Characterization of orphan mitochondrial proteins via quantitative interaction mapping

by

Brendan J. Floyd

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

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

2015

Date of final oral examination: May 7, 2015

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

David J. Pagliarini, Assistant Professor, Biochemistry
Mark E. Burkard, Assistant Professor, Medicine
Joshua J. Coon, Professor, Biomolecular Chemistry and Chemistry
John L. Markley, Professor, Biomolecular Structure
Sushmita Roy, Assistant Professor, Biostatistics and Medical Informatics

To Sally, for her love, support, and joy.

ACKNOWLEDGMENTS

I am extremely thankful to a whole host of people for their support throughout the course of my PhD and leading up to it. As may be the case now more than ever, science is a collaborative process, and my path has been particularly blessed with great teammates. Here, I'll highlight some particular groups and individuals that have made a great and positive difference in my graduate experience, while acknowledging that there are too many that fit the description to fit into even several theses.

To Team Interactome, especially Mike, Holly, Catie, Emily, and Kelly. The rolling stone of this project has gathered no moss, but it has rather accumulated a wonderful team. Thank you for helping with the process of getting this project off the ground, for bringing it to its current state, and for carrying the torch into the future. Best of luck as you all move forward!

To Brendan, Sarah, and Alexa, for being willing to carry out my crazy ideas. You joined me as undergraduate and high school students, and reminded me of the joys of early days in lab. Brendan, as my first ever trainee, I'm terribly grateful for your dedication and loyalty, even when I clearly didn't know what I was doing.

To the entire Pagliarini Lab, for being great friends and colleagues. I look forward to the great things you'll continue to achieve. Natalie, you have been a great source of ideas, and it has been terribly helpful to bounce

my thoughts off of you in our varied chats. Thanks as well for helping to edit the first chapter of this work in these final weeks. Jon, your strong work ethic and conscientiousness has been a model for me and many others in the lab, and will continue to serve you well.

To Dave, for building such a collaborative, highly-motivated environment in which to work. The team-based approach is a difficult one to manage, but it can yield wonderful results when done well, and I've benefited greatly. You took a chance on me at a difficult time, and that has made all the difference in my experience of grad school. I've enjoyed our chats and glimpses into the life of a PI. Most of all, I'm proud to have been a member of your lab.

To the Coon Group, especially Prof. Coon, Catie, Emily, Alex, Greg, Clay, and Emma. Your proteomics efforts have been a key differentiating feature of my projects, and have been absolutely essential to any success I've had. You have been great collaborators, bringing the highest levels of interest, urgency, and technical expertise!

To the Attie Lab, especially Alan, Kiki, Angie, Lindsay, and Melkam, for giving me my start in graduate school and laying the foundation for the researcher I came to be.

To my MSTP classmates, especially Josh, Rachel, Farsh, Anna, Jason, Heather, and Ben. It has been a long road together, and we still have a ways to go, but our time together is always fun and fulfilling. You've made my years here a joy.

To the MSTP directors, especially Deane Mosher and Anna Huttenlocher. Time and again, you have had faith in my abilities and gave me the support I needed to make tough decisions. My time in this program has been outstanding, and I owe that to a great deal to your confidence in me.

To my committee members: Prof. Josh Coon and Drs. Mark Burkard, John Markley, Sushmita Roy, Dawn Davis, Jon Audhya, Luigi Puglielli, and the late Paul Bertics. I wholeheartedly appreciate your support and guidance.

To all my mentors and teachers, past and present, including Mr. Graham, Dr. Block, Ms. Deichstetter, Holly Goodson, Paul Helquist, and Doc Doyle. Each of you have fanned the flame of my scientific interest in your own way, from showing that math and science can be fun to giving me a rigorous environment in which to expand that passion into a lifelong pursuit. Dr. Goodson, I'll never forget the joy I felt that first day when you gave me a bit of bench space in your lab. Doc and the ACE program, you gave me real responsibilities and supported my growth and maturation alongside all the bumps and bruises along the way while showing me what it means to be a conscientious professional.

To my family, for all their love and support. While it may not always be clear why I'm still in school, or what it is I'm actually doing, you've been there for me through it all. I owe Tom, Meg, Kate, and Kevin a great deal for charging ahead and showing me the way forward. To my parents and parents-in-law, your love and support for me and for one another is a

guide along the path that lay ahead. To my grandparents, you've all played a special role in bringing me to where I am today, and I will continue to strive to live up to your example.

And, most importantly, to Sally. You've been with me through highs and lows, always having confidence and faith in me even when I don't. You gave me the courage to make tough choices, and the strength to follow through. You inspire me to be my best each day by the way you live your life. I'm proud of you for the choices you've made and the work you've done, and I know this will all lead to your greater happiness. I've been blessed over these past few years of our life together, and look forward to many more wonderful things to come!

CONTENTS

LIST OF FIGURES

CHARACTERIZATION OF ORPHAN MITOCHONDRIAL PROTEINS VIA QUANTITATIVE INTERACTION MAPPING

Brendan J. Floyd

Under the supervision of Associate Professor David J. Pagliarini

At the University of Wisconsin-Madison

Mitochondria are essential for nearly all eukaryotic cells and play a key role in more than 300 human diseases. Over the past several decades, in part due to advancements in proteomics and the integration of orthogonal approaches, the components of the mitochondrial proteome have been revealed to near completion. In contrast to this depth of knowledge, we find that 20% of the mitochondrial proteome is comprised of uncharacterized, orphan mitochondrial proteins (OMPs). Similarly, many patients with overt mitochondrial disease lack molecular diagnoses. To illuminate key areas of mitochondrial biology, we apply mass spectrometry-based proteomics to quantify the dynamic nature of mitochondrial post-translational modifications (PTMs) and protein-protein interactions (PPIs).

While the known diversity and prevalence of mitochondrial protein PTMs have expanded rapidly of late, the regulatory role of these modifications, if any, is still poorly understood. To evaluate the extent and importance of one such PTM, lysine acetylation, we quantified the mouse liver mitochondrial acetylproteome under two pairs of contrasting states - obese vs. lean and fasted vs. refed. Our proteomic analyses support

the identification of thousands of acetyl sites on hundreds of proteins. Acetyl abundance at more than 500 sites is dynamically regulated across the aforementioned conditions. Furthermore, we apply state-of-the-art protein purification techniques to establish the inhibitory capacity of acetylation of the key metabolic enzyme acetyl-CoA acetyltransferase 1 (Acat1) at each of two dynamically-acetylated active site lysine residues.

To accelerate the functional characterization of the more than 200 orphan mitochondrial proteins, we applied affinity enrichment coupled to mass spectrometry (AE-MS) to map OMP-related PPIs across two cell types and nutrient states. We establish the first unbiased evidence of a mammalian coenzyme Q biosynthetic (CoQ) complex involving several OMPs. Furthermore, we associate many OMPs with known pathways and processes for the first time. We find that the OMP LYRM5, forms a specific, direct, and stable interaction with the electron transfer flavoprotein (ETF) complex. As mutations in ETF lead to disease, we prioritize LYRM5 as a possible cause of glutaric acidemia type II in patients lacking molecular diagnoses.

In this work, we evaluate the current state of mitochondrial protein and disease annotation, we expand the understanding of the regulatory nature of mitochondrial protein acetylation, we find evidence for and map the architecture of the mammalian CoQ biosynthetic complex, and we accelerate the functional annotation of more than a dozen OMPs.

1 CHAPTER ONE

Quantitative analysis of mitochondrial proteins and processes

Brendan J. Floyd and David J. Pagliarini

1.1 Abstract

With the aid of mass spectrometry-based proteomics, the mitochondrial proteome has been established to more than 85% completion. This compendium has facilitated the systematic analysis of the role of mitochondria in health and disease, while it simultaneously has elucidated the presence of hundreds of uncharacterized, orphan mitochondrial proteins. Recent advancements in focused, quantitative proteomics have catalyzed protein functional analysis via the elucidation of protein interactions and sub-organellar localization. Together, these tools expand our knowledge of the proteome beyond a parts list and toward a network diagram of mitochondrial structure and function. Continued efforts and integration of proteomics with orthogonal approaches will lead us toward the establishment of a map of this essential, dynamic organelle, and lay the ground work for similar approaches in other subcellular domains.

1.2 Introduction

Over the past decade, systematic approaches have facilitated the annotation and characterization of a host of pathways and processes. Among a variety of other methods, mass spectrometry (MS)-based proteomics has enabled the analysis of protein abundance, subcellular localization, and post-translational modifications, much of which could not be assessed by other approaches. For example, in comparison to nucleotide-based methods, proteomics is particularly effective at revealing new aspects of cellular biochemistry due to the fact that many cellular processes are carried out by proteins. Recent proteomics studies have been uniquely successful in illuminating our understanding of the mitochondrion. The goal of this review is to describe advancements in the identification and annotation of mitochondrial proteins over the past several decades, with an emphasis on the role of proteomics in this endeavor. These approaches may serve as an example for other cellular and organellar systems. Further implementation of such quantitative approaches will continue to expand our knowledge of protein biochemistry and open new avenues for study.

Establishment of the mitochondrial proteome

Following the first isolation of purified mitochondria in 1946 [1], biochemical studies of this organelle established its essential role in eukaryotic cells. This early work sought to identify the components responsible for mito-

chondrial energy metabolism, and successfully revealed the complexes of the electron transport chain [2]. Over the next several decades, it became clear that mitochondria not only provide more than 90% of cellular energy requirements, but also play key roles in cell signaling, apoptosis, response to viral infection, and many other processes [3]. In parallel, mitochondrial dysfunction was observed in more than 150 unique diseases, including many inborn errors of metabolism, diabetes, and cancer [3–5]. These studies led to the need for a deeper understanding of the role of specific mitochondrial components in health and disease states.

Early on, our incomplete knowledge of the mitochondrial proteome was an obstacle to the systematic study of this organelle. The lack of a discrete list of mitochondrial components made it difficult or impossible to associate mitochondrial phenotypes with the appropriate protein. For example, even though the biochemical nature of the calcium transporter had been determined to great extent, the lack of a mitochondrial proteome inhibited the search for the responsible polypeptide [6]. As such, it became clear that the first step toward a comprehensive understanding of the mitochondrion was to catalog the organellar proteome.

Despite several decades of biochemical characterization, the proteome of the mitochondrion has only recently been established to near-complete coverage. The direct identification of mitochondrial proteins has largely been the domain of focused studies of one or a handful of proteins at a time, until advances in technology allowed for the analysis of hundreds

or thousands of proteins simultaneously as with mass spectrometry (MS)-based proteomics.

MS-based approaches have proven particularly instrumental in the comprehensive cataloging of the mitochondrial proteome. Initial studies were extremely labor-intensive, requiring a week to identify 100 proteins from a 2-D gel [7, 8]. With improvements in technology came the possibility of identifying several hundred mitochondrial proteins in a single study. This ignited an explosion in the knowledge of the proteome (Fig. 1.1).

A quick succession of work began to more broadly establish the yeast, mouse, and human mitochondrial proteomes. In 2003, three groups worked out initial maps of mitochondrial proteins in yeast, mice, and humans [11, 16, 19]. These efforts were supported to a great extent by advancements in proteomics that allowed for greater sensitivity and throughput.

In 2006, the protein correlation profiling (PCP) technique allowed for the analysis of many organelles simultaneously [20]. Originally designed for centrosome analysis [21], this method relies on the co-appearance of organellar proteins across a density centrifugation gradient. While the depth of coverage for mitochondrial proteins was not as great as for previous studies, this method was put to great use in establishing subtle differences in the profile of proteins in similar subcellular compartments. It also became the foundation for complexome profiling, as described later in this review.

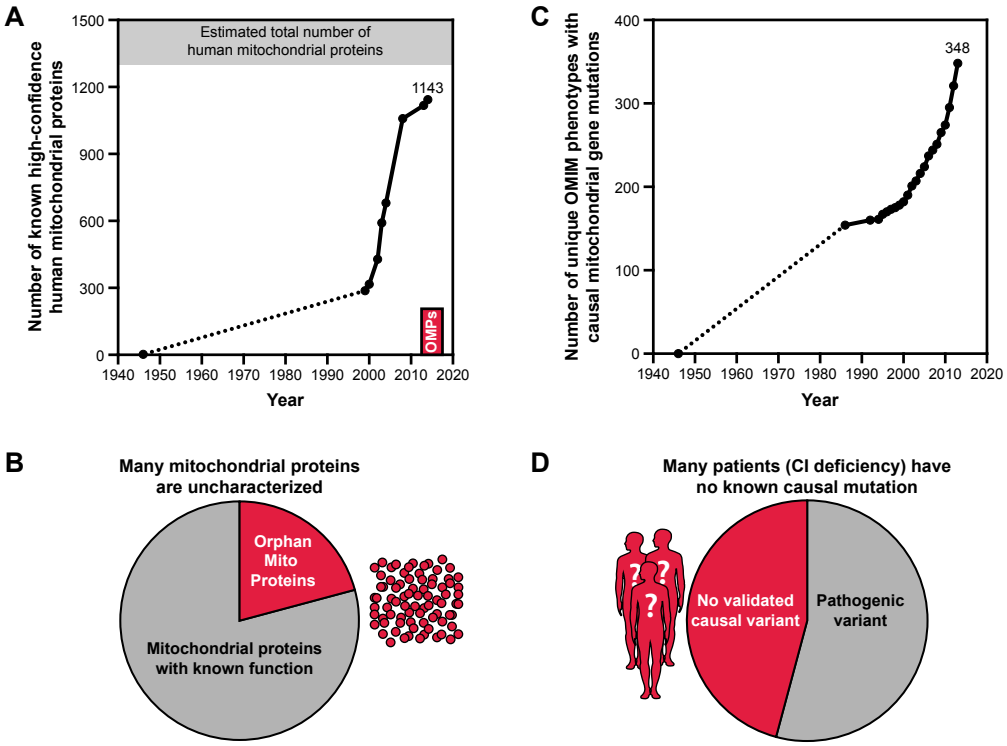


Figure 1.1: *Published mitochondrial proteins and diseases over time.* A) Number of known, high-confidence, resident human mitochondrial proteins over time. Data from [9–15]. Gray bar delineates the predicted total number of human mitochondrial proteins at 1300 [13] to 1500 [16]. The red bar shows the proportion of orphan mitochondrial proteins (OMPs) estimated to date. B) Orphan mitochondrial proteins, having no known function, comprise 20% of the mitochondrial proteome. C) Number of unique human diseases with known causal mutations in mitochondrial proteins over time. Data from OMIM [17]. D) Many patients with mitochondrial disease lack a genetic diagnosis, including nearly 50% of patients with overt complex I deficiency, according to a recent study [18].

