefaciens 5A. Microorganisms 2020, Vol 8, Page 1339 8:1339.

22. Carlin A, Shi W, Dey S, Rosen BP. 1995. The ars operon of Escherichia coli confers arsenical and antimonial resistance. J Bacteriol 177:981–986.

23. Komorowicz I, Sajnóg A, Barałkiewicz D. 2019. Total Arsenic and Arsenic Species Determination in Freshwater Fish by ICP-DRC-MS and HPLC/ICP-DRC-MS Techniques. Molecules 24.

24. Luvonga C, Rimmer CA, Yu LL, Lee SB. 2020. Organoarsenicals in Seafood: Occurrence, Dietary Exposure, Toxicity, and Risk Assessment Considerations - A Review. J Agric Food Chem 68:943–960.

25. Rawle RA, Kang YS, Bothner B, Wang G, McDermott TR. 2019. Transcriptomics analysis defines global cellular response of Agrobacterium tumefaciens 5A to arsenite exposure regulated through the histidine kinases PhoR and AioS. Environ Microbiol 21:2659–2676.

26. Parvez F, Wasserman GA, Factor-Litvak P, Liu X, Slavkovich V, Siddique AB, Sultana R, Sultana R, Islam T, Levy D, Mey JL, van Geen A, Khan K, Kline J, Ahsan H, Graziano JH. 2011. Arsenic Exposure and Motor Function among Children in Bangladesh. Environ Health Perspect 119:1665.

27. Kempf B, Bremer E. 1998. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Arch Microbiol 170:319–330.

28. Gaion A, Sartori D, Scuderi A, Fattorini D. 2014. Bioaccumulation and biotransformation of arsenic compounds in Hediste diversicolor (Muller 1776) after exposure to spiked sediments. Environmental Science and Pollution Research 21:5952–5959.

29. Cramer JD, Jarrett JT. 2018. Purification, Characterization, and Biochemical Assays of Biotin Synthase From Escherichia coli. Methods Enzymol 606:363–388.

30. Satiaputra J, Shearwin KE, Booker GW, Polyak SW. 2016. Mechanisms of biotin-regulated gene expression in microbes. Synth Syst Biotechnol 1:17–24.

31. Waldrop GL, Holden HM, Maurice MS. 2012. The enzymes of biotin dependent CO2 metabolism: What structures reveal about their reaction mechanisms. Protein Science 21:1597–1619.

32. Chen J, Rosen BP. 2020. The Arsenic Methylation Cycle: How Microbial Communities Adapted Methylarsenicals for Use as Weapons in the Continuing War for Dominance. Front Environ Sci 8:43.

CONCLUDING REMARKS

This dissertation outlines two cases of systems biology omics investigations. One project involves a challenging, non-traditional system, the other investigates a classic model organism. These two examples showcase how omics can be used to study a system holistically, tracking many of the variables that bring about life. They also highlight the most powerful application, in the opinion of the author, of modern omics – the generation of new, directly testable hypotheses. Hypothesis generation can be one of the most difficult aspects of science (especially in relatively uncharacterized systems) and many would argue it to be the most important step in the scientific method. The studies presented here test some hypotheses, however, while doing so generate a much larger number of new questions and directly testable hypotheses.

The methanogen *M.barkeri* holds great promise, industrially speaking, for its ability to extract iron and sulfur from mineral surfaces. The exact mechanism utilized to do so remains unknown, but the work outlined here suggests it to be much more complex than previously thought. Going into this project, we hypothesized the presence of a conserved, concerted mechanism used to reduce the metal surface, import the resulting products, and process them into the various biologically relevant Fe/S containing molecules. At present, an exact answer to this hypothesis remains elusive. The multi-omics analysis does suggest, however, that the cells undergo a phenotypic shift in order to facilitate this process. There is likely specific machinery that are carrying out the consumption of mineral surfaces, but this data imply that the related processes are i) relatively biochemically expensive, ii) spread out across many different pathway families (central carbon, amino acid biosynthesis, etc.), and/or iii) require extensive cellular remodeling to proceed. While this does not elucidate the entirety of the mechanism on its own, it

provides a myriad of ways to directly probe the mechanism that were not available previously. For example- dozens of uncharacterized proteins change in response to the presence of the mineral, prompting an ongoing bioinformatic investigation (and eventual knockout studies/expression trials/enzymatic assays). Many other hypotheses have emerged as a result of the metabolomics/lipidomics as well: there appears to be a disconnect with form of stored S between mineral grown cells and those in soluble Fe/S. Specifically, cells grown on FeS_2 and the analogous HS group appear to be accumulating S in sulfoxide form- this observation needs to be confirmed and expanded on further through direct investigation. The most important next step for this project lies in the true statistical multi-omics integration of all the different data sets. This will require additional experiments to identify the metabolites and lipids that are regulated, through the use of targeted LC-MS/MS experiments. Once more features from the dataset have confirmed identifications, the project can move towards a true systems scale interpretation. Analyses like this are more than just cross-dataset interpolation as has been done to this point; direct correlations are drawn between the regulatory patterns in the transcripts, proteins, metabolites, and lipids to allow for true metabolic modeling.

While many of the mechanisms underlying microbial arsenic detoxification are characterized, how arsenic is incorporated into biomolecules and the nature of those molecules is not clearly understood. *Escherichia coli* and other organisms in the human gut microbiome play a large part in the dangerous effects of arsenic on human health. The data presented here suggests that arsenic, depending on the form present, distributes among different classes of molecules; some ends up in proteins, some remains soluble in the cytosol, and some associated with nonpolar molecules like lipids. Efforts are ongoing to characterize this distribution fully, but

some potential bio-arsenicals that have not been previously documented were identified in the analysis. Interestingly, these appear to be methylated arsenicals, which are not able to be synthesized by any (characterized) proteins present in *E. coli*. More than that, the *E.coli* lipidome as a whole appears to change in response to arsenic, again depending on the type of arsenic as well as the presence of the machinery to detoxify it. As with the analyses above, this serves as an excellent foundation for more specific, testable hypotheses and further analysis. Targeted LC-MS/MS analyses to identify key features observed as changing significantly will provide additional biological inferences, and will also allow for bioinformatic pathway enrichment analyses designed specifically for *E.coli*. This analysis will unlock future avenues for probing the gut microbiome community response to arsenic, as the K12 strain is a proxy for *E.coli* present in the gut and AW3110 can serve as a representative for organisms lacking the Ars operon.

All in all, these projects present excellent case studies for the usefulness of systems biology omics-based scientific inquiry. They also highlight some of the shortcomings of this approach, namely the difficulty of extracting pertinent biological inferences. As time progresses, and the techniques continue to mature, these shortcomings will inevitably disappear as scientists get better at looking at comprehensive analyses. This is because the pitfalls discussed here ultimately emerge from the *overabundance* of data. When a biological system is probed in a comprehensive manner, the results are inherently comprehensive! This creates a figurative haystack of data in which pertinent information can sometimes be a needle. Efforts to mitigate this, as seen in the dozens of published bioinformatic processing tools and software, are well underway, but support for less characterized systems like *M.barkeri* is still lacking. The final

frontier in systems biology, therefore, lies in computation. The technology exists to acquire the necessary data, the community only needs to be able to fully interpret it.

Despite the current shortcomings, the future of systems biology, and in turn science as a whole, is very bright. In the coming decades, this school of thought will propel the human race to previously inconceivable innovations in medicine, microbiology, and all things biological. Once fully fleshed out, scientists will finally start to match the scale of complexity of biological systems with nuance of their own. Granularity will continue to improve to subcellular resolution, and also expand beyond the individual cellular system. To this point the "system" has remained intracellular, but in reality, life exists in additional hierarchical levels far above this. Multicellular assemblies, biofilms, organs, all the way up to a system like the human body are perfect for systems biology omics investigations. Imagine a world where colorectal cancer prognosis is linked to changes in post translational modifications found on proteins in muscular tissue. Or biofilms growing on catheters and at the bottom of sewage sludge can be fully assessed, including each of their constituents individually and all subsequent interactions. All living organisms, and the chemistry contained therein, function together as a system. It makes sense, then, to assess them from a systemic perspective through the use of modern omics technologies.

REFERENCES CITED

Ajees, A. A., & Rosen, B. P. (2015). As(III) S-Adenosylmethionine Methyltransferases and Other Arsenic Binding Proteins. Https://Doi.Org/10.1080/01490451.2014.908983, 32(7), 570–576. https://doi.org/10.1080/01490451.2014.908983

Amenabar, M. J., & Boyd, E. S. (2018). Mechanisms of mineral substrate acquisition in a thermoacidophile. Applied and Environmental Microbiology, 84(12). https://doi.org/10.1128/AEM.00334-18

Amenabar, M. J., & Boyd, E. S. (2019a). A review of the mechanisms of mineral-based metabolism in early Earth analog rock-hosted hydrothermal ecosystems. World Journal of Microbiology and Biotechnology, 35(2), 29. https://doi.org/10.1007/s11274-019-2604-2

Amenabar, M. J., & Boyd, E. S. (2019b). A review of the mechanisms of mineral-based metabolism in early Earth analog rock-hosted hydrothermal ecosystems. World Journal of Microbiology & Biotechnology, 35(2). https://doi.org/10.1007/S11274-019-2604-2

Archibald, F. (1983). Lactobacillus plantarum, an organism not requiring iron. FEMS Microbiology Letters, 19(1), 29–32. https://doi.org/10.1111/j.1574-6968.1983.tb00504.x

Argos, M., Ahsan, H., & Graziano, J. H. (2012). Arsenic and human health: Epidemiologic progress and public health implications. Reviews on Environmental Health, 27(4), 191–195. https://doi.org/10.1515/REVEH-2012-0021/MACHINEREADABLECITATION/RIS

Ayala-Castro, C., Saini, A., & Outten, F. W. (2008). Fe-S cluster assembly pathways in bacteria. Microbiology and Molecular Biology Reviews : MMBR, 72(1), 110–125, table of contents. https://doi.org/10.1128/MMBR.00034-07

Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R., & Wolfe, R. S. (1979). Methanogens: reevaluation of a unique biological group. Microbiological Reviews, 43(2), 260–296. https://doi.org/10.1128/mmbr.43.2.260-296.1979

Barrett, J. R. (2017). Rice versus drinking water: Estimating the primary source of arsenic in the U.S. Diet. Environmental Health Perspectives, 125(6). https://doi.org/10.1289/EHP2096

Beinert, H., Holm, R. H., & Münck, E. (1997a). Iron-sulfur clusters: nature's modular, multipurpose structures. Science (New York, N.Y.), 277(5326), 653–659. https://doi.org/10.1126/science.277.5326.653

Beinert, H., Holm, R. H., & Münck, E. (1997b). Iron-sulfur clusters: nature's modular, multipurpose structures. Science (New York, N.Y.), 277(5326), 653–659. https://doi.org/10.1126/SCIENCE.277.5326.653

Belaidi, A. A., & Bush, A. I. (2016a). Iron neurochemistry in Alzheimer's disease and Parkinson's disease: targets for therapeutics. Journal of Neurochemistry, 179–197. https://doi.org/10.1111/JNC.13425

Belaidi, A. A., & Bush, A. I. (2016b). Iron neurochemistry in Alzheimer's disease and Parkinson's disease: targets for therapeutics. Journal of Neurochemistry, 179–197. https://doi.org/10.1111/JNC.13425

Berner, R. A. (1967). Thermodynamic Stability of Sedimentary Iron Sulfides. American Journal of Science, 265(9), 773–785. https://doi.org/10.2475/ajs.265.9.773

Berner, R. A. (1984a). Sedimentary pyrite formation: An update. Geochimica et Cosmochimica Acta, 48(4), 605–615. https://doi.org/https://doi.org/10.1016/0016-7037(84)90089-9

Berner, R. A. (1984b). Sedimentary pyrite formation: An update. Geochimica et Cosmochimica Acta, 48(4), 605–615. https://doi.org/10.1016/0016-7037(84)90089-9

Bilinski, T. (1991). Oxygen toxicity and microbial evolution. Biosystems, 24(4), 305–312. https://doi.org/https://doi.org/10.1016/0303-2647(91)90049-Q

Blochl, E., Keller, M., Wachtershauser, G., & Stetter, K. O. (1992). Reactions depending on iron sulfide and linking geochemistry with biochemistry. Proceedings of the National Academy of Sciences of the United States of America, 89(17), 8117–8120. https://doi.org/10.1073/pnas.89.17.8117

Boyd, E., Payne, D., Shepard, E., Spietz, R., Guo, Y., Broderick, W., & Broderick, J. (2022a). Biomining metals from pyritic ores: physiological, biochemical, and biophysical considerations. Biophysical Journal, 121(3, Supplement 1), 296a. https://doi.org/https://doi.org/10.1016/j.bpj.2021.11.1269

Boyd, E., Payne, D., Shepard, E., Spietz, R., Guo, Y., Broderick, W., & Broderick, J. (2022b). Biomining metals from pyritic ores: physiological, biochemical, and biophysical considerations. Biophysical Journal, 121(3), 296a. https://doi.org/10.1016/j.bpj.2021.11.1269

Brzóska, K., Meczyńska, S., & Kruszewski, M. (2006). Iron-sulfur cluster proteins: electron transfer and beyond. Acta Biochimica Polonica, 53(4), 685–691.