The most comprehensive mitochondrial proteome was compiled in 2008 [13]. The integration of MS-based proteomic analysis of crude and highly-purified mitochondria from 14 mouse tissues, along with evolu-

tionary conservation, fluorescence microscopy, and other genome-scale datasets in a Bayesian framework, allowed for the establishment of the mouse mitochondrial proteome to approximately 85% completion. This study nearly doubled the known mammalian mitochondrial proteome and its list, termed the MitoCarta, serves as a benchmark to this day.

Additional work has further expanded the known mitochondrial proteome by 87 members, while simultaneously revealing the sub-mitochondrial localization of a large number of these proteins [14, 15]. The current breadth of coverage stands at 1,143 unique mitochondrial proteins by our count. This number is consistent with earlier predictions that the true size of the mitochondrial proteome is approximately 1,300-1,500 proteins, as estimated based on 2-D gel and computational analyses [13, 16].

This nearly-complete list of mitochondrial proteins achieves three key goals. First, it acts as a filter, placing bounds on the terrain to be mapped by subsequent mitochondria-related studies (Section 1.3). Second, it offers guidance in the selection of targets for organelle-wide functional profiling (Section 1.4). Third, it reveals unexplored regions of the proteome that warrant further characterization (Section 1.5). MS-based proteomic methods have had great success in catalyzing the annotation of such uncharacterized proteins (Section 1.6). In sum, the compendium of mitochondrial proteins has served as an invaluable map of the known and unknown, and will continue to chart the course of studies for years to come.

1.3 Separating the wheat from the chaff

Identification of mitochondrial gene mutations

Owing in part to its small size, mitochondrial DNA (mtDNA) was the first human genome ever sequenced [22]. This enabled the identification of disease-causing variations in all 13 mtDNA-encoded genes [4]. As these genes encode for just 1% of all mitochondrial proteins, it prompted investigators to search for mutations in nuclear DNA (nDNA)-encoded genes, as these account for the remaining 99% of mitochondrially-localized proteins.

While the impact of mutations in mtDNA-encoded proteins was clear, the path forward was difficult for nDNA-encoded proteins. The cost and difficulty of gene sequencing has, until quite recently, prohibited the analysis of the entire genome of a patient. Instead, use of pathway-specific lists limited the search-space and enabled the identification of many causal mutations in such tractable pathways as beta-oxidation and oxidative phosphorylation [23].

It became clear, however, that many mutations exist in genes not yet associated with the disease-related pathway. As such, and to allow for the identification of novel mutations in unexpected genes, subsequent work has widened the net to include genes encoding for all mitochondrial proteins and their regulators. Only recently has it become feasible to examine all nuclear-encoded mitochondrial genes in patient samples. Aided by

advancements in exome sequencing technologies as well as the catalog of mitochondrial proteins, several studies have applied mitochondria-focused or genome-wide exome sequencing in a search for variants.

The mitochondrial proteome has served as an important filter, both for the selection of genes to sequence as well as for the prioritization of variants in follow-up analyses. Using the list of mitochondrial proteins as targets is another successful way to prioritize genes while still allowing for the identification of mutations in unexpected genes. Based on the list of mitochondrial protein-encoding genes, several studies have assessed variation in all known mtDNA and nDNA-genes encoding mitochondrial proteins, termed MitoExome sequencing [18, 24, 25]. In this way, nearly double the number of patients were given molecular diagnoses compared to the standard of care at the time [18].

Even with the advent of faster and more economical whole-exome sequencing, the mitochondrial proteome still plays an important role in prioritizing gene variants of patients with mitochondrial disease, as mutations in mitochondrial genes are of greatest likelihood in explaining the disease [26–28]. Despite these advances, almost half of patients with biochemically-validated mitochondrial disease have no known disease-causing mutations [18, 23], emphasizing the need for deeper exploration of the components and function of mitochondria, for these may directly contribute to the diseases lacking molecular diagnoses.

Mitochondria-focused proteomics

As with gene sequencing, the MitoCarta has played a similar role of target prioritization for proteomics studies. Akin to the history of gene sequencing, proteomics approaches and technology have expanded a great deal in the past decades. The greater sensitivity and speed of modern instrumentation, along with newly devised protocols, allows for the simultaneous interrogation of thousands of proteins and tens of thousands of unique peptides [29, 30].

With the knowledge of the number and diversity of mitochondrial proteins came with it renewed questions of how this organelle responds to states of stress, and how dysfunction may lead to disease. Using various models of cellular and organismal stress, researchers have applied quantitative proteomics to compare the mitochondrial proteome across contrasting states.

Just as gene sequencing has benefited from the use of the mitochondrial proteome as a way to filter the data and obtain an organellar focus, this has also been the case with proteomics studies. In many cases, the compendium of mitochondrial proteins has been utilized for the *in silico* identification and functional analysis of the mitochondrion. Use of the mitochondrial proteome as a filter has facilitated the global and mitochondria-specific analysis of mouse, rat, and human tissues, and the elucidation of the role of mitochondrial post-translational modifications in acute and chronic conditions like obesity, diabetes, caloric restriction, and

exercise [31–37] (See also Chapter 2). As an example, our lab determined that mitochondrial protein abundance is elevated in the liver of obese mice, and that this corresponds to activation of mitochondrial metabolism.

1.4 Systematic analyses of mitochondria

A second benefit of a well-defined mitochondrial proteome is that it can help guide systems-based approaches. Not only can data be filtered after the fact by looking for proteins specific to this organelle, but the list of mitochondrial proteins has proven useful for the design of experiments from the outset. Integration of the MitoCarta with phylogenetic profiling enabled the identification of several complex I related genes, termed Complex One Phylogenetic Profiling [13].

Several studies have looked at the mRNA expression of mitochondrial genes to find correlations among them [38, 39]. Recently, quantitative proteomics via stable-isotope labeling was integrated with RNA sequencing data to establish the temporal nature of RNA and protein turnover in infected cells, revealing that mitochondrial turnover via mitophagy plays an important role [40].

An integrative approach, combining the MitoCarta compendium with comparative physiology and evolutionary genomics of kinetoplastids and *S. cerevisiae*, was sufficient to focus the search for the calcium transporter upon just 13 candidate genes. RNA interference then narrowed the field

to a single protein, MICU1 [6]. While not the core calcium transporter, it was then used as a bait in further analyses that subsequently found the transporter MCU [41, 42]. Such an approach harnesses the power of exclusion, as it was presumed that this transporter was likely hiding in the list of known mitochondrial proteins.

Defining the MitoCarta has also allowed for the specific targeting of each mitochondrial gene product. In one study, siRNA oligonucleotides were designed against each mitochondrial transcript. Cells were then grown in each of four nutrient conditions, and the concentration of ATP per cell was measured after knockdown. This work identified two regulators of phosphorylation, the kinase AK4 and the phosphatase PPTC7, that also play an important role in cellular energetics [43]. Such an organelle-wide analysis would have been impossible without a nearly complete compendium of mitochondrial proteins.

1.5 Orphan mitochondrial proteins

A third benefit of the compendium of mitochondrial proteins is that it highlights organellar proteins of unknown function that are ripe for further characterization. Uncharacterized proteins are a problem for many facets of biomedicine, from the bench to the bedside. Mutations in uncharacterized genes offer little insight to physicians and patients. Pathways that lack enzymes to fulfill key steps are resistant to pharmacological manipulation.

Systems approaches such as Gene Ontology (GO) term analysis and other networks of functional annotation depend upon the characterization of all or most proteins in a dataset, so incomplete annotation leaves gaps and reduces the power of these analyses. And finally, without a protein to study, biochemists are unable to apply their tools to extend our understanding of the chemical transformations within our cells. As such, uncharacterized proteins form a bottleneck that impedes progress in a host of important areas.

In the process of defining the MitoCarta, it was determined that many mitochondrial proteins have no annotated function [13]. To re-evaluate the state of mitochondrial protein annotation, we recently performed in-depth analysis of each mitochondrial protein. First, we integrated the MitoCarta with subsequent proteomics-based studies of sub-mitochondrial localization to identify an up-to-date list of 1,143 genes encoding mitochondrial proteins [14, 15]. We then queried both large-scale and focused annotation efforts to assess the extent of functional characterization of each protein. Next, we hand-annotated of each of the 1,143 genes encoding mitochondrial proteins via careful analysis of the most recent literature. We observed that approximately 20% of all mitochondrial proteins have no known function, which we term orphan mitochondrial proteins (OMPs) [44] (See also Chapter 3). Upon integrating our list of the 209 OMPs with disease-associated genes culled from OMIM [17], Orphanet [45], and GeneCards/MalaCards [46, 47], we further identified at least 32 OMPs

associated with, and 20 causal for, human disease. These figures are likely underestimating the problem of OMPs, as many proteins are only superficially characterized, and the disease-relevance of OMPs is often revealed alongside their annotation.

A key counterpoint to the proteins of unknown function are the patients of mitochondrial disease with no genetic diagnosis. In two studies of patients with biochemically-confirmed complex I deficiency, a molecular diagnosis was elusive for nearly 50% of patients with mitochondria-focused gene or exome sequencing [18, 23]. One explanation for this discrepancy is that mutations underlying undiagnosed disease may lie in OMPs. These assessments of the lack of characterization of both mitochondrial proteins and diseases have shown that our knowledge of mitochondrial proteins and their functions is incomplete.

Together, the annotation of protein function and causal mutations has identified gaps in our knowledge of mitochondrial components and their role in disease. Clearly what is needed is a way to efficiently annotate the remaining regions of the proteome, as this will help catalyze systematic biomedical research efforts (Fig. 1.2).

1.6 Protein-protein interactions

One common method for systematically annotating uncharacterized proteins relies on the analysis of interaction networks based on physical or

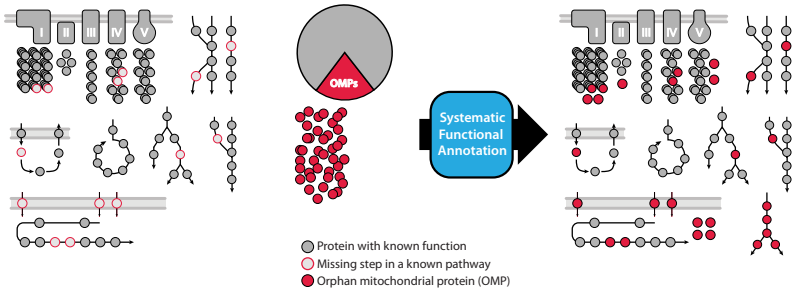


Figure 1.2: *Systematic approach to uncharacterized proteins.* Left: While much is known about the pathways within the mitochondrion (filled gray circles), many gaps in knowledge persist (open red circles). Center: Simultaneously, 20% of the mitochondrial proteome is comprised of orphan proteins (OMPs, filled red circles). Right: Systematic approaches are needed to fill in pathway gaps with OMPs

genetic interactions between proteins or genes. The network of protein-protein interactions (PPIs), termed the interactome, has facilitated the functional annotation of uncharacterized proteins [48]. The revelation that proteins of a given pathway interact with an uncharacterized protein can facilitate the functional annotation of the latter. This "guilt-by-association" approach has successfully connected genes with diseases and functions many times over the past two decades [48–51].

Organized units below the mitochondrial level

Mitochondrial protein-protein interactions have been known for more than 60 years, when it was surmised that the surprisingly large sizes of PDH and α-KGDH may indicate the existence of "organized units below the mitochondrial level" [52]. Interestingly, many of the techniques for the

purification and analysis of native, active complexes were developed for the complexes of oxidative phosphorylation (OxPhos) [53, 54], molecular machines that are quite familiar to any student of biochemistry. Despite this long history of study, many mitochondrial proteins lack known PPIs, which is likely an artifact of limited or nonexistent study rather than a biological reality [55]. Yet, when they are known, PPIs have served an important role in our understanding of mitochondrial function and in the annotation of OMPs.

Network analysis of mitochondrial proteins via several types of data has been predictive of protein characteristics and functions. In one example, a mitochondrial network was built based on correlated protein and transcript abundance overlaid on known protein-protein and functional interactions [56]. The positions of nodes within this network were found to correlate with protein sub-organellar localization and function, and the evolutionary emergence of new mitochondrial genes. Centrally-located nodes were found to be more ancient and localized to the mitochondrial matrix. Furthermore, network edges were sufficient to prioritize candidate disease-genes. Similarly, yeast mitochondrial PPIs were assembled into a network, facilitating the annotation of a set of uncharacterized proteins including several involved in lipid metabolism [51].

Orphan PPIs: the missing pieces

A key prior assumption for the mapping of functions via network approaches, however, is that the important genes are sufficiently captured in the underlying dataset. A sparse, incomplete dataset will hinder or even completely impede progress toward these aims. Thus, network-based approaches break down for a given uncharacterized protein if it truly lacks bioinformatic data. Recent analysis of the known human interactome, however, has revealed it to be woefully incomplete [57] and categorically biased against unstudied proteins [55]. This sparsity prohibits further computational annotation of uncharacterized proteins, as they exist as "islands," lacking connections to the greater mass of annotated proteins.

As such, there is a great need to expand the interactome in an organized fashion. While we are still a ways off from a complete human interactome, recent efforts which focus on just a subset of the proteome overcome the issue of sparsity by filling out a local region [58]. In the same way that the first step toward a systematic understanding of the mitochondrion was the cataloging of its protein components, so, too, will an OMP-focused PPI network catalyze annotation of uncharacterized proteins.

Once again, MS-based proteomics has played a significant role in the expansion of our knowledge of the interactome. Salient points regarding the history, current use, and considerations of proteomics-based network analysis will be discussed below. Chapter 3 will further discuss our efforts toward the annotation of uncharacterized mitochondrial proteins via their

PPIs.

Proteomic methods of defining protein-protein interactions: Affinity purification coupled to mass spectrometry (AP-MS)

Affinity purification coupled to mass spectrometry (AP-MS) has emerged as an important, systematic way to establish PPIs. Many large-scale AP-MS studies have utilized well-annotated baits as a way to extend their annotation to other uncharacterized proteins [59–64]. Work in press has utilized over 2000 bait proteins, most of known function, in the beginnings of a global mammalian PPI-screen. Another focused, HIV-centric study performed AP-MS with each of the 18 virally-encoded polypeptides to identify their host cell interactors [58].

Other AP-MS studies with mitochondrial protein baits have successfully linked a small number of OMPs to key cellular pathways like cardiolipin metabolism [65, 66], the mitochondrial ribosome [67, 68] and OxPhos [69]. Thus, the method of AP-MS is efficient at connecting uncharacterized proteins to novel annotations, thereby catalyzing subsequent focused biochemical analysis of these proteins.

AP-MS considerations

Early on the mere presence of a certain protein in an enriched sample was considered sufficient to identify it as an interactor. With time and greater sophistication, it was determined that many of these interactions were spurious, likely due to common contaminant proteins. This marked the need for quantitative data collection and scoring systems to isolate high-confidence interactions from among the mass of uninformative background.

Background, nonspecific, and uninformative interactions continue to plague AP-MS studies. The greater sensitivity of modern MS instruments has proven to be a double-edged sword, for while significant interactors of low abundance can now be observed, these come alongside other non-specifically interacting proteins. Indeed, in some studies, more than 90% of the protein mass in the "purified" sample corresponds to nonspecific and therefore uninformative prey proteins (see also Chapter 3) [70].

This issue of background is recognized as a widespread problem in AP-MS studies, and that one of the best tools to overcome this is the quantitative nature of mass spectrometry. There are several methods to quantify interactors, coarsely divided into labeling and label-free approaches. Label-free methods such as the counting of peptide spectral matches (PSMs) were some of the earliest to be put to use. More recently, *in vivo* labels and *in vitro* tags have facilitated the direct comparison of purifications in the same MS run. With new methods of label-free quantification, the ap-

proach to quantifying AP-MS has come full-circle. In particular, extracted ion current-based quantification, such as the MaxQuant MaxLFQ system, has been shown to have the cost and ease-of-use benefits of PSM counting with improved accuracy and dynamic range [71].

Just as there are several ways to quantify prey proteins, so too are there many scoring systems to differentiate between high- and low-confidence interactions. Most of these systems rely on some approximation of the specificity, reproducibility, and abundance of each potentially interacting protein. This topic has previously been covered in depth [72–77], and an exhaustive list is beyond the scope of this review, however, a brief overview of some of the more popular tools will be provided. The MiST algorithm was designed to elucidate virus-host interactions which are predicted to have little overlap, and thus prioritizes specificity over almost all other characteristics [58, 78]. CompPASS, on the other hand, takes specificity into consideration, but recognizes that high-confidence interactors are also likely to be highly-abundant in the enriched sample [59, 60, 69, 79, 80]. SAINT, a third commonly used method, uses mixture modeling and Bayesian statistics to generate distributions of true and false interactions, and then estimates the probability of a *bona fide* interaction [81–83].

There are many other important considerations which have a dramatic impact on the results of these experiments. These include but are not limited to: the method of bait enrichment, the presence, position, and nature of the tag, whether to overexpress the bait protein or enrich for

endogenous levels, the cell type, the lysis and purification buffer system, the method of elution, and MS sample preparation. While no best method has risen to the top, different approaches may yield non-overlapping yet still biologically relevant results [77, 84].

Proteomic methods of defining protein-protein interactions: Beyond AP-MS

While AP-MS has been quite successful, the method has certain distinct drawbacks. As bait proteins need to be enriched or purified, a consistent scheme is necessary, so investigators often tag and overexpress their protein of interest. This can, of course, lead to spurious interactions and localization via unfolding or the occlusion of binding sites. Additionally, the cellular system must be tractable for such manipulations, often meaning only cell-culture can be used. To combat these issues, several other methods of enriching for and identifying protein-protein interactions have come to the fore.

Recently, proximity labeling techniques have facilitated the identification of sub-organellar proteomes. APEX, a modified peroxidase, was modified to activate phenyl-biotin, which then can label cellular proteins in the nearby environment [85]. Fusion of APEX to a mitochondrial matrix protein allowed the specific labeling, isolation, and proteomic analysis of the mitochondrial matrix proteome [14]. Additionally, because of the

impermeable nature of the mitochondrial inner membrane, transmembrane proteins were only labeled on the matrix face. This allowed the establishment of the spatial orientation of these proteins. This technique has subsequently been used to elucidate the proteome of the mitochondrial inter-membrane space [15]. Together, these studies localized 51% of known mitochondrial proteins to submitochondrial compartments, while simultaneously identifying 87 new mitochondrial proteins.

Not only can proximity labeling identify proteins localized within a sub-organellar compartment, but it can also be applied to the identification of *in vivo* PPIs. In the BioID approach, a variant of the BirA biotin ligase engineered for increased promiscuity is fused to a protein of interest . This fusion protein then biotinylates the local environment, thereby labeling binding partners and nearby proteins. When BirA was fused to Lamin A, other nuclear envelope proteins were identified, including a novel member of this compartment [86]. Fusion of BirA to yeast Om45p, a protein localized to the outer mitochondrial membrane, facilitated proximity-specific ribosome profiling to assess the dynamic nature of mitochondrial protein synthesis [87]. A recent analysis has shown that the techniques of AP-MS and BioID are complementary and identify biologically relevant unique and overlapping PPIs [84].

The observation that interacting proteins maintain correlated solubilities across certain methods of native protein separation has led to several large-scale methods of complex identification. In the complexome pro-

filing approach, a whole-cell or organellar lysate is separated via native PAGE. Proteins from each gel slice are identified and quantified by MS-based proteomics [66, 69, 88, 89]. As with protein correlation profiling, the abundance of a given protein in each slice is used to generate a profile, and proteins with correlated abundance profiles are potentially members of the same complex. This method has its roots in several decades of mitochondrial complex native PAGE methods that efficiently separates each of the OxPhos complexes and supercomplexes, each maintained in an active state [53, 54].

Several groups have successfully applied complexome profiling to identify functions for orphan mitochondrial proteins. The protein APOOL was found to interact with the mitochondrial contact site (MICOS) complex, thus implicating its function in the regulation of mitochondrial contacts and cristae junction formation [66]. TMEM126B was found to be a complex I assembly factor (CIAF), and interacts with several other CIAFs to form the MCIA complex [88]. Another study further characterized the mitochondrial ribosome and identified a new ribonucleotide complex of five proteins [89].

Complexome profiling has several advantages over AP-MS. In one experiment, thousands of proteins across hundreds of complexes can be interrogated simultaneously. The protein of interest need not be overexpressed. This technology is compatible with membrane-associated proteins, especially when using non-ionic and weak detergents. Unfortunately, as the

enrichment is limited in comparison to AP-MS, proteins of low abundance may escape the limits of detection [20].

Recently, the Harper group combined AP-MS with complexome profiling to establish TIMMDC1 as an assembly factor of complex I (CI) of oxidative phosphorylation [69]. First, they expressed tagged versions of several members of CI and found them to interact with TIMMDC1. Then, they observed that TIMMDC1 was sufficient to enrich for many of the CI subunits. Finally, they applied complexome profiling to wild-type and TIMMDC1 knock down cells. Cells lacking TIMMDC1 had a reduction in the abundance of mature CI, and a build-up of an immature CI subcomplex, consistent with its role as a CI assembly factor. Thus, the benefits and drawbacks of each method can be used to suggest and validate a functional annotation for an orphan protein.

Similar to complexome profiling, a recent study applied biochemical protein fractionation to the entire cell to identify co-appearing proteins within more than 1,000 soluble fractions [90]. Each fraction was analyzed by quantitative MS to generate protein profiles as in PCP and complexome profiling, and correlated profiles suggest interacting proteins. This work identified hundreds of new complexes, many containing five or fewer subunits. They were also able to reconstruct the stoichiometries of certain complexes, including the proteasome. Furthermore, this approach requires limited *a priori* knowledge of the cellular system, and is amenable to the application to understudied or unwieldy cell or tissue samples.

1.7 Conclusion

The recent advent of a nearly-complete list of mitochondrial proteins has been a boon to all methods of understanding this organelle. This information has been successfully used to prioritize results after a large-scale experiment, to guide the selection of targets when planning, and to identify uncharacterized proteins for further analyses. Mass spectrometry-based proteomics has played a large role in the generation and application of these data, and this will likely continue as focus turns from the identification of organellar proteins to their network interactions as with AP-MS. Finally, these systematic approaches have been proven in the mitochondrion and will almost certainly be as effective in other organelles and across the cell. Given the importance of PPIs and mitochondrial proteins to health and human disease, it is likely that these methods laid out here will have a great impact on biomedical research and, in time, patient care and disease management.

2 CHAPTER TWO

Quantification of Mitochondrial Acetylation Dynamics Highlights Prominent Sites of Metabolic Regulation

Amelia J. Nestler*, Brendan J. Floyd*, Alexander S. Hebert*, Craig A. Bingman, Joshua J. Carson, Drew R. Gunderson, Brendan K. Dolan, Paul A. Grimsrud, Kristin E. Dittenhafer-Reed, Donald S. Stapleton, Mark P. Keller, Michael S. Westphall, John M. Denu, Alan D. Attie, Joshua J. Coon, and David J. Pagliarini

2.1 Abstract

Lysine acetylation is rapidly becoming established as a key post-translational modification for regulating mitochondrial metabolism. Nonetheless, distinguishing regulatory sites from among the thousands identified by mass spectrometry and elucidating how these modifications alter enzyme function remain primary challenges. Here, we performed multiplexed quantitative mass spectrometry to measure changes in the mouse liver mitochondrial acetylproteome in response to acute and chronic alterations in nutritional status, and integrated these data sets with our compendium of predicted Sirt3 targets. These analyses highlight a subset of mitochondrial proteins with dynamic acetylation sites, including acetyl-CoA acetyltransferase 1 (Acat1), an enzyme central to multiple metabolic pathways. We performed *in vitro* biochemistry and molecular modeling to demonstrate that acetylation of Acat1 decreases its activity by disrupting the binding of coenzyme A. Collectively, our data reveal an important new target of regulatory acetylation and provide a foundation for investigating the role of select mitochondrial protein acetylation sites in mediating acute and chronic metabolic transitions.

2.2 Introduction

Protein acetylation was first identified 50 years ago [91, 92] and has since become established as a key regulator of gene transcription through its role in chromatin remodeling. In recent years, fueled by technological advances in mass spectrometry-based proteomics, lysine acetylation has become recognized as a widespread post-translational modification (PTM) that rivals phosphorylation and ubiquitination in its prevalence [93–95]. Metabolic enzymes, including many in mitochondria, are among the most heavily acetylated proteins [96–98]. Mechanistic and physiological studies have begun to reveal that acetylation of these proteins plays a key role in regulating metabolic flux, and to suggest that aberrant levels of this PTM might contribute to disorders associated with metabolic inflexibility, including obesity, type 2 diabetes, and cancer [99–105].

Despite mounting evidence that acetylation is a prevalent protein modification, much work remains to establish it as a pervasive mitochondrial regulatory mechanism. This includes the need to distinguish *bona fide* functional acetylation sites from what may be widespread spurious or adventitious modifications [106]. This is highlighted by the considerable gap between the large number of identified mitochondrial lysine acetylation sites (~2,200) and the few with a validated regulatory function (~12 proteins) [107]. The reasons for this disparity stem in part from both the relative youth of the field and from technical limitations of most previ-

ous mass spectrometry-driven investigations, which were largely unable to produce accurate quantitative information on how acetylation levels change between contrasting biological states.

Here, we applied our newly developed quantitative acetyl-proteomics methods to measure changes in the mouse liver mitochondrial acetylproteome between two pairs of contrasting nutritional states [32]. First, because PTMs are a primary mechanism of regulation at acute time scales, we assessed changes in the mitochondrial acetylproteome between fasted and refed mice. Second, because chronic dysregulation of PTMs can be an important factor in disease, we compared the mitochondrial acetylproteome of obese (Leptin-deficient, *Leptin*$^{ob/ob}$) and lean (*Leptin*$^{+/+}$) C57/Bl6 (B6) mice. Our analyses reveal that, among a background of largely static acetylation sites, 237 mitochondrial proteins have a significant change ($q \leqslant 0.1$, Welch's t-test with Storey correction) in the relative occupancy of one or more acetylation sites during these transitions in nutritional status. These include a range of enzymes from core mitochondrial metabolic processes, such as oxidative phosphorylation, fatty acid β-oxidation, branched chain amino acid metabolism, and ketogenesis.

To further prioritize mitochondrial lysine acetylation sites that are the most likely to be important for metabolic flexibility, we integrated these large-scale data sets with our compendium of predicted targets of the mitochondrial deacetylase, Sirt3 [32]. Together, these analyses highlighted a select set of mitochondrial proteins as being likely targets for regula-

tory acetylation, including acetyl-CoA acetyltransferase 1 (Acat1), which possessed several sites of highly dynamic acetylation. By performing site-specific acetyllysine incorporation and *in vitro* biochemical enzyme activity assays, we discovered that Sirt3-driven deacetylation of K260ac and K265ac significantly enhances Acat1 activity. Molecular modeling revealed that the inhibitory effects of acetylation on Acat1 activity are likely due to decreased affinity for CoA caused by a loss of a favorable electrostatic interaction between one or more positively charged lysines with the negatively charged 3'-phosphate of CoA. Collectively, our work reveals Acat1 as an important new metabolic target of reversible acetylation and provides a resource of prioritized acetylation sites for future investigations into the role of this PTM in regulating mitochondrial protein function [31].