C., T. H., E., B. N., & B., P. P. (2005). Microbial origin of excess methane in glacial ice and implications for life on Mars. Proceedings of the National Academy of Sciences, 102(51), 18292–18296. https://doi.org/10.1073/pnas.0507601102

Carlin, A., Shi, W., Dey, S., & Rosen, B. P. (1995). The ars operon of Escherichia coli confers arsenical and antimonial resistance. Journal of Bacteriology, 177(4), 981–986. https://doi.org/10.1128/JB.177.4.981-986.1995

Chaudhary, P. P., Conway, P. L., & Schlundt, J. (2018). Methanogens in humans: potentially beneficial or harmful for health. Applied Microbiology and Biotechnology, 102(7), 3095–3104. https://doi.org/10.1007/s00253-018-8871-2

Chaudhuri, S., Vyas, K., Kapasi, P., Komar, A. A., Dinman, J. D., Barik, S., & Mazumder, B. (2007). Human ribosomal protein L13a is dispensable for canonical ribosome function but indispensable for efficient rRNA methylation. RNA (New York, N.Y.), 13(12), 2224–2237. https://doi.org/10.1261/rna.694007

Chávez-Capilla, T. (2022). The Need to Unravel Arsenolipid Transformations in Humans. DNA and Cell Biology, 41(1), 64–70. https://doi.org/10.1089/DNA.2021.0476/ASSET/IMAGES/LARGE/DNA.2021.0476_FIGURE2.JPEG

Chen, J., & Rosen, B. P. (2020). The Arsenic Methylation Cycle: How Microbial Communities Adapted Methylarsenicals for Use as Weapons in the Continuing War for Dominance. Frontiers in Environmental Science, 8, 43. https://doi.org/10.3389/FENVS.2020.00043/BIBTEX

Christina, J., Alexis, E., Mason, M.-E., R., C. D., L., D. J., S., B. E., & W., M. W. (2022). Pathways of Iron and Sulfur Acquisition, Cofactor Assembly, Destination, and Storage in Diverse Archaeal Methanogens and Alkanotrophs. Journal of Bacteriology, 203(17), e00117-21. https://doi.org/10.1128/JB.00117-21

Ciccarelli, F. D., Copley, R. R., Doerks, T., Russell, R. B., & Bork, P. (2002). CASH – a β-helix domain widespread among carbohydrate-binding proteins. Trends in Biochemical Sciences, 27(2), 59–62. https://doi.org/https://doi.org/10.1016/S0968-0004(01)02046-1

Cohen, S. M., Arnold, L. L., Eldan, M., Lewis, A. S., & Beck, B. D. (2008). Methylated Arsenicals: The Implications of Metabolism and Carcinogenicity Studies in Rodents to Human Risk Assessment. Http://Dx.Doi.Org/10.1080/10408440500534230, 36(2), 99–133. https://doi.org/10.1080/10408440500534230

Colman, D. R., Feyhl-Buska, J., Fecteau, K. M., Xu, H., Shock, E. L., & Boyd, E. S. (2016). Ecological differentiation in planktonic and sediment-associated chemotrophic microbial populations in Yellowstone hot springs. FEMS Microbiology Ecology, 92(9). https://doi.org/10.1093/femsec/fiw137

Coryell, M., McAlpine, M., Pinkham, N. V., McDermott, T. R., & Walk, S. T. (2018). The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. Nature Communications, 9(1). https://doi.org/10.1038/S41467-018-07803-9

Cramer, J. D., & Jarrett, J. T. (2018). Purification, Characterization, and Biochemical Assays of Biotin Synthase From Escherichia coli. Methods in Enzymology, 606, 363–388. https://doi.org/10.1016/BS.MIE.2018.06.003

Cullen, W. R., & Reimer, K. J. (1989). Arsenic Speciation in the Environment. Chemical Reviews, 89(4), 713–764. https://doi.org/10.1021/CR00094A002/ASSET/CR00094A002.FP.PNG_V03

Cutter, G. A., & Kluckhohn, R. S. (1999). The cycling of particulate carbon, nitrogen, sulfur, and sulfur species (iron monosulfide, greigite, pyrite, and organic sulfur) in the water columns of Framvaren Fjord and the Black Sea. Marine Chemistry, 67(3–4), 149–160. https://doi.org/10.1016/S0304-4203(99)00056-0

De Vrieze, J., Hennebel, T., Boon, N., & Verstraete, W. (2012a). Methanosarcina: The rediscovered methanogen for heavy duty biomethanation. Bioresource Technology, 112, 1–9. https://doi.org/10.1016/j.biortech.2012.02.079

De Vrieze, J., Hennebel, T., Boon, N., & Verstraete, W. (2012b). Methanosarcina: the rediscovered methanogen for heavy duty biomethanation. Bioresource Technology, 112, 1–9. https://doi.org/10.1016/J.BIORTECH.2012.02.079

Delaney, S. R., Nandakumar, V., & Jannetto, P. J. (2020). Trust your gut: a case of persistent gastrointestinal disturbances. Toxicology Cases for the Clinical and Forensic Laboratory, 433–435. https://doi.org/10.1016/B978-0-12-815846-3.00086-7

Devon, P., M., S. E., L., S. R., Katherine, S., Sue, B., Mark, Y., Brian, B., E., B. W., B., B. J., S., B. E., & Y., G. M. (2022). Examining Pathways of Iron and Sulfur Acquisition, Trafficking, Deployment, and Storage in Mineral-Grown Methanogen Cells. Journal of Bacteriology, 203(19), e00146-21. https://doi.org/10.1128/JB.00146-21

Dlouhy, A. C., & Outten, C. E. (2013a). The iron metallome in eukaryotic organisms. Metal Ions in Life Sciences, 12, 241–278. https://doi.org/10.1007/978-94-007-5561-1_8

Dlouhy, A. C., & Outten, C. E. (2013b). The iron metallome in Eukaryotic organisms. Metal Ions in Life Sciences, 12, 241–278. https://doi.org/10.1007/978-94-007-5561-1_8/FIGURES/12

Dong, X., Shulzhenko, N., Lemaitre, J., Greer, R. L., Peremyslova, K., Quamruzzaman, Q., Rahman, M., Ibn Hasan, O. S., Joya, S. A., Golam, M., Christiani, D. C., Morgun, A., & Kile, M. L. (2017). Arsenic exposure and intestinal microbiota in children from Sirajdikhan, Bangladesh. PloS One, 12(12). https://doi.org/10.1371/JOURNAL.PONE.0188487

E., P. J., & C., G. F. (2000). Lack of a Role for Iron in the Lyme Disease Pathogen. Science, 288(5471), 1651–1653. https://doi.org/10.1126/science.288.5471.1651