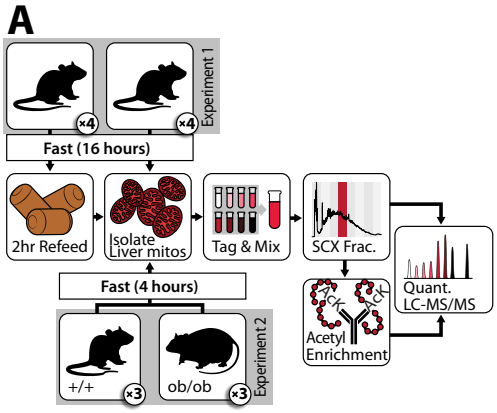
2.3 Results

Quantitative analysis of mitochondrial acetylation dynamics during fasting and refeeding

To quantify the dynamics of mitochondrial acetylation with high resolution, we leveraged our recently developed strategy for efficient acetyl peptide enrichment [32] and our QuantMode method [108] for highly accurate quantitative proteomics with isobaric tags to profile mouse liver mitochondrial protein abundance and acetylation levels across the transi-

tion between fasting and refeeding. Eight mice were fasted overnight for 16 h, after which half were allowed to feed *ad libitum* for 2 h (Fig. 2.1A), a time point featuring marked PTM changes with limited protein abundance changes [31]. We identified 1,915 unique acetylation isoforms, of which 1,796 were quantified (Fig. 2.1B and Fig. 2.3). Unsupervised average linkage hierarchical clustering of the data based on relative acetyl isoform occupancy (site-specific acetyl fold change divided by protein abundance fold change) grouped the refed and fasted mice separately (Fig. 2.3A), indicating that acetylation changes are a reproducible and distinguishing feature of the transition between fasting and refeeding [109, 110]. Importantly, our quantitative approach reveals that ~10% (188) of identified acetyl isoforms exhibited a statistically significant change ($q \leqslant 0.1$) in acetyl occupancy in response to this acute perturbation of metabolic status.

We hypothesize that the subset of mitochondrial protein acetylation sites with changing occupancy upon refeeding represent those most likely to have a regulatory function during this transition. For example, we detect acetylation on all four mitochondrial members of the malate-aspartate shuttle (Mdh2, Got2, Slc25a11, and Slc25a12), which is responsible for relaying the reducing potential of NADH generated from glycolysis into the mitochondrial matrix (Fig. 2.3B) [111]. Rat liver Mdh2 activity is increased upon feeding [112], and previous studies have reported that Mdh2 activity can be increased either by acetylation (on Lys-185, Lys-301,



B

			IDs	Quant	Δ
Refed/Fasted	Proteins	Total	1980	1514	458
		MitoCarta	**635**	632	**193**
	Acetyl-proteins	Total	481	453	98
		MitoCarta	**329**	324	**94**
	Acetyl-isoforms	Total	1915	1796	188
		MitoCarta	**1568**	1495	**185**
Obese/Lean	Proteins	Total	3513	3189	2623
		MitoCarta	**722**	745	**632**
	Acetyl-proteins	Total	605	580	389
		MitoCarta	**277**	271	**212**
	Acetyl-isoforms	Total	1723	1601	826
		MitoCarta	**1020**	975	**526**

Figure 2.1: *Schematic of workflow for quantitative proteomic and acetylomic analyses of mouse liver mitochondria.* A) in experiment one, eight lean mice were fasted for 16 h and either sacrificed immediately or refed for 2 h before being sacrificed. In experiment two, three obese mice and three lean mice were fasted for 4 h before being sacrificed. In each case, liver mitochondria were enriched and analyzed by LC-MS/MS. SCX Frac., strong cation exchange fractionation. B) summary of protein and acetylation data. IDs, unique identifications at 1% false discovery rate; Quant, measurements quantified with isotopic reporter ions in all mice assessed; Δ, measurements significantly changing with $q \leqslant 0.1$.

Lys-307, and Lys-314) [113] or by Sirt3-dependent deacetylation (of Lys-239) [32]. Here, we detect a large decrease in acetylation occupancy of Mdh2 Lys-239 with little or no changes to the other sites. This indicates that the Sirt3-controlled acetylation site (Lys-239) is likely the dominant regulator of Mdh2 during this acute metabolic perturbation, consistent with an increased flux of reducing equivalents from the cytosol into the mitochondria for energy production upon feeding.

Quantifying the magnitude of acetylation changes across an acute time scale can provide insights into key acetyl sites without the confounding

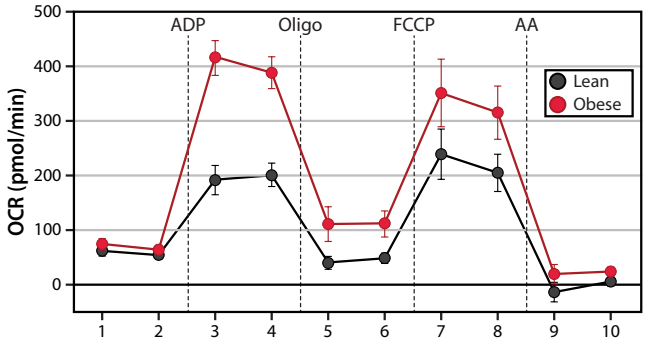


Figure 2.2: *Mitochondrial stress test.* Mitochondria isolated from either lean (black) or obese (red) mice were attached to a XF96 plate, and a stress test was performed in the presence of glutamate/malate. Black dashed vertical lines denote injections of indicated compounds. Each dot represents the oxygen consumption rate over the course of 3 min, and a minimum of ten wells were averaged for each condition. Error bars represent the standard deviation.

changes in protein levels that accompany longer perturbations (Fig. 2.3C). For example, despite unchanging protein levels among almost all enzymes involved in mitochondrial fatty acid oxidation (FAO), more than 90 distinct acetyl isoforms within this pathway are altered significantly upon refeeding (Fig. 2.3D). These include four acetylation sites on carnitine O-palmitoyltransferase 2 (Cpt2; Lys-104, Lys-544, Lys-537, and Lys-453), an enzyme involved in the transport of fatty acids from the cytosol into the mitochondrial matrix for β-oxidation [114]. Similarly, energy production from amino acid degradation is robustly induced in the postprandial state [115], and several enzymes in amino acid catabolic processes exhibit large changes in relative acetylation occupancy without alteration to the underlying protein abundance levels. As one example, acetylation on Lys-329 of

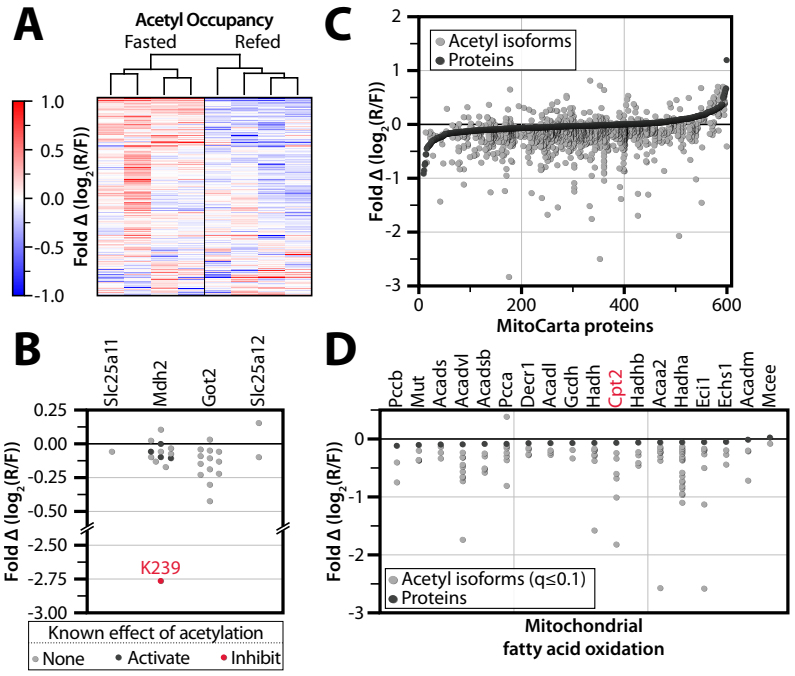


Figure 2.3: *Global analysis of acute protein and acetylation changes during the transition from fasting to refeeding.* A) Heat map showing unsupervised average linkage hierarchical clustering of four fasted and four refed mice groups them according to nutrient status. Values are colored based on relative acetyl occupancy, normalized to the average of all eight mice, on a $\log_2$ scale from $\leqslant$ -1 to $\geqslant$ 1. R/F, refed/fasted. B) Relative acetyl occupancy fold change for proteins in the malate-aspartate shuttle. The acetylation site previously shown to inhibit Mdh2 activity is shown in red, whereas those previously shown to activate Mdh2 are shown in black. Other sites we detected as acetylated but that have no prior functional data are shown in gray. C) Changes in acetylation are greater than changes in protein abundance. All quantified MitoCarta proteins are ranked on the x axis by protein abundance fold change, refed/fasted (black dots). Relative acetyl isoform changes are plotted in the same position on the x axis as the corresponding protein measurement (gray dots). D) Acetyl and protein abundance changes in mitochondrial fatty acid oxidation. Proteins are marked in black, and significantly changing ($q \leqslant 0.1$) acetyl isoforms are marked in gray.

glutaminase (Gls2) decreases >2-fold following refeeding, which is consistent with this PTM serving an inhibitory function on this key enzyme of amino acid breakdown. Overall, the vast majority of dynamic acetylation sites decreases in relative acetyl occupancy upon refeeding, consistent with previous reports that fasting can lead to hyperacetylation of liver mitochondrial proteins [116].

Quantitative analysis of mitochondrial acetylation dynamics due to chronic obesity

Previous studies have noted general changes to the acetylproteome during the onset of obesity [103, 117]; however, these studies were also limited by the semi-quantitative nature of their applied mass spectrometry methods. To evaluate and measure the key obesity-induced changes in acetylation of mitochondrial proteins with high accuracy and depth of coverage, we performed a large-scale quantitative comparison of the liver mitochondrial acetylproteome from three lean (*Leptin*$^{+/+}$) and three obese (*Leptin*$^{ob/ob}$) mice at 10 weeks of age, each fed standard chow diet *ad libitum* (Fig. 2.1A). This approach provides a comparison that is similar in nature to the first perturbation (contrasting nutrient abundance) but markedly different in time scale (hours versus weeks). Leptin deficiency in B6 mice is known to cause hyperphagia-induced obesity, which is accompanied by a dramatic response to over-nutrition (*e.g.* hyperinsulinemia, hepatic steatosis) by

10 weeks [118]. These obese mice, however, have only a mild, controlled elevation in blood glucose and model the prediabetic state, providing a model of obesity that is not confounded by diabetes. Here, we identified 1,723 unique acetyl isoforms, 1,601 of which were quantified, and 826 that were significantly changing ($q \leqslant 0.1$; Fig. 2.1B, and Fig. 2.4).

Not surprisingly, given the relative magnitudes and durations of the perturbations involved, the changes in protein abundance and acetyl occupancy due to chronic obesity encompass a greater dynamic range than those seen following a mere 2-h refeeding (Fig. 2.1B and Fig. 2.4, A and B). This comparison could reveal key acetylation events potentially important for regulating mitochondrial function in states of chronic overnutrition. For instance, it is known that the protein abundance and the activity of complexes in the oxidative phosphorylation (OxPhos) pathway are increased in certain models of obesity [31, 119], and acetylation has been implicated in inhibiting all five complexes of OxPhos [117, 120–123]; however, the specific regulatory sites are as yet unclear. To confirm the increase in OxPhos flux in our model of obesity, we isolated liver mitochondria from obese and lean livers and measured oxygen consumption using a Seahorse extracellular flux analyzer. As expected, mitochondria from obese mice exhibit an elevated state 3 oxygen consumption rate compared with their lean counterparts (Fig. 2.4D and Fig. 2.2). In our obese versus lean acetylome comparison, we identified 156 acetyl isoforms on OxPhos proteins (Fig. 2.4C). Ninety-seven of these sites change significantly (q

$\leqslant 0.1$) in acetyl occupancy, 91 of which decrease in the obese state. We therefore predict that a subset of these 91 sites (which we further prioritize below) plays an important role in the stimulation of OxPhos activity in obesity, irrespective of any protein abundance changes.

As with our fasting and refeeding study, we find that catabolic enzymes, including those of FAO and amino acid degradation, appear particularly susceptible to alterations in acetylation levels. For example, acetylation of Lys-42 on long chain acyl-CoA dehydrogenase (Acadl) is significantly decreased with obesity. This PTM has previously been shown to decrease enzymatic activity [124], consistent with elevated mitochondrial FAO in obese animals. Aminoadipate aminotransferase (Aadat), a protein involved in tryptophan degradation to kynurenine acid [125], has increased acetylation on Lys-263, which is involved in binding the pyridoxal phosphate cofactor [126, 127]. We predict that acetylation of Lys-263 reduces aminoadipate aminotransferase activity by preventing this enzyme-cofactor interaction. This decrease in activity could shift tryptophan degradation from the kynurenine-kynurenic acid pathway toward the kynurenine-NAD^+ pathway, which has previously been linked to hypertension, diabetes, atherosclerosis, obesity, and immunodeficiency [128].