Elrod, V. A., Berelson, W. M., Coale, K. H., & Johnson, K. S. (2004). The flux of iron from continental shelf sediments: A missing source for global budgets. Geophysical Research Letters, 31(12). https://doi.org/10.1029/2004GL020216

Feng, J., Tian, H., Huang, Y., Ding, Z., & Yin, Z. (2019). Pyrite oxidation mechanism in aqueous medium. Journal of the Chinese Chemical Society, 66(4), 345–354. https://doi.org/10.1002/jccs.201800368

Fernandes-Martins, M. C., Keller, L. M., Munro-Ehrlich, M., Zimlich, K. R., Mettler, M. K., England, A. M., Clare, R., Surya, K., Shock, E. L., Colman, D. R., & Boyd, E. S. (2021). Ecological dichotomies arise in microbial communities due to mixing of deep hydrothermal waters and atmospheric gas in a circumneutral hot spring. Applied and Environmental Microbiology, 87(23). https://doi.org/10.1128/AEM.01598-21

Fike, D. A., Bradley, A. S., & Rose, C. V. (2015). Rethinking the ancient sulfur cycle. Ann Rev Earth Planet Sci, 43, 593–622. https://doi.org/10.1146/annurev-earth-060313-054802

Fukuda, S., Terasawa, M., & Shiomi, K. (2011). Phosphatidylarsenocholine, one of the major arsenolipids in marine organisms: Synthesis and metabolism in mice. Food and Chemical Toxicology, 49(7), 1598–1603. https://doi.org/10.1016/J.FCT.2011.03.053

Furbank, R. T., & Tester, M. (2011). Phenomics - technologies to relieve the phenotyping bottleneck. Trends in Plant Science, 16(12), 635–644. https://doi.org/10.1016/j.tplants.2011.09.005

Gaion, A., Sartori, D., Scuderi, A., & Fattorini, D. (2014). Bioaccumulation and biotransformation of arsenic compounds in Hediste diversicolor (Muller 1776) after exposure to spiked sediments. Environmental Science and Pollution Research, 21(9), 5952–5959. https://doi.org/10.1007/S11356-014-2538-Z/TABLES/3

Garbinski, L. D., Rosen, B. P., & Chen, J. (2019). Pathways of arsenic uptake and efflux. Environment International, 126, 585–597. https://doi.org/10.1016/J.ENVINT.2019.02.058

Ghaisas, S., Maher, J., & Kanthasamy, A. (2016). Gut microbiome in health and disease: Linking the microbiome–gut–brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. Pharmacology & Therapeutics, 158, 52–62. https://doi.org/10.1016/J.PHARMTHERA.2015.11.012

Giles, B. H., & Mann, K. K. (2022). Arsenic as an immunotoxicant. Toxicology and Applied Pharmacology, 454, 116248. https://doi.org/10.1016/J.TAAP.2022.116248

Grau-Perez, M., Kuo, C. C., Gribble, M. O., Balakrishnan, P., Spratlen, M. J., Vaidya, D., Francesconi, K. A., Goessler, W., Guallar, E., Silbergeld, E. K., Umans, J. G., Best, L. G., Lee, E. T., Howard, B. V., Cole, S. A., & Navas-Acien, A. (2017). Association of low-moderate arsenic exposure and arsenic metabolism with incident diabetes and insulin resistance in the strong heart family study. Environmental Health Perspectives, 125(12). https://doi.org/10.1289/EHP2566

Hall, A. J. (1986a). Pyrite-pyrrhotine redox reactions in nature. Mineralogical Magazine, 50(356), 223–229.

Hall, A. J. (1986b). Pyrite-pyrrhotine redox reactions in nature. Mineralogical Magazine, 50(356), 223–229. https://doi.org/10.1180/MINMAG.1986.050.356.05

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Almagro Armenteros, J. J., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. BioRxiv. https://doi.org/10.1101/2022.04.08.487609

Hidese, R., Inoue, T., Imanaka, T., & Fujiwara, S. (2014). Cysteine desulphurase plays an important role in environmental adaptation of the hyperthermophilic archaeon Thermococcus kodakarensis. Molecular Microbiology, 93(2), 331–345. https://doi.org/10.1111/mmi.12662

Holloway, P., McCormick, W., Watson, R. J., & Chan, Y. K. (1996). Identification and analysis of the dissimilatory nitrous oxide reduction genes, nosRZDFY, of Rhizobium meliloti. Journal of Bacteriology, 178(6), 1505–1514. https://doi.org/10.1128/jb.178.6.1505-1514.1996

Hook, S. E., Wright, A. D. G., & McBride, B. W. (2010a). Methanogens: Methane producers of the rumen and mitigation strategies. Archaea, 2010. https://doi.org/10.1155/2010/945785

Hook, S. E., Wright, A.-D. G., & McBride, B. W. (2010b). Methanogens: Methane Producers of the Rumen and Mitigation Strategies. Archaea, 2010, 945785. https://doi.org/10.1155/2010/945785

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Hughes, J., Ward, C. J., Peral, B., Aspinwall, R., Clark, K., San Millán, J. L., Gamble, V., & Harris, P. C. (1995). The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. Nature Genetics, 10(2), 151–160. https://doi.org/10.1038/ng0695-151

Imlay, J. A. (2006). Iron-sulphur clusters and the problem with oxygen. Molecular Microbiology, 59(4), 1073–1082. https://doi.org/https://doi.org/10.1111/j.1365-2958.2006.05028.x

Iron | Linus Pauling Institute | Oregon State University. (n.d.-a). Retrieved April 20, 2023, from https://lpi.oregonstate.edu/mic/minerals/iron

Iron | Linus Pauling Institute | Oregon State University. (n.d.-b). Retrieved April 10, 2023, from https://lpi.oregonstate.edu/mic/minerals/iron

Jing, H., Takagi, J., Liu, J., Lindgren, S., Zhang, R., Joachimiak, A., Wang, J., & Springer, T. A. (2002). Archaeal Surface Layer Proteins Contain β Propeller, PKD, and β Helix Domains and Are Related to Metazoan Cell Surface Proteins. Structure, 10(10), 1453–1464. https://doi.org/https://doi.org/10.1016/S0969-2126(02)00840-7

Johnson, C., England, A., Munro-Ehrlich, M., Colman, D. R., DuBois, J. L., & Boyd, E. S. (2021). Pathways of Iron and Sulfur Acquisition, Cofactor Assembly, Destination, and Storage in

Diverse Archaeal Methanogens and Alkanotrophs. Journal of Bacteriology, 203(17). https://doi.org/10.1128/JB.00117-21

Johnson, D. C., Dean, D. R., Smith, A. D., & Johnson, M. K. (2005a). Structure, function, and formation of biological iron-sulfur clusters. Annual Review of Biochemistry, 74, 247–281. https://doi.org/10.1146/ANNUREV.BIOCHEM.74.082803.133518

Johnson, D. C., Dean, D. R., Smith, A. D., & Johnson, M. K. (2005b). Structure, function, and formation of biological iron-sulfur clusters. Annual Review of Biochemistry, 74, 247–281. https://doi.org/10.1146/annurev.biochem.74.082803.133518

Johnston, D. T. (2011). Multiple sulfur isotopes and the evolution of Earth's surface sulfur cycle. Earth-Science Reviews, 106(1–2), 161–183. https://doi.org/10.1016/j.earscirev.2011.02.003

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Science, 28(11), 1947–1951. https://doi.org/https://doi.org/10.1002/pro.3715

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2020). KEGG: integrating viruses and cellular organisms. Nucleic Acids Research, 49(D1), D545–D551. https://doi.org/10.1093/nar/gkaa970

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kempf, B., & Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Archives of Microbiology, 170(5), 319–330. https://doi.org/10.1007/S002030050649/METRICS

Keren, R., Méheust, R., Santini, J. M., Thomas, A., West-Roberts, J., Banfield, J. F., & Alvarez-Cohen, L. (2022). Global genomic analysis of microbial biotransformation of arsenic highlights the importance of arsenic methylation in environmental and human microbiomes. Computational and Structural Biotechnology Journal, 20, 559–572. https://doi.org/10.1016/J.CSBJ.2021.12.040

Kitano, H. (2002). Systems biology: A brief overview. Science, 295(5560), 1662–1664. https://doi.org/10.1126/science.1069492

Klein, C. (2005a). Some Precambrian banded iron-formations (BIFs) from around the world: Their age, geologic setting, mineralogy, metamorphism, geochemistry, and origin. American Mineralogist, 90(10), 1473–1499. https://doi.org/10.2138/am.2005.1871

Klein, C. (2005b). Some Precambrian banded iron-formations (BIFs) from around the world: Their age, geologic setting, mineralogy, metamorphism, geochemistry, and origins. American Mineralogist, 90(10), 1473–1499. https://doi.org/10.2138/AM.2005.1871

Komorowicz, I., Sajnóg, A., & Barałkiewicz, D. (2019). Total Arsenic and Arsenic Species Determination in Freshwater Fish by ICP-DRC-MS and HPLC/ICP-DRC-MS Techniques. Molecules (Basel, Switzerland), 24(3). https://doi.org/10.3390/MOLECULES24030607

Konhauser, K. O., Lalonde, S. V, Planavsky, N. J., Pecoits, E., Lyons, T. W., Mojzsis, S. J., Rouxel, O. J., Barley, M. E., Rosìere, C., Fralick, P. W., Kump, L. R., & Bekker, A. (2011). Aerobic bacterial pyrite oxidation and acid rock drainage during the Great Oxidation Event. Nature, 478(7369), 369–373. https://doi.org/10.1038/nature10511

Laganowsky, A., Reading, E., Hopper, J. T. S., & Robinson, C. V. (2013). Mass spectrometry of intact membrane protein complexes. Nature Protocols, 8(4), 639–651. https://doi.org/10.1038/NPROT.2013.024

Laham, N., & Ehrlich, R. (2004). Manipulation of iron to determine survival. Immunologic Research, 30(1), 15. https://doi.org/10.1385/IR:30:1:015

Lashof, D. A., & Ahuja, D. R. (1990). Relative contributions of greenhouse gas emissions to global warming. Nature, 344(6266), 529–531. https://doi.org/10.1038/344529A0

Lau, C. K. Y., Krewulak, K. D., & Vogel, H. J. (2016a). Bacterial ferrous iron transport: the Feo system. FEMS Microbiology Reviews, 40(2), 273–298. https://doi.org/10.1093/FEMSRE/FUV049

Lau, C. K. Y., Krewulak, K. D., & Vogel, H. J. (2016b). Bacterial ferrous iron transport: the Feo system. FEMS Microbiology Reviews, 40(2), 273–298. https://doi.org/10.1093/femsre/fuv049

Law, K. P., He, W., Tao, J., & Zhang, C. (2021). A Novel Approach to Characterize the Lipidome of Marine Archaeon Nitrosopumilus maritimus by Ion Mobility Mass Spectrometry. Frontiers in Microbiology, 12, 3416. https://doi.org/10.3389/FMICB.2021.735878/BIBTEX

Li, H., Condon, F., Kessler, D., Nandi, V., Rebosa, M., Westerman, M., Shaz, B. H., & Ginzburg, Y. (2016a). Evidence of relative iron deficiency in platelet- and plasma-pheresis donors correlates with donation frequency. Journal of Clinical Apheresis, 31(6), 551–558. https://doi.org/10.1002/JCA.21448

Li, H., Condon, F., Kessler, D., Nandi, V., Rebosa, M., Westerman, M., Shaz, B. H., & Ginzburg, Y. (2016b). Evidence of relative iron deficiency in platelet- and plasma-pheresis donors correlates with donation frequency. Journal of Clinical Apheresis, 31(6), 551–558. https://doi.org/10.1002/JCA.21448

Lill, R., & Freibert, S.-A. (2020). Mechanisms of Mitochondrial Iron-Sulfur Protein Biogenesis. Annual Review of Biochemistry, 89(1), 471–499. https://doi.org/10.1146/annurev-biochem-013118-111540

Lindley, P. F. (1996). Iron in biology: a structural viewpoint. Reports on Progress in Physics, 59(7), 867–933. https://doi.org/10.1088/0034-4885/59/7/002

Liu, Y., Beer, L. L., & Whitman, W. B. (2012a). Methanogens: a window into ancient sulfur metabolism. Trends in Microbiology, 20(5), 251–258. https://doi.org/https://doi.org/10.1016/j.tim.2012.02.002

Liu, Y., Beer, L. L., & Whitman, W. B. (2012b). Methanogens: a window into ancient sulfur metabolism. Trends in Microbiology, 20(5), 251–258. https://doi.org/10.1016/J.TIM.2012.02.002

Liu, Y., Sieprawska-Lupa, M., Whitman, W. B., & White, R. H. (2010a). Cysteine is not the sulfur source for iron-sulfur cluster and methionine biosynthesis in the methanogenic archaeon Methanococcus maripaludis. The Journal of Biological Chemistry, 285(42), 31923–31929. https://doi.org/10.1074/jbc.M110.152447

Liu, Y., Sieprawska-Lupa, M., Whitman, W. B., & White, R. H. (2010b). Cysteine is not the sulfur source for iron-sulfur cluster and methionine biosynthesis in the methanogenic archaeon Methanococcus maripaludis. The Journal of Biological Chemistry, 285(42), 31923–31929. https://doi.org/10.1074/JBC.M110.152447