Notably, for our proteomics analyses we purified mitochondria only to the extent required to achieve near comprehensive coverage of the liver MitoCarta protein list. (MitoCarta is a compendium of mitochondrial proteins [13].) As such, we were also able to profile the acetylproteomes of other co-

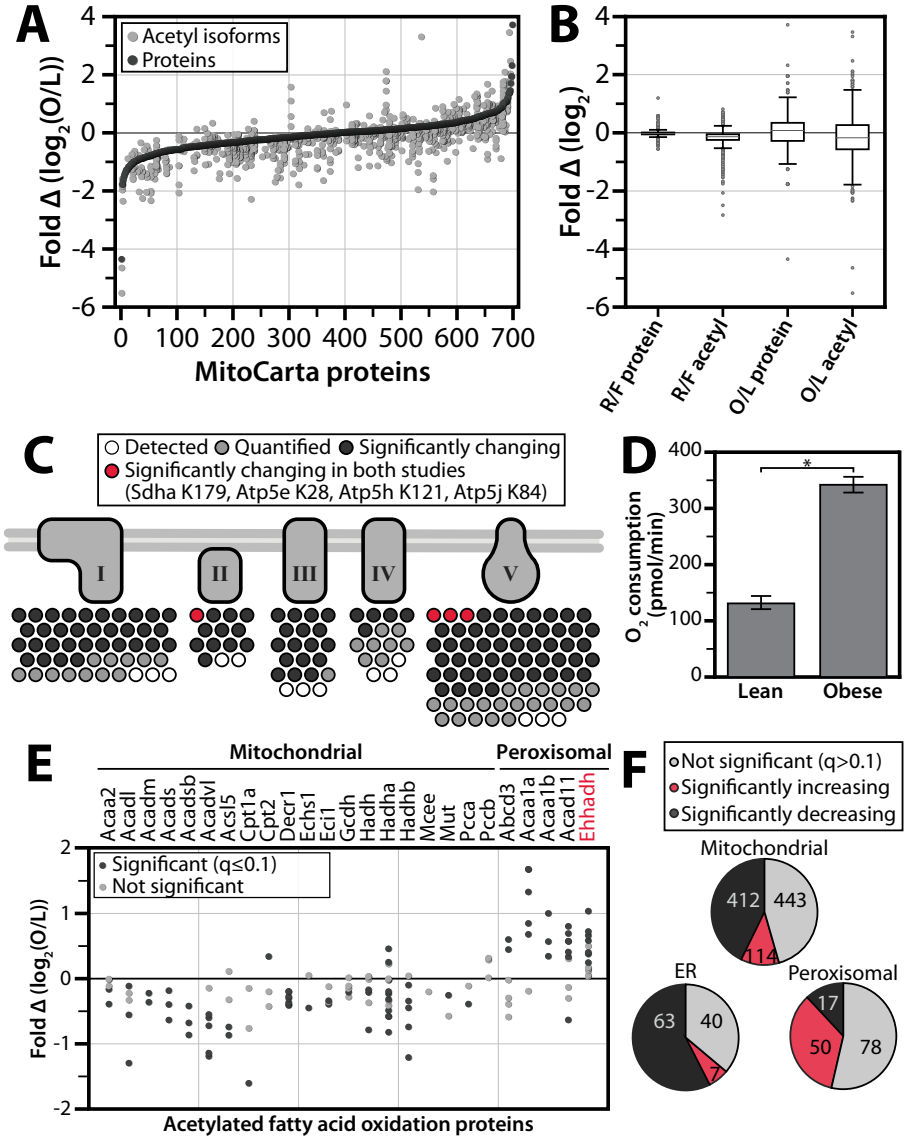

A
Acetyl isoforms
Proteins
Fold Δ (log2(O/L))
MitoCarta proteins
B
Fold Δ (log2)
R/F protein
R/F acetyl
O/L protein
O/L acetyl
C
Detected Quantified Significantly changing
Significantly changing in both studies
(Sdha K179, Atp5e K28, Atp5h K121, Atp5j K84)
I
II
III
IV
V
D
O2 consumption (pmol/min)
Lean
Obese
E
Mitochondrial
Peroxisomal
Aca2 Acadl Acadm Acads Acadsb Acadvl Acsl5 Cpt1a Cpt2 Decr1 Echs1 Eci1 Gcdh Hadh Hadha Hadhb Mcee Mut Pcca Pccb Abcd3 Aca1a Aca1b Acad11 Ehhadh
Significant (q≤0.1)
Not significant
Fold Δ (log2(O/L))
Acetylated fatty acid oxidation proteins
F
Not significant (q>0.1)
Significantly increasing
Significantly decreasing
Mitochondrial
412
443
114
ER
63
40
7
Peroxisomal
17
50
78

Figure 2.4: *Comparison of the acetylproteome response to acute and chronic perturbations.* A) All quantified MitoCarta proteins are ranked on the x axis by protein abundance fold change (black dots) for the chronic comparison of obese (O) and lean (L) mice. Relative acetyl isoform changes are plotted in the same position on the x axis as the corresponding protein measurement (gray dots). B) Dynamic ranges of protein and acetyl abundance are greater in obese/lean study than in refed/fasted (R/F) study. A box plot comparing the ranges of protein abundance and acetyl isoform fold change in the acute (refed/fasted) study versus the chronic (obese/lean) study. Whiskers represent data within 50% of the range of the first and third quartiles, and remaining outlier data is represented as a gray dots. C) Prioritization of acetylation sites in oxidative phosphorylation. Acetyl isoforms exhibiting significant changes ($q \leqslant 0.1$) in both studies are in red; those significantly changing in the obese versus lean study only are in black; those isoforms that are not changing significantly are in gray; isoforms that were identified but not quantified are in white. D) ADP stimulated increase in oxygen consumption rate is greater in mitochondria isolated from obese livers than from lean livers. An asterisk indicates significance (p-value = $3.9\text{x}10^{-14}$). E) Relative acetyl occupancy changes in fatty acid oxidation. Sites with significantly changing ($q \leqslant 0.1$) occupancy are marked in black, those that were quantified but not significantly changing in gray, and those that colocalize to peroxisomes are grouped. F) Acetylation by organelle. Pie charts show the number of mitochondrial, peroxisomal, or endoplasmic reticular acetyl isoforms from the obese/lean study that are not significant ($q > 0.1$) (gray), significantly increasing (red), and significantly decreasing (black).

purifying organelles, including peroxisomes and endoplasmic reticulum. Interestingly, FAO enzymes with evidence of mitochondrial/peroxisomal co-localization are generally increasing in relative occupancy with obesity (38 detected; 24 increase, one decreases, $q \leqslant 0.1$) (Fig. 2.4E). In comparison, acetylation sites on mitochondria-specific FAO enzymes are mostly decreasing (89 detected; three increase, 49 decrease, $q \leqslant 0.1$). Among the peroxisomal enzymes with increasing acetylation is Ehhadh (enoyl-CoA, hydratase/3-hydroxyacyl CoA dehydrogenase), which is activated by this PTM [113]. We detect increasing acetylation on seven lysines, with the greatest change on Lys-344, consistent with the known increase peroxisomal fatty oxidation (*i.e.* α-oxidation) in obesity[112, 129]. Overall, acetylation increases on most peroxisomal proteins in the obese state but decreases on most mitochondrial and endoplasmic reticulum proteins (Fig. 2.4F).

Integration of disparate data sets highlights key regulatory acetylation sites

The analyses above clarify the role of previously identified acetylation events and also quantify the acetyl occupancy of a wide range of uncharacterized acetylation sites. To further prioritize this extensive list for subsequent mechanistic studies, we compared the dynamic acetylation sites observed in both datasets. Of the thousands of mitochondrial sites

we identify across the two studies, only 48 change significantly ($q \leqslant 0.1$) due to both refeeding and obesity (Fig. 2.5C). We hypothesized that these isoforms likely represent key PTMs that serve as common mechanisms for mediating metabolic responses to a variety of different perturbations in nutrient status. Included in these are four of the OxPhos isoforms noted above (Sdha Lys-179 in complex II, and Atp5e Lys-28, Atp5h Lys-121, and Atp5j K84 in complex V), further highlighting them as high-likelihood regulatory sites from among the hundreds of OxPhos sites identified (Figs. 2.4C and 2.5C).

We reasoned that many *bona fide* regulatory acetylation sites would also be targets of Sirt3, the only well-established mitochondrial deacetylase [107]. Therefore, we further integrated our data sets with a compendium of predicted Sirt3 targets, which we recently developed from analysis of wild type and *Sirt3*$^{-/-}$ mouse liver mitochondria [32]. Sirt3 target sites are among those with the largest decrease in relative acetyl occupancy due to acute refeeding or chronic obesity, indicative of an induction of Sirt3 activity in these states (Fig. 2.5, A and B). Strikingly, putative Sirt3 targets represent over half of the acetyl isoforms that exhibited a significant change ($q \leqslant 0.1$) in relative acetyl occupancy in both studies despite accounting for <7% of all sites quantified. These results suggest a major, Sirt3-driven remodeling of the acetylproteome that is secondary to genetic obesity or to 2 h of refeeding after an overnight fast. Measurements for Sirt3 protein using both quantitative MS (Fig. 2.5B, inset) and immunoblotting

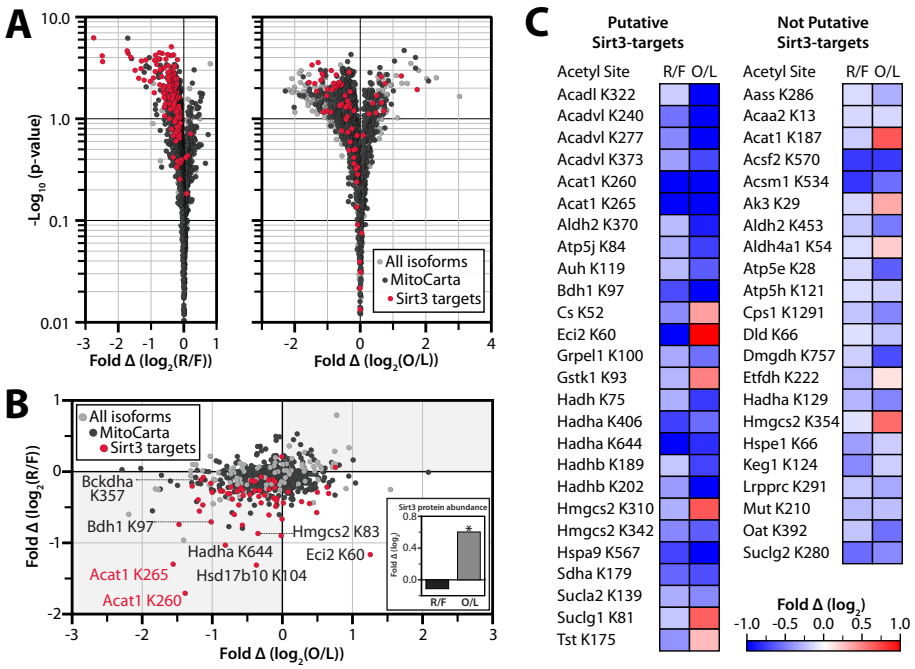


Figure 2.5: *Integration of different metabolic comparisons identifies candidate regulatory acetylation events.* A) Volcano plots of fold change in relative acetyl occupancy versus -$\log_{10}$ (p value). B and C) Comparison of the relative acetyl occupancy fold change across the refed/fasted versus obese/lean comparisons. Only isoforms that were quantified in both studies are represented. B) Putative Sirt3 targets are shown in red, other MitoCarta proteins are in black, and non-MitoCarta proteins are colored gray. Inset, Sirt3 protein abundance fold change by mass spectrometry. An asterisk indicates significance ($p \leqslant 0.05$). C) Acetylation sites that are significantly changing in both studies ($q \leqslant 0.1$). Fold changes ($\log_2$) are shown in red (positive) and blue (negative) for refed/fasted (R/F) and obese/lean (O/L).

(Fig. 2.6) indicate that Sirt3 abundance is increased in B6 10-week-old animals with obesity, consistent with our activity predictions. We note that this obesity-induced increase in Sirt3 abundance mirrors the induction of Sirt3 levels observed at early time points in the regimen of high fat diet feeding [103], as opposed to the decrease seen after many weeks of this diet. We did not observe a significant change in Sirt3 abundance following refeeding, despite the significant decrease in acetylation levels of predicted Sirt3 sites (Fig. 2.5, A-C, and Fig. 2.6). This is consistent with previous reports that alterations in Sirt3 activity are not always correlated with its protein abundance levels [117, 130, 131], suggesting that Sirt3 itself might be subject to post-translational regulation following acute perturbations. It is also consistent with indirect measures of mitochondrial redox state that suggest an increase in the NAD^+/NADH ratio in the refed state [132].

Among the dynamic Sirt3 targets observed in both large-scale analyses are multiple sites on enzymes from core metabolic pathways involved in responding to changing nutrient status. Three of these processes, branched chain amino acid metabolism, fatty acid catabolism, and ketogenesis, exhibit acetylation changes in many pathway members and have in common a single enzyme: acetyl-coenzyme A acetyltransferase 1 (Acat1, NCBI gene ID 110446) [133, 134]. Acat1 catalyzes two major reversible reactions: the cleavage of acetoacetyl-CoA into two acetyl-CoA, and the cleavage of 2-methyl acetoacetyl-CoA into propionyl-CoA and acetyl-CoA. We identified 18 total acetylation sites on this protein, nine of which are changing signif-