Luvonga, C., Rimmer, C. A., Yu, L. L., & Lee, S. B. (2020a). Organoarsenicals in Seafood: Occurrence, Dietary Exposure, Toxicity, and Risk Assessment Considerations - A Review. Journal of Agricultural and Food Chemistry, 68(4), 943–960. https://doi.org/10.1021/ACS.JAFC.9B07532

Luvonga, C., Rimmer, C. A., Yu, L. L., & Lee, S. B. (2020b). Organoarsenicals in Seafood: Occurrence, Dietary Exposure, Toxicity, and Risk Assessment Considerations - A Review. Journal of Agricultural and Food Chemistry, 68(4), 943–960. https://doi.org/10.1021/ACS.JAFC.9B07532/ASSET/IMAGES/LARGE/JF9B07532_0007.JPEG

Luvonga, C., Rimmer, C. A., Yu, L. L., & Lee, S. B. (2020c). Organoarsenicals in Seafood: Occurrence, Dietary Exposure, Toxicity, and Risk Assessment Considerations - A Review. Journal of Agricultural and Food Chemistry, 68(4), 943–960. https://doi.org/10.1021/ACS.JAFC.9B07532

Lyons, T. W., & Severmann, S. (2006a). A critical look at iron paleoredox proxies: New insights from modern euxinic marine basins. Geochimica et Cosmochimica Acta, 70(23 SPEC. ISS.), 5698–5722. https://doi.org/10.1016/j.gca.2006.08.021

Lyons, T. W., & Severmann, S. (2006b). A critical look at iron paleoredox proxies: New insights from modern euxinic marine basins. Geochimica et Cosmochimica Acta, 70(23 SPEC. ISS.), 5698–5722. https://doi.org/10.1016/J.GCA.2006.08.021

M., B., M., G., I., F., G., P., G., C., M., R., & G., P. (2011). Structural and Functional Insights into Aeropyrum pernix OppA, a Member of a Novel Archaeal OppA Subfamily . Journal of Bacteriology, 193(3), 620–630. https://doi.org/10.1128/JB.00899-10

Major, T. A., Burd, H., & Whitman, W. B. (2004a). Abundance of 4Fe–4S motifs in the genomes of methanogens and other prokaryotes. FEMS Microbiology Letters, 239(1), 117–123. https://doi.org/10.1016/j.femsle.2004.08.027

Major, T. A., Burd, H., & Whitman, W. B. (2004b). Abundance of 4Fe–4S motifs in the genomes of methanogens and other prokaryotes. FEMS Microbiology Letters, 239(1), 117–123. https://doi.org/10.1016/J.FEMSLE.2004.08.027

Martin, D. R., & Matyushov, D. V. (2017). Electron-transfer chain in respiratory complex I. Scientific Reports, 7(1), 5495. https://doi.org/10.1038/s41598-017-05779-y

Martin, W. F., & Sousa, F. L. (2015). Early Microbial Evolution: The Age of Anaerobes. Cold Spring Harbor Perspectives in Biology, 8(2), a018127. https://doi.org/10.1101/cshperspect.a018127

McMichael, A. J., Powles, J. W., Butler, C. D., & Uauy, R. (2007). Food, livestock production, energy, climate change, and health. The Lancet, 370(9594), 1253–1263.

Mihara, H., & Esaki, N. (2002). Bacterial cysteine desulfurases: their function and mechanisms. Applied Microbiology and Biotechnology, 60(1–2), 12–23. https://doi.org/10.1007/s00253-002-1107-4

Murphy, R., & Strongin, D. R. (2009). Surface reactivity of pyrite and related sulfides. Surface Science Reports, 64(1), 1–45. https://doi.org/10.1016/j.surfrep.2008.09.002

Palma-Lara, I., Martínez-Castillo, M., Quintana-Pérez, J. C., Arellano-Mendoza, M. G., Tamay-Cach, F., Valenzuela-Limón, O. L., García-Montalvo, E. A., & Hernández-Zavala, A. (2020). Arsenic exposure: A public health problem leading to several cancers. Regulatory Toxicology and Pharmacology, 110, 104539. https://doi.org/10.1016/J.YRTPH.2019.104539

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.-É., Li, S., & Xia, J. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Research, 49(W1), W388–W396. https://doi.org/10.1093/nar/gkab382

Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N., & Xia, J. (2022). Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. Nature Protocols. https://doi.org/10.1038/s41596-022-00710-w

Parvez, F., Wasserman, G. A., Factor-Litvak, P., Liu, X., Slavkovich, V., Siddique, A. B., Sultana, R., Sultana, R., Islam, T., Levy, D., Mey, J. L., van Geen, A., Khan, K., Kline, J., Ahsan, H., & Graziano, J. H. (2011). Arsenic Exposure and Motor Function among Children in Bangladesh. Environmental Health Perspectives, 119(11), 1665. https://doi.org/10.1289/EHP.1103548

Patti, G. J., Yanes, O., & Siuzdak, G. (2012). Innovation: Metabolomics: the apogee of the omics trilogy. Nature Reviews Molecular Cell Biology, 13(4), 263–269. https://doi.org/10.1038/nrm3314

Pattus, F., & Abdallah, M. A. (2000a). Siderophores and Iron-Transport in Microorganisms. Journal of the Chinese Chemical Society, 47(1), 1–20. https://doi.org/10.1002/JCCS.200000001

Pattus, F., & Abdallah, M. A. (2000b). Siderophores and Iron-Transport in Microorganisms. Journal of the Chinese Chemical Society, 47(1), 1–20. https://doi.org/https://doi.org/10.1002/jccs.200000001

Payne, D., Shepard, E. M., Spietz, R. L., Steward, K., Brumfield, S., Young, M., Bothner, B., Broderick, W. E., Broderick, J. B., & Boyd, E. S. (2021). Examining Pathways of Iron and Sulfur Acquisition, Trafficking, Deployment, and Storage in Mineral-Grown Methanogen Cells. Journal of Bacteriology, 203(19). https://doi.org/10.1128/JB.00146-21

Payne, D., Spietz, R. L., & Boyd, E. S. (2021a). Reductive dissolution of pyrite by methanogenic archaea. The ISME Journal, 15(12), 3498–3507. https://doi.org/10.1038/s41396-021-01028-3

Payne, D., Spietz, R. L., & Boyd, E. S. (2021b). Reductive dissolution of pyrite by methanogenic archaea. The ISME Journal 2021 15:12, 15(12), 3498–3507. https://doi.org/10.1038/s41396-021-01028-3

Percak-Dennett, E., He, S., Converse, B., Konishi, H., Xu, H., Corcoran, A., Noguera, D., Chan, C., Bhattacharyya, A., Borch, T., Boyd, E., & Roden, E. E. (2017). Microbial acceleration of aerobic pyrite oxidation at circumneutral pH. Geobiology, 15(5), 690–703. https://doi.org/https://doi.org/10.1111/gbi.12241

Perkins, D. N., Pappin, D. J. C., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. ELECTROPHORESIS, 20(18), 3551–3567. https://doi.org/https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

Peters, J. W., Schut, G. J., Boyd, E. S., Mulder, D. W., Shepard, E. M., Broderick, J. B., King, P. W., & Adams, M. W. W. (2015). [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1853(6), 1350–1369. https://doi.org/https://doi.org/10.1016/j.bbamcr.2014.11.021

Podgorski, J., & Berg, M. (2020). Global threat of arsenic in groundwater. Science, 368(6493), 845–850. https://doi.org/10.1126/SCIENCE.ABA1510/SUPPL_FILE/ABA1510_PODGORSKI_SM.PDF

Powell, L. W., Seckington, R. C., & Deugnier, Y. (2016). Haemochromatosis. The Lancet, 388(10045), 706–716. https://doi.org/10.1016/S0140-6736(15)01315-X

Qi, Y., Li, X., Chang, C., Xu, F., He, Q., Zhao, Y., & Wu, L. (2017). Ribosomal protein L23 negatively regulates cellular apoptosis via the RPL23/Miz-1/c-Myc circuit in higher-risk myelodysplastic syndrome. Scientific Reports, 7(1), 2323. https://doi.org/10.1038/s41598-017-02403-x

Rawle, R. A., Kang, Y. S., Bothner, B., Wang, G., & McDermott, T. R. (2019). Transcriptomics analysis defines global cellular response of Agrobacterium tumefaciens 5A to arsenite exposure regulated through the histidine kinases PhoR and AioS. Environmental Microbiology, 21(8), 2659–2676. https://doi.org/10.1111/1462-2920.14577

Rawle, R. A., Tokmina-Lukaszewska, M., Shi, Z., Kang, Y. S., Tripet, B. P., Dang, F., Wang, G., McDermott, T. R., Copie, V., & Bothner, B. (2020). Metabolic Responses to Arsenite Exposure Regulated through Histidine Kinases PhoR and AioS in Agrobacterium tumefaciens 5A. Microorganisms 2020, Vol. 8, Page 1339, 8(9), 1339. https://doi.org/10.3390/MICROORGANISMS8091339

Rawle, R., Saley, T. C., Kang, Y. S., Wang, Q., Walk, S., Bothner, B., & McDermott, T. R. (2021). Introducing the ArsR-Regulated Arsenic Stimulon. Frontiers in Microbiology, 12, 162. https://doi.org/10.3389/FMICB.2021.630562/BIBTEX

Rickard, D., & Luther, G. W. (2007a). Chemistry of iron sulfides. In Chemical Reviews (Vol. 107, Issue 2). https://doi.org/10.1021/cr0503658

Rickard, D., & Luther, G. W. (2007b). Chemistry of iron sulfides. Chemical Reviews, 107(2), 514–562. https://doi.org/10.1021/CR0503658/ASSET/CR0503658.FP.PNG_V03

Rocha, E. R., Bergonia, H. A., Gerdes, S., & Jeffrey Smith, C. (2019). Bacteroides fragilis requires the ferrous-iron transporter FeoAB and the CobN-like proteins BtuS1 and BtuS2 for assimilation of iron released from heme. MicrobiologyOpen, 8(4), e00669. https://doi.org/10.1002/mbo3.669

Rohlin, L., Leon, D. R., Kim, U., Loo, J. A., Ogorzalek Loo, R. R., & Gunsalus, R. P. (2012). Identification of the Major Expressed S-Layer and Cell Surface-Layer-Related Proteins in the Model Methanogenic Archaea: Methanosarcina barkeri Fusaro and Methanosarcina acetivorans C2A. Archaea, 2012, 873589. https://doi.org/10.1155/2012/873589

Rohwerder, T., Gehrke, T., Kinzler, K., & Sand, W. (2003). Bioleaching review part A: Progress in bioleaching: Fundamentals and mechanisms of bacterial metal sulfide oxidation. Applied Microbiology and Biotechnology, 63(3), 239–248. https://doi.org/10.1007/s00253-003-1448-7

Rouault, T. A., & Tong, W.-H. (2005). Iron–sulphur cluster biogenesis and mitochondrial iron homeostasis. Nature Reviews Molecular Cell Biology, 6(4), 345–351. https://doi.org/10.1038/nrm1620

Satiaputra, J., Shearwin, K. E., Booker, G. W., & Polyak, S. W. (2016). Mechanisms of biotin-regulated gene expression in microbes. Synthetic and Systems Biotechnology, 1(1), 17–24. https://doi.org/10.1016/J.SYNBIO.2016.01.005

Schoonen, M. A. A. (2004a). Mechanisms of sedimentary pyrite formation. Special Paper of the Geological Society of America, 379, 117–134. https://doi.org/10.1130/0-8137-2379-5.117

Schoonen, M. A. A. (2004b). Mechanisms of sedimentary pyrite formation. Geological Society of America Special Papers, 379, 117–134.

Science | AAAS. (n.d.). Retrieved April 14, 2023, from https://www.science.org/journal/science/no-first-release

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Research, gkac194. https://doi.org/10.1093/nar/gkac194

Shibata, M., & Terada, F. (2010). Factors affecting methane production and mitigation in ruminants. Animal Science Journal, 81(1), 2–10. https://doi.org/10.1111/J.1740-0929.2009.00687.X

Smethurst, D. G. J., & Shcherbik, N. (2021). Interchangeable utilization of metals: New perspectives on the impacts of metal ions employed in ancient and extant biomolecules. Journal of Biological Chemistry, 297(6), 101374. https://doi.org/10.1016/j.jbc.2021.101374

Spietz, R. L., Payne, D., Kulkarni, G., Metcalf, W. W., Roden, E. E., & Boyd, E. S. (2022a). Investigating Abiotic and Biotic Mechanisms of Pyrite Reduction. In Frontiers in microbiology. (Vol. 13). https://doi.org/10.3389/fmicb.2022.878387

Spietz, R. L., Payne, D., Kulkarni, G., Metcalf, W. W., Roden, E. E., & Boyd, E. S. (2022b). Investigating Abiotic and Biotic Mechanisms of Pyrite Reduction. Frontiers in Microbiology, 13, 1421. https://doi.org/10.3389/FMICB.2022.878387/BIBTEX

Spietz, R. L., Payne, D., Szilagyi, R., & Boyd, E. S. (2022a). Reductive biomining of pyrite by methanogens. Trends in Microbiology. https://doi.org/https://doi.org/10.1016/j.tim.2022.05.005