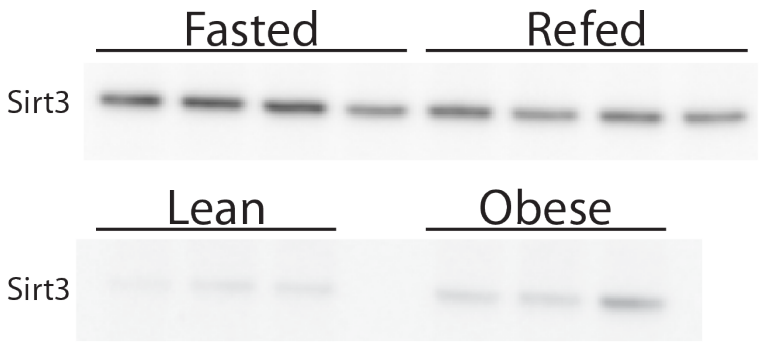


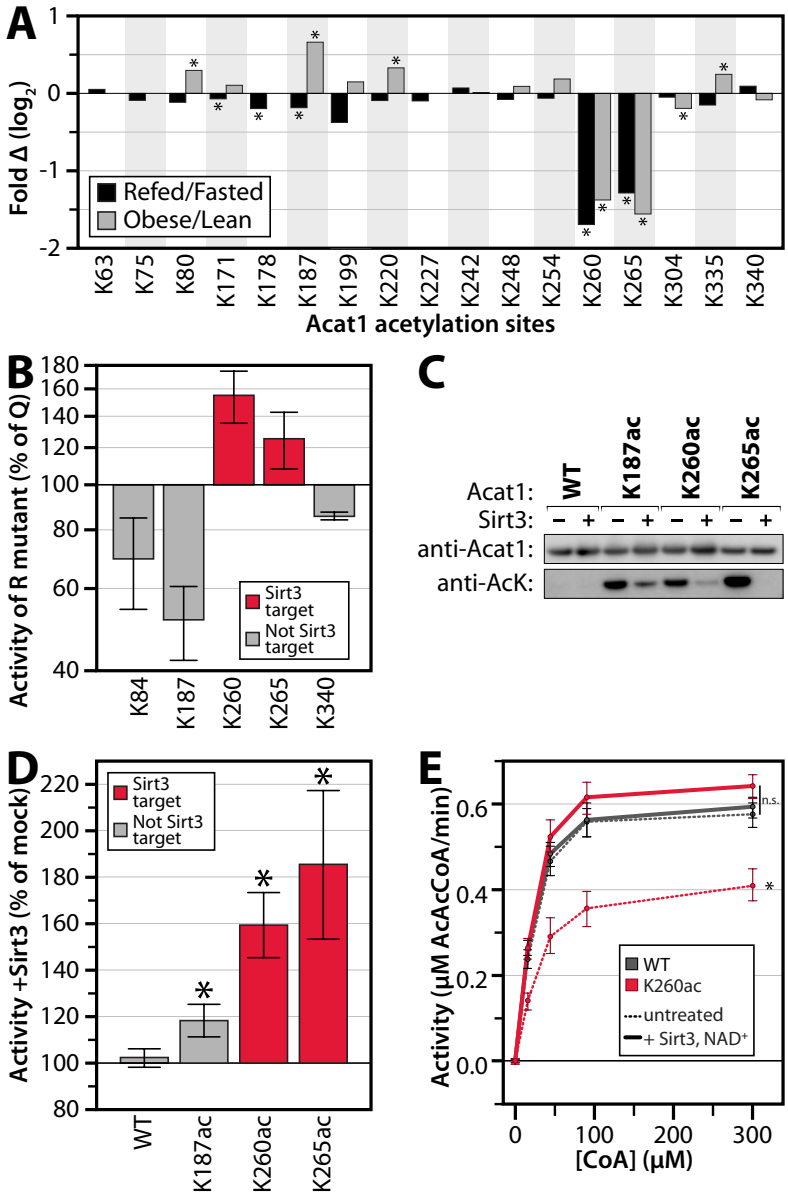
Figure 2.6: *Quantification of Sirt3 protein abundance.* Immunoblot of crude mitochondrial preparations for Sirt3 protein for each sample used in experiment 1 (top) and experiment 2 (bottom).

icantly ($q \leqslant 0.1$) in at least one comparison. Relative acetyl occupancy at three of these sites changed significantly in both analyses (Lys-187, Lys-260, and Lys-265) (Fig. 2.7A), the latter two of which are predicted Sirt3 targets. Another recent Sirt3 knock-out study also identifies Lys-260 and Lys-265, but not Lys-187, as Sirt3 targets [105]. In the present study, the acetylation levels of both putative Sirt3-target sites on Acat1 decrease markedly with obesity (relative to lean mice) and with refeeding (relative to fasted mice), whereas acetylation on Lys-187 increases due to obesity. These data led us to hypothesize that acetylation on Lys-187, Lys-260, and/or Lys-265 alters the activity of Acat1. Moreover, as Sirt3 most often activates its target enzymes [107], we further hypothesized that deacetylation of Lys-260 and Lys-265 increases Acat1 activity in response to nutrient excess.

Acetylation of lysine 260 and 265 on Acat1 inhibits enzymatic activity

Our analyses suggest that acetylation of select Acat1 lysines is important for regulating enzyme activity. To begin testing this possibility, we individually mutated five acetylated lysine residues (Lys-84, Lys-187, Lys-260, Lys-265, and Lys-340) to glutamine to mimic the charge state of an acetylated lysine or mutated to arginine to mimic a deacetylated state [135]. These sites were selected based on the structure of human ACAT1 [136] and on the dynamic changes seen in both data sets. Acat1 variants and wild type protein were purified from *E. coli* and tested by an *in vitro* activity assay. Mutation of lysines 260 and 265 to arginine (K260R, K265R) resulted in increased activity compared with the respective glutamine mutant (K260Q, K265Q), suggesting that acetylation would decrease activity (Fig. 2.7B). In contrast, K187R led to a loss of activity compared with K187Q. Interestingly, mutation of the other acetylation sites did not significantly change the activity of Acat1, further suggesting that Lys-187, Lys-260, and Lys-265 may be the most important acetylation sites for regulating the activity of this enzyme.

Although mutation of lysine to glutamine or arginine can approximate the effect of acetylation on protein function, these are imperfect analogs. To validate the ability of acetylation to regulate Acat1 enzymatic activity, we used site-specific acetyllysine incorporation, a technique that allows for the



F

Isoform	Treatment	V_{max} (μM/min)	$K_{M\ effective}$ (μM CoA)
WT	Mock	0.62 ± 0.06	17.75 ± 1.69
	+Sirt3, NAD+	0.63 ± 0.05	16.74 ± 1.94
K187ac	Mock	0.37 ± 0.03	22.20 ± 2.42
	+Sirt3, NAD+	0.44 ± 0.05 *	21.56 ± 2.22
K260ac	Mock	0.44 ± 0.07	24.44 ± 2.00
	+Sirt3, NAD+	0.68 ± 0.06 *	17.68 ± 1.89 *
K265ac	Mock	0.22 ± 0.06	22.06 ± 2.02
	+Sirt3, NAD+	0.38 ± 0.02 *	30.35 ± 2.63

Figure 2.7: *Reversible, Sirt3-regulated acetylation of Acat1 Lys-260 inhibits activity.* A) Relative acetyl occupancy fold change for 17 quantified sites on Acat1. An asterisk indicates $q \leqslant 0.1$. B) *In vitro* enzyme activity assays of Acat1 mutants generated by mutating lysine to glutamine (Q, acetyl mimic) or arginine (R, deacetyl mimic). Data are expressed as the average activity of the Arg mutant as a percentage of that for the Gln mutant, at 300 μM CoA (± S.E., $n \geqslant 2$). C) Immunoblot for Acat1 and acetyllysine on samples of purified Acat1 protein used in kinetic assays. Proteins were treated with or without NAD^+ and Sirt3. AcK, acetyllysine. C is representative of samples used in D. D-F), *in vitro* enzyme activity assays of Acat1 with site-specific incorporation of acetyllysine at the indicated residues, with or without deacetylase (Sirt3) treatment. Data are expressed as the average activity after Sirt3 treatment as a percent of mock-treated enzyme, at 300 μM CoA (± S.E., $n \geqslant 3$; an asterisk indicates a significant difference from mock treatment, $p < 0.05$). E) Aggregated enzyme kinetic analysis ($n = 4$; an asterisk indicates a significant difference from all other treatments, $p < 0.05$; n.s. indicates no significant difference between the highlighted treatments). F) Table of kinetic parameters of Acat1 isoforms ± Sirt3 treatment (± S.E., $n \geqslant 3$; an asterisk indicates a significant difference from mock treatment, $p < 0.05$).

production of a protein with nearly full acetylation occupancy on a specific lysine [137, 138]. Acat1 variants were generated and purified that each possessed a single acetylated lysine at position 187, 260, or 265 (K187ac, K260ac, K265ac). Acetyllysine incorporation was verified by immunoblot and MS analyses, and *in vitro* treatment of these proteins with recombinant Sirt3 effectively removed this acetylation in an NAD^+-dependent manner (Fig. 2.7C and Fig. 2.8). Interestingly, *in vitro* Sirt3 treatment more readily removed the acetyl groups on Lys-260 and Lys-265 than that on Lys-187, consistent with the prediction that K260ac and K265ac are Sirt3 targets *in vivo*.

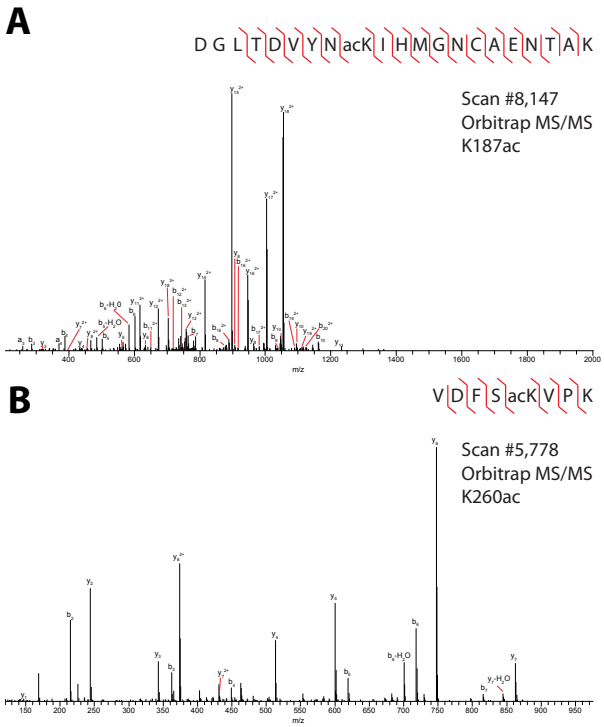


Figure 2.8: *MS/MS analysis confirms incorporation of acetyllysine at K187 (A) and K260 (B).* CAD spectra from isolated peptide precursors having monoisotopic m/z = 798.3714 (6 ppm mass error from theoretical m/z) (A) or 481.2728 (3 ppm mass error from theoretical m/z) (B). N- and C-terminal fragment ions matching theoretical fragments, within 0.01 Th, are annotated as b- or y- along with the corresponding amino acid fragment number and charge state.

In vitro Acat1 activity assays of each isoform revealed that Sirt3-mediated deacetylation of K260ac and K265ac dramatically increased enzyme activity in an NAD^+-dependent manner (Fig. 2.7 D-F and Fig. 2.9). In contrast, we observed minimal change in activity with identically treated wild type or K187ac proteins. We further analyzed the kinetic properties of the Acat1 K260 in its acetylated and deacetylated forms across several concentra-

tions of substrate and determined that acetylation decreases the V_{max} of the enzyme while increasing its K_M for CoA (Fig. 2.7F). As physiological levels of CoA in liver are markedly higher than the Acat1 K_M [139], these alterations in V_{max} will likely alter metabolic flux *in vivo*.

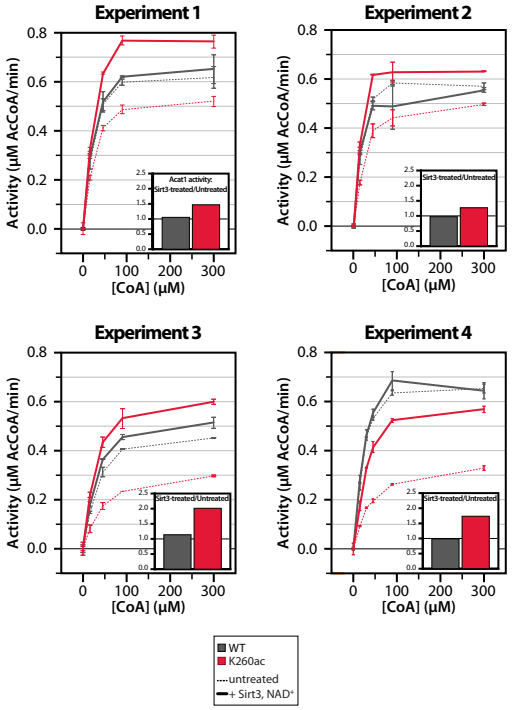


Figure 2.9: In vitro *enzyme activity assays of Acat1 with site-specific incorporation of acetyl-lysine at the indicated residues, with or without NAD$^+$ and deacetylase (Sirt3) treatment across four replicate experiments*. Proteins were treated with or without the combination of Sirt3 and NAD$^+$. Activity was assessed from 0-300 µM CoA. Inset: Data are expressed as the activity after Sirt3 and NAD$^+$-treatment as a percent of mock-treated enzyme, at 300 µM CoA.

Consistent with the hypothesis that regulatory residues are likely to be conserved throughout evolution [106, 140], Acat1 Lys-260 and Lys-265

are highly invariant and correspond to lysines 263 and 268 in the human ACAT1 ortholog (K263.Hs and K268.Hs). Recently, Haapalainen and colleagues [136] solved the crystal structure of human ACAT1 in complex with its CoA substrate. This structure reveals that positively charged K263.Hs is in close proximity to the negatively charged 3'-phosphate of CoA, which suggests that charge complementarity for CoA in the active site is important for substrate binding. Using this structure (Protein Data Bank code 2IBW) (Fig. 2.10), we modeled a neutralized lysine residue at K263.Hs to simulate the effect of acetylation (Fig. 2.10B). We then calculated electrostatic isosurfaces extending through space at +1 kT/e with all lysines fully charged (Fig. 2.10C), or with K263.Hs neutralized (Fig. 2.10D). This analysis clearly reveals that the negatively charged 3'-phosphate of CoA is present within the positively charged space enclosed by this electrostatic isosurface in the fully charged enzyme, but not when K263.Hs is neutralized. These data suggest that acetylation of Acat1 Lys-260 (ACAT1 K263.Hs) likely reduces the favorable electrostatic interactions between CoA and the Acat1 binding pocket. This is similar to reports of the Sirt3-dependent regulation of other enzymes by an electrostatic repulsion mechanism, including Mn-SOD [141, 142]. Together, in conjunction with our *in vitro* activity data, we propose that Sirt3 activates Acat1 by removing lysine acetylation that reduces the affinity of Acat1 for CoA.

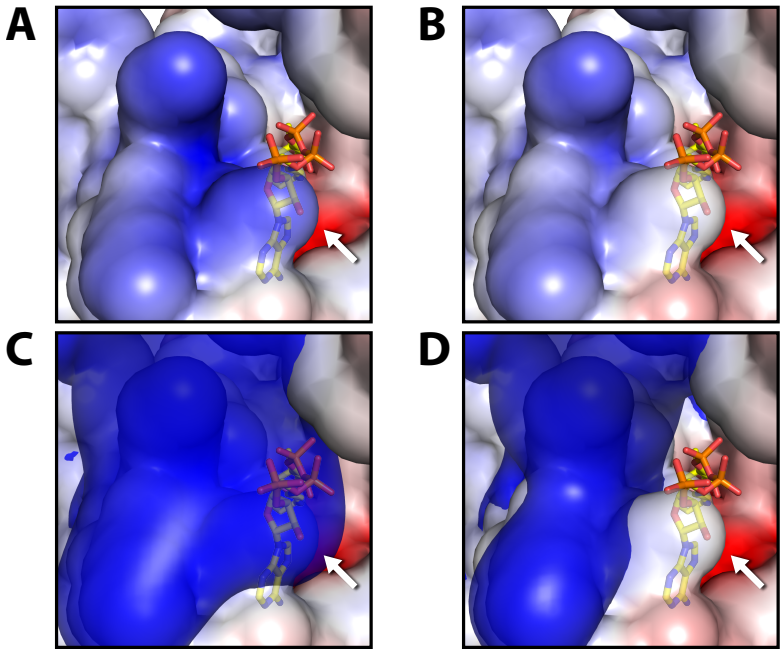


Figure 2.10: *Electrostatic modeling of acetylation effects using the crystal structure of human ACAT1 (PDB ID 2IBW).* Local electrostatic environment of ACAT1 active site surrounding K263.Hs (K260.Mm, identified by the white arrow). A and C represent wild-type ACAT1, whereas B and D mimic acetylation on K263.Hs by neutralizing its charge. The electrostatic environment is visualized as a continuous red-blue gradient from -4 to +4 kT/e on the solvent accessible surface of the protein. In C and D, it is overlaid with an electrostatic isosurface extending through space at +1 kT/e. The negatively charged 3'-phosphate of CoA is located within the space enclosed by the positive isosurface when K263.Hs is fully charged, but it is located outside when this lysine is neutralized.

2.4 Discussion

Reversible lysine acetylation is rapidly becoming recognized as a pervasive PTM that regulates cellular processes ranging from gene transcription to intermediary metabolism. The impact of this modification extends to

mitochondria, where a strikingly high percentage of proteins are acetylated on one or more lysines. A few of these sites are now known to modify metabolic flux by altering enzymatic activity or substrate availability, adding to the growing notion that PTMs are key to regulating mitochondrial function.

However, although thousands of mitochondrial acetylation sites have now been identified, little is known about the enzymes that control acetylation in mitochondria. There is only a single well established mitochondrial protein deacetylase, Sirt3 [107]. Likewise, there is no known acetyltransferase within this organelle, although emerging evidence suggests that GCN5L1 may be a regulator of a yet to be characterized protein acetyltransferase complex [143]. Further confounding this issue is the fact that even at low concentrations of acetyl-CoA, acetylation can occur nonenzymatically [144, 145]. Coupled with the knowledge that mitochondrial enzymes tend to become acetylated following nutritional changes that produce an increase in acetyl-CoA [146] and that there are a very limited number of validated regulatory sites, this has led to the speculation that many of the numerous mitochondrial acetylation events may be nonenzymatic and, thus, likely nonregulatory. As such, rigorous efforts are needed to spotlight and validate *bona fide* regulatory acetylation sites from a potentially noisy background of non-regulatory PTM "decorations" [147].

To prioritize candidate acetyl-regulated sites on mitochondrial proteins, we performed large-scale quantitative acetyl-proteomics across multiple

contrasting biological states. Our data reveal that only a select subset of sites significantly change in occupancy during drastic acute and chronic alterations in nutritional status. From our recent study [32], we predict that many of these dynamic acetylation sites are targets of Sirt3, consistent with the role of this enzyme as a primary regulator of mitochondrial acetylation. Together, our integrated analyses highlight a collection of proteins with high-confidence candidate regulatory sites (Fig. 2.5C). Although other acetylation sites may prove to be regulatory under distinct metabolic transitions, our data nonetheless caution that many others might represent unregulated and likely non-regulatory targets of acetylation.

A second priority in the field is to rigorously validate and define the functions of this modification in regulating protein activity [107, 148]. In this study, we investigated the importance of acetylation on Acat1. Our quantitative analyses nominated three of the 18 acetylated lysines on this protein as likely regulatory sites. Using site-specific acetyllysine incorporation, *in vitro* enzymatic analyses and molecular modeling, we demonstrated that acetylation of Lys-260 and Lys-265 inhibits Acat1 activity by disrupting CoA binding.

Acat1 is found at the cross-roads of branched chain amino acid degradation, ketogenesis, and fatty acid catabolism. Interestingly, more than one-third (19/48) of our high-confidence candidate regulatory sites are on proteins involved in these three pathways. As such, we hypothesize that the role of regulatory acetylation is not limited to isolated control points,

but rather that it impacts multiple enzymes of a pathway to collectively alter net flux, often alongside other PTMs. Indeed, other key regulators are known for each of these pathways (*e.g.* Bckdha phosphorylation [149], Hmgcs2 acetylation [150] and phosphorylation [31], and Acadl acetylation [124]). Interestingly, Acat1 Lys-265 was also recently identified as a prominent site of reversible succinylation, further suggesting that this is an unusually important site of post-translational regulation [35]. Based on our work here, we predict that this modification would likewise dampen Acat1 activity.

Although each of the three pathways mentioned above occurs in the liver, they are not simultaneously induced. For example, branched chain amino acid degradation is robustly induced by refeeding, as assessed by Bckdha phosphorylation status [31, 117] but fatty acid oxidation and ketogenesis are inhibited in that same state [133]. Conversely, fatty acid oxidation and ketogenesis are highly active in the obese state [151]. Thus, it seems reasonable to speculate that the overabundance of PTMs on ketogenesis, FAO, and branched chain amino acid enzymes noted above, including those on Acat1, operate combinatorially to help properly calibrate the individual fluxes through these interrelated but distinct pathways. Furthermore, although our biochemical data support a role for deacetylation of Acat1 in enhancing flux through any these pathways, its acetylation alone would likely not be sufficient for complete inactivation of any one of them. Further integration of our data sets with other protein PTM

data and quantitative metabolic flux analyses will be necessary to better define the physiological ramifications of these modifications in collectively modifying metabolic output.

The results of our analyses both mirror current trends in the acetylation field and also raise new issues. Consistent with previous work, we find that acetylation is widespread in mitochondria and that its levels generally tend to track with predicted acetyl-CoA abundance [146]. Our data also suggest that Sirt3 is more active in the refed state and early during the onset of obesity and that it tends to enhance the activity of its target enzymes. However, our data reveal that acetyl occupancy at several putative Sirt3-regulated sites (*e.g.* Eci2 Lys-60, Gstk1 Lys-93, Cs Lys-52, and Tst Lys-175) was significantly decreased due to refeeding but significantly increased due to obesity. This suggests that Sirt3 may possess an adjustable target preference under different nutritional states. Moreover, our data indicate that Sirt3 activity cannot necessarily be predicted from its abundance, implying that other regulatory post-transcriptional or post-translational features might be important for titrating its activity. Last, our data reveal that many probable non-Sirt3 sites also change considerably in acetylation occupancy, supporting the notion that other mitochondrial regulatory acetylation machinery awaits discovery.

Collectively, our work provides a quantitative map of the changing mitochondrial protein acetylation landscape during acute and chronic metabolic transitions, defines the role of deacetylation in stimulating Acat1

activity, and highlights many other high-priority acetylation sites for mechanistic follow-up work [31].

2.5 Experimental Procedures

Animal models

Breeding, sacrificing, and tissue harvesting of mice were described previously for the fasting versus refeeding study [31] and the obese versus lean study [152]. Briefly, male mice were bred and housed in an environmentally controlled facility with a 12-h light-dark cycle (6 am-6 pm light cycle) and provided water and standard rodent chow (Purina number 5008). For the fasting versus refeeding study, chow was removed for 16 h (5 pm-9 am), after which half of the animals were sacrificed, whereas the other half were allowed to feed *ad libitum* an additional 2 h before being sacrificed. Liver tissue was dissected and immediately enriched for mitochondria using differential centrifugation [31]. For the obese versus lean study, chow was removed for 4 h (8 am-12 pm) before sacrifice, and the liver was flash-frozen in liquid nitrogen before being thawed and enriched for mitochondria.

Quantitative acetylproteome analysis

Crude mitochondria were subjected to quantitative proteomics/acetylomics using recently developed methods [32]. Briefly, mitochondrial pellets were resolubilized, and the proteins were reduced and alkylated and digested enzymatically and labeled with isobaric tags (6-plex TMT or 8-plex iTRAQ). Samples were mixed and subjected to strong cation exchange chromatography, followed by immunoprecipitation with pan-acetyllysine antibody (Immunechem), and non-bound peptides were further fractionated by high pH reverse phase chromatography. All samples were subjected to data-dependent Nano-LC-MS/MS analysis on an ETD-enabled LTQ Orbitrap Velos (Thermo Fisher Scientific) using our previously described QuantMode instrumentation method [108] with higher energy collision-activated dissociation fragmentation. Our custom software COMPASS [153] was utilized to search the MS/MS spectra against a concatenated target-decoy mouse database (UniProt), to trim the identified peptides (1% false discovery rate), to normalize the intensities of isobaric tag reporter ions, to group all peptides to parsimonious protein groups (1% false discovery rate), to localize acetylation sites to specific residues (95% probability), and to sum reporter ion intensities for spectra identifying the same protein or acetyl-isoform.

Measurement of oxygen consumption rate

The oxygen consumption rate was measured using an XF-96 Extracellular Flux Analyzer (Seahorse Bioscience) with methods based on previous work [116]. Briefly, isolated mitochondria were quantified by wet weight, and 40 μg were loaded per well and supplemented with 10 mM glutamate and 2 mM malate. Basal respiration was measured for 3 min, followed by ADP injection to a final concentration of 2 mM. ADP-stimulated oxygen consumption was calculated by subtracting basal respiration, after which the wells were averaged. Data analysis was performed using the XF96 software (version 1.4.1.4) and the "Level(Direct)Akos" algorithm.

Biochemical assessment of Acat1 activity

Recombinant Acat1 (amino acids 31-424) with a C-terminal hexahistidine tag was overexpressed in *Escherichia coli* and purified using metal affinity resin based on previous methods [136, 154]. For site-specific acetyllysine incorporation, an acetyl-lysyl-tRNA synthetase/$tRNA_{CUA}$ pair that recognizes an amber codon was co-expressed with Acat1 in the presence of 10 mM acetyllysine and 20 mM nicotinamide [137, 138, 155]. Briefly, a plasmid encoding Acat1 was cotransformed with pBK AcKRS (acetyllysine tRNA synthase) and pCDF PyIT ($tRNA_{CUA}$) into BL21(DE3) *E. coli*. Cells were grown to an A_{600} of 0.5 to 0.7 and then supplemented with 10 mM acetyllysine and 20 mM nicotinamide for 30 min before the addition of 0.3

mM isopropyl 1-thio-β-*D*-galactopyranoside. Cells were purified using metal affinity resin as above with the inclusion of 20 mM nicotinamide in the lysis buffer. Incorporation was verified by mass spectrometry and immunoblots for Acat1 and acetyllysine. Sirt3 was purified as described previously [156].

Acat1 activity assays

Activity was assessed as described previously [136, 157]. Briefly, the reaction mixture contained 50 mM Tris-HCl (pH 8.1), 20 mM $MgCl_2$, 10 μM acetoacetyl-CoA, 40 mM KCl, and 8 ng of purified Acat1 enzyme (diluted into a buffer of 50 mM HEPES (pH 6.6), 9.5% glycerol, and 0.5 mg/ml gelatin). The reaction was initiated with the addition of CoA. The final volume of the reaction was 300 μL. We measured the change in absorbance at 303 nm over 30 min at room temperature (~ 25°C). For the determinations of K_M and V_{max} for CoA, the above procedure was followed, and the CoA concentration was varied between 15 and 300 μM. All absorbance measures were obtained using the BioTek Synergy 2 microplate spectrophotometer with Gen5 software. We analyzed the mean velocity over the first 5 min, whereas the reaction velocity curve remained linear over the first 10 min. K_M and V_{max} calculations were obtained using the Enzkin package in Matlab and SigmaPlot (version 12).

Sirt3 treatments

Sirt3 treatments were performed at 37°C for 30 min and contained 16 ng/μL Acat1, 0.375 μM Sirt3, 1 mM NAD^+, 0.5 mM DTT in dilution buffer (50 mM HEPES, pH 8.23, 0.5 mg/mL gelatin, 9% glycerol). Deacetylation was verified with immunoblots for Acat1 and acetyllysine, and activity was measured as above.

Electrostatic modeling

To assess the changes in charge complementarity for CoA in the active site of Acat1 caused by lysine acetylation, the tetrameric coordinates of Protein Data Bank code 2IBW [136] were prepared for electrostatic calculations. Water and other ligands were removed from the model. Lysine residues were either fully charged in the calculation or neutralized to simulate the effect of acetylation. An electrostatic map was calculated using the Adaptive Poisson-Boltzmann Solver [158] plugin in PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4, Schodinger, LLC) using default parameters. Local electrostatic environment was visualized as a continuous red-blue gradient running from -4 to +4 kT/e on the solvent accessible surface of the protein, or as an isosurface extending through space at +1 kT/e.