Spietz, R. L., Payne, D., Szilagyi, R., & Boyd, E. S. (2022b). Reductive biomining of pyrite by methanogens. Trends in Microbiology, 30(11), 1072–1083. https://doi.org/10.1016/j.tim.2022.05.005

Steward, K. F., Payne, D., Kincannon, W., Johnson, C., Lensing, M., Fausset, H., Németh, B., Shepard, E. M., Broderick, W. E., Broderick, J. B., Dubois, J., & Bothner, B. (2022a). Proteomic Analysis of Methanococcus voltae Grown in the Presence of Mineral and Nonmineral Sources of Iron and Sulfur. Microbiology Spectrum, 10(4), e01893-22. https://doi.org/10.1128/spectrum.01893-22

Steward, K. F., Payne, D., Kincannon, W., Johnson, C., Lensing, M., Fausset, H., Németh, B., Shepard, E. M., Broderick, W. E., Broderick, J. B., Dubois, J., & Bothner, B. (2022b). Proteomic Analysis of Methanococcus voltae Grown in the Presence of Mineral and Nonmineral Sources of Iron and Sulfur. Microbiology Spectrum, 10(4). https://doi.org/10.1128/SPECTRUM.01893-22

Stüeken, E. E., Catling, D. C., & Buick, R. (2012). Contributions to late Archaean sulphur cycling by life on land. Nature Geoscience, 5(10), 722–725. https://doi.org/10.1038/ngeo1585

Tam, L. M., Price, N. E., & Wang, Y. (2020). Molecular Mechanisms of Arsenic-Induced Disruption of DNA Repair. Chemical Research in Toxicology, 33(3), 709–726. https://doi.org/10.1021/ACS.CHEMRESTOX.9B00464/ASSET/IMAGES/LARGE/TX9B00464_0002.JPEG

Technologies, O. (2020). Overview of Systems Biology and Omics Technologies. January 2016. https://doi.org/10.2174/0929867323666160926

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., & Hedderich, R. (2008a). Methanogenic archaea: Ecologically relevant differences in energy conservation. Nature Reviews Microbiology, 6(8), 579–591. https://doi.org/10.1038/nrmicro1931

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., & Hedderich, R. (2008b). Methanogenic archaea: ecologically relevant differences in energy conservation. Nature Reviews. Microbiology, 6(8), 579–591. https://doi.org/10.1038/NRMICRO1931

Tokmina-Lukaszewska, M., Shi, Z., Tripet, B., McDermott, T. R., Copié, V., Bothner, B., & Wang, G. (2017). Metabolic response of Agrobacterium tumefaciens 5A to arsenite. Environmental Microbiology, 19(2), 710–721. https://doi.org/10.1111/1462-2920.13615

Tribovillard, N., Algeo, T. J., Lyons, T., & Riboulleau, A. (2006). Trace metals as paleoredox and paleoproductivity proxies: An update. Chemical Geology, 232(1–2), 12–32. https://doi.org/10.1016/j.chemgeo.2006.02.012

Truche, L., Berger, G., Destrigneville, C., Guillaume, D., & Giffaut, E. (2010a). Kinetics of pyrite to pyrrhotite reduction by hydrogen in calcite buffered solutions between 90 and 180 C: Implications for nuclear waste disposal. Geochimica et Cosmochimica Acta, 74(10), 2894–2914.

Truche, L., Berger, G., Destrigneville, C., Guillaume, D., & Giffaut, E. (2010b). Kinetics of pyrite to pyrrhotite reduction by hydrogen in calcite buffered solutions between 90 and 180°C: Implications for nuclear waste disposal. Geochimica et Cosmochimica Acta, 74(10), 2894–2914. https://doi.org/10.1016/J.GCA.2010.02.027

Tung, H. C., Bramall, N. E., & Price, P. B. (2005). Microbial origin of excess methane in glacial ice and implications for life on Mars. Proceedings of the National Academy of Sciences of the United States of America, 102(51), 18292–18296. https://doi.org/10.1073/PNAS.0507601102

Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S., & Isozaki, Y. (2006). Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. Nature, 440(7083), 516–519. https://doi.org/10.1038/nature04584

Vinet, L., & Zhedanov, A. (2011). A 'missing' family of classical orthogonal polynomials. Journal of Physics A: Mathematical and Theoretical, 44(8), 085201. https://doi.org/10.1088/1751-8113/44/8/085201

Waldrop, G. L., Holden, H. M., & Maurice, M. S. (2012). The enzymes of biotin dependent CO2 metabolism: What structures reveal about their reaction mechanisms. Protein Science, 21(11), 1597–1619. https://doi.org/10.1002/PRO.2156

Weber, K. A., Achenbach, L. A., & Coates, J. D. (2006). Microorganisms pumping iron: Anaerobic microbial iron oxidation and reduction. Nature Reviews Microbiology, 4(10), 752–764. https://doi.org/10.1038/nrmicro1490

Welz, D., & Braun, V. (1998). Ferric citrate transport of Escherichia coli: functional regions of the FecR transmembrane regulatory protein. Journal of Bacteriology, 180(9), 2387–2394. https://doi.org/10.1128/JB.180.9.2387-2394.1998

Wolfe, J. M., & Fournier, G. P. (2018). Horizontal gene transfer constrains the timing of methanogen evolution. Nature Ecology & Evolution, 2(5), 897–903. https://doi.org/10.1038/s41559-018-0513-7

Wu, B., Liu, F., Fang, W., Yang, T., Chen, G. H., He, Z., & Wang, S. (2021a). Microbial sulfur metabolism and environmental implications. The Science of the Total Environment, 778. https://doi.org/10.1016/J.SCITOTENV.2021.146085

Wu, B., Liu, F., Fang, W., Yang, T., Chen, G.-H., He, Z., & Wang, S. (2021b). Microbial sulfur metabolism and environmental implications. Science of The Total Environment, 778, 146085. https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.146085

Xue, X. M., Xiong, C., Yoshinaga, M., Rosen, B., & Zhu, Y. G. (2021). The enigma of environmental organoarsenicals: Insights and implications. Https://Doi.Org/10.1080/10643389.2021.1947678, 52(21), 3835–3862. https://doi.org/10.1080/10643389.2021.1947678

Zhao, F. J. (2016). A novel pathway of arsenate detoxification. Molecular Microbiology, 100(6), 928–930. https://doi.org/10.1111/MMI.13395

Zheng, L., White, R. H., Cash, V. L., Jack, R. F., & Dean, D. R. (1993). Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. Proceedings of the National Academy of Sciences, 90(7), 2754–2758. https://doi.org/10.1073/pnas.90.7.2754