Chemicals and supplies

8-plex iTRAQ Reagents were purchased from AB Sciex. 6-plex Tandem Mass Tags (TMT) reagents were purchased from Thermo-Pierce. The bicinchoninic acid assay (BCA) Protein Assay Kit was purchased from Pierce Biotechnology (Rockford, IL). Trypsin Gold, was purchased from Promega (Madison, WI). Lysyl Endopeptidase (LysC) was purchased from Wako Chemicals (Richmond, VA). Sep-Pak tC18 and C18 cartridges were purchased from Waters (Milford, MA). A Poly SULFOETHYL A column (200 x 9.4 mm, 5 mm, 200 Å was purchased from PolyLC (Columbia, MD). C18 resin (5 μm pore size) was purchased from Alltech (Deerfield, IL). Inline MicroFilters and MicroTight column unions, fittings, and sleeves were purchased from Upchurch Scientific (Oak Harbor, WA). Fused-silica capillary tubing was purchased from Polymicro Technologies (Phoenix, AZ). LITHISIL lithium silicate was purchased from PQ Corporation (Valley Forge, PA). Formic acid and trifluoroacetic acid ampoules were purchased from Thermo Scientific (Rockford, IL). Pan-acetyl lysine antibody-agarose conjugate was purchased from Immunechem. TALON metal affinity resin was purchased from ClonTech. The deacetylase inhibitor entinostat/MS-275 was purchased from Fisher Scientific. Protease (Complete mini EDTA-free) and phosphatase (PhosSTOP) inhibitors were purchased from Roche (Mannheim, Germany). The goat anti-Acat1 antibody was purchased from Abcam (ab99190). The donkey anti-goat HRP antibody was purchased from Santa Cruz (sc-2020). The rabbit anti-acetyllysine antibody was pur-

chased from GeneTel (AC002). The goat anti-rabbit HRP antibody was purchased from Millipore (12-348). All other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Coenzyme A, acetoacetyl CoA, Nϵ-Acetyl-*L*-lysine, and the deacetylase inhibitors SAHA, trichostatin A, nicotinamide, and sodium butyrate were purchased from Sigma-Aldrich (St. Louis, MO).

Mitochondrial enrichment

Crude mitochondrial enrichment was performed using previously described methods [31], modified for acetylproteomics, with all steps carried out at 4°C. Frozen (−80°C) liver sections (~150 mg wet tissue weight) were placed into a Potter-Elvehjem glass/Teflon homogenizer along with 8 mL of MSHE Buffer (220 mM Mannitol, 70 mM sucrose, 5 mM HEPES, pH 7.41 mM EGTA, 1x protease inhibitor tablet, and 1x deacetylase inhibitors: 5 μM entinostat/MS-275, 2μM SAHA, 10 mM sodium butyrate, 10 mM nicotinamide, and 10 μM trichostatin A) supplemented with 0.5% BSA. The fasting/refeeding study also included 1x phosphatase inhibitors. The tissue was homogenized with 4 homogenizer strokes at 1000 rpm, and the resulting homogenate was decanted into a new tube. The homogenizer was rinsed with an additional 2 mL MSHE (supplemented with 0.5% BSA), which was added to the homogenate. The sample was centrifuged at 800 x g for 10 min in a bench-top conical centrifuge. The small amount of lipid that formed at the top of supernatant was carefully aspirated. The super-

natant, containing the mitochondria, was gently drawn off with a pipet-aid and transferred to an ultra-clear 12 mL centrifuge tube. The samples were centrifuged at 8000 x g for 10 min, and the resulting supernatant and any loose material was aspirated and discarded, leaving a dark brown pellet with a light brown halo around the center. An additional 1 mL of MSHE Buffer (supplemented with 0.5% BSA) was added to the pellet, which was disrupted by washing it from the side of the tube until homogenous using as few pipetting strokes as possible. Crude mitochondria were transferred to a 1.5 mL microfuge tube and centrifuged at 8,000 x g for 10 min in bench-top centrifuge. The supernatant was aspirated and discarded. The pellet was re-suspended in MSHE (without BSA) and centrifuged at 8,000 x g for 10 min in bench- top centrifuge. The supernatant was aspirated and discarded, and the mitochondrial pellet was flash frozen in liquid N_2 and stored at $-80°C$ until ready for use.

Sample preparation

Purified mitochondria were suspended in 8 M urea, 50 mM Tris pH 8.0, 5 mM $CaCl_2$, 100 mM NaCl, 1x protease inhibitor tablet, and 1x deacetylase inhibitors. Each samples was sonicated with a probe sonicator on ice to extract protein. Extracted protein was quantified by BCA assay. Protein from each sample, 1 mg and 0.5 mg per sample for obesity and fasting experiments respectively, was reduced and alkylated with iodoacetimide as described previously [31]. The samples were then digested with LysC

(100:1, protein:enzyme) for 2 hrs at ambient temperature. Following dilution to 1.5 M urea with 50 mM Tris pH 8.0, 100 mM NaCl, 5 mM $CaCl_2$ the samples were further digested with trypsin (50:1 protein:enzyme) overnight at ambient temperature. Digestions were quenched by TFA acidification, desalted with a tC18 sep-Pak (Waters), dried and re-suspended in 200 mM triethylammonium bicarbonate pH 8.5. Peptides from obesity experiments were labeled with 6-plex TMT and peptides from the fasting re-fed experiment were labeled with 8-plex iTRAQ per manufacturer's instructions. For each experiment labeled peptides were combined and desalted. Labeling efficiency was evaluated by analyzing a test mixture from each experiment by LC/MS/MS. Each experiment had $\geqslant$ 95% labeling efficiency, calculated by the number of N-terminal labeled peptides divided by the total number of peptide identifications.

Strong cation exchange fractionation and acetyl lysine enrichment

Labeled peptides were fractionated by strong cation exchange (SCX) on a polysulfoethyl A column (9.4 mm x 200 mm; PolyLC) with mobile phases A: 5 mM KH_2PO_4 pH 2.65 and 30% acetonitrile; B: 5 mM KH_2PO_4 pH 2.65, 350 mM KCl, and 30% acetonitrile; C: 5 mM KH_2PO_4 pH 6.5 and 500 mM KCl. The gradient was generated by a Surveyor LC quaternary pump (Thermo) at 3 mL/min flow rate. Collected fractions were desalted and a

small portion was retained for protein analysis (fasted-re-fed experiment) resuspended in 50 mM HEPES pH 7.5 and 100 mM KCl. Approximately 100 μL pan-acetyl lysine antibody-agarose conjugate (Immunechem) was added to each fraction. The samples were rotated overnight at 4°C. Unbound peptides were collected and saved for further fractionation (obesity experiments). The conjugated beads were rinsed eight times with 50 mM HEPES pH 7.5 and 100 mM KCl, followed by elution with 0.1% TFA. Eluted peptides were desalted prior to analysis.

High pH reverse phase fractionation (obesity experiments)

Peptides not bound to anti-acetyl lysine beads were further fractionated across a Gemini C18 reversed phase column (4.6 mm x 250 mm; Phenomenex) with mobile phases A: 20 mM ammonium formate pH 10; B: 20 mM ammonium formate pH 10 and 80% acetonitrile. The gradient was generated by a Surveyor LC quaternary pump (Thermo) at 0.8 mL/min flow rate. Collected fractions were dried and saved for protein analysis.

LC/MS/MS

All samples were analyzed by reverse phase liquid chromatography on a nanoAcuity (Waters) coupled to a Velos Orbitrap (Thermo). Samples were loaded onto a 75 μm inner diameter column packed with 5 μm C18 particles (Michrome) and eluted with an acetonitrile gradient in 0.2%

formic acid. Mass spectrometry instrument methods started with one MS1 survey scan (resolution = 30,000; 300 Th - 1,500 Th) followed by data dependent MS2 fragmentation and analysis (resolution = 15,000) of the ten most intense precursors. Precursors with charge state $\geqslant 2$ were isolated in a $\pm$ 1.5 Th window and fragmented using the previously described QuantMode instrument acquisition method [108]. Precursors were excluded from re-sampling for 30 seconds.

Database search and FDR filtering

Spectra were converted to searchable text files using DTA generator. Generated text files were searched as fully tryptic peptides, with mass tolerances of 20 ppm for precursors and 0.01 Th for fragment ions, against a UniProt target-decoy mouse canonical and isoforms database (downloaded Nov. 28, 2011) using OMSSA v2.2.8 [159] (54,314 target and 54,314 decoy protein entries). Carbamidomethylation of cysteine, isobaric labeling of lysine and the peptide N-terminus were searched as fixed modifications for all samples. Methionine oxidation and isobaric labeling of tyrosine were searched as variable modifications for all samples. Enriched fractions were additionally search for variable acetylation. The acetylation mass shift was set to the difference between an acetyl group and an isobaric label, -187.1523 Da and -262.1948 for TMT and iTRAQ respectively [32]. With the COMPASS v1.2.1.0 software suite search results were filtered to 1% FDR at the unique peptide level [153]. TMT quantitation on iden-

tified peptides was performed within COMPASS v1.2.1.0 as previously described [160]. Peptides were grouped into proteins according to previously reported rules and protein identifications were further filtered to 1% FDR [161]. Protein quantification was performed by summing all reporter ion intensities within each channel for each protein. Acetylated peptides and peptides in multiple protein groups were excluded from this step.

Acetylation analysis

All possible patterns of acetylation, on lysine or the protein N-terminus (considering variable cleavage of the initiator methionine), for each identified peptide were considered according to previously described rules [32]. If the calculated localization confidence is 95% or greater, then the acetylation pattern is considered localized. All peptides with identical acetylation patterns were grouped together and their reporter ion intensities were summed. Acetylation sites that localize to the C-terminus of a peptide are excluded from this calculation.

Protein normalization

All reporter ion intensities were $\log_2$ transformed and mean normalized for every acetyl isoform and protein. To normalize acetylation for protein differences, the value for each acetyl isoform reporter ion channel had subtracted from it the quantitative value of that channel from the

corresponding protein. This gives a protein normalized acetylation mean normalized value which is further interrogated for statistical significant fold change between conditions across experiments. Fold change calculations were made by averaging the protein normalized values for each condition and then calculating the difference of averages. For each of these comparisons a p-value was calculated (Welch's t-test with Storey correction for multiple hypotheses testing) [162].

Measurement of oxygen consumption

All respiration assays were performed using an XF-96 Extracellular Flux Analyzer (Seahorse Bioscience) with a 4-port injection system. The experimental design was based on previous analysis of isolated mitochondria with some modifications [116]. Mitochondria were isolated as described above using MSHE buffer with the addition of 1%BSA. Mitochondria were quantified by wet weight using an analytical scale. Isolated mitochondria were resuspended to 100 μg/μL in MAS assay buffer (70 mM sucrose, 220 mM mannitol, 10 mM KH_2PO_4, 5 mM $MgCl_2$, 2 mM HEPES, 1 mM EGTA and 0.2% (w/v) BST, pH 7.2 at 37°C). Mitochondria were diluted to 2 μg/μL in MAS buffer supplemented with 10 mM glutamate and 2 mM malate. Mitochondria from lean and obese B6 mice were added to at least 10 wells from each condition to a final concentration of 40 μg/well (20 μL of diluted mitochondria) and were spun down using a benchtop refrigerated centrifuge for 20 min at 2000 g and 4°C to adhere the isolated

mitochondria to the bottom of the wells. After the spin MAS buffer with glutamate and malate was added to each well to bring the total volume to 120 μL. Injection stocks of ADP (20 mM), Oligomycin (100 μg/mL), FCCP (100 μM) and Antimycin A (250 μM) (all from Sigma) were diluted in water at 10x the final concentration and 20 μL were loaded into the injection ports of an XF96 plate. After calibration of the assay plate using the XF96 software, mitochondria from lean and obese B6 mice were loaded into the instrument and subjected to a standard mitochondrial stress test protocol. Respiration rates were measured for 3 minutes to establish a baseline, then measured again after the addition of ADP, Oligomycin, FCCP and Antimycin A to record State 3, State 4o, State 3u and non-mitochondrial respiration rates, respectively. All data analysis was performed using the XF96 software (version 1.4.1.4), and the "Level(Direct)Akos" algorithm that is a built-in factor of the software package.

Protein purification

Production of recombinant mouse Acat1 (MmCD00319142) was based on previous work [136, 154, 155]. BL21 (Rosetta2(DE3)pLysS) cells were transformed with pET28 encoding residues 31-424 followed by a c-terminal histidine tag, AAALEHHHHHH. An overnight culture was diluted 1:50 into TB and autoinduced at RT using a glucose/lactose mix for 24 hours. The final media contained 12 g/L tryptone, 24 g/L yeast extract, 12.8 g/L glycerol, 8.9 mM potassium phosphate mix, 0.375% asparate, 2 mM $MgSO_4$,

0.15 g/L glucose, 5 g/L alpha-lactose monohydrate, and antibiotics. All purification steps were performed at 4°C. BME and protease inhibitors were added the same day. Cells were pelleted at 5000 x g for 10 min, re-suspended in 25 mM HEPES pH 6.6, 30 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 10 mM imidazole, 5 mM BME, protease inhibitors, and 2 mg/mL lysozyme, and incubated for one hour before lysis using a French press or sonication. Insoluble material was pelleted at 15,000 x g for 30 min, and the solute was incubated with TALON Metal Affinity Resin for two hours. The unbound was removed and the beads were washed thoroughly with lysis buffer, then wash buffer (25 mM HEPES pH 6.6, 150 mM NaCl, 10% glycerol, 25 mM imidazole, and 1 mM BME), and the protein was eluted in 25 mM HEPES pH 6.6, 150 mM NaCl, 10% glycerol, and 500 mM imidazole. Protein was concentrated using Amicon Ultra centrifugal filters with a 30 kDa cutoff, after which DTT was added to 0.1 mM. Proteins were aliquoted, flash frozen using liquid nitrogen, and stored at -80°C. Mutants were generated using site-directed mutagenesis and were purified on the same day as wild-type protein.

Site-specific acetyllysine incorporation was performed as previously described [137, 138, 155]. An ampicillin-resistant version of the above plasmid was cotransformed with pBK AcKRS (acetyllysine tRNA synthase) and pCDF PyIT ($tRNA_{CUA}$) into BL21(DE3) cells. An overnight culture was diluted 1:50 into TB and grown to OD_{600} = 0.5 to 0.7. Acetyllysine was added to 10 mM and nicotinamide was added to 20 mM while the culture

was at RT for 30 min before the addition of IPTG to 0.3 mM. The culture was induced overnight at RT. Purification was performed as above, except with the inclusion of 20 mM nicotinamide in the lysis buffer. Incorporation was verified by mass spectrometry and Immunoblot analysis.

Recombinant Sirt3 was purified as previously described [156]. Briefly, Sirt3 was expressed on pQE-80 in BL21-DE3 cells and induced via IPTG. Cells were lysed via sonication in 25 mM Tris pH 8, 150 mM NaCl, 1 mM BME, 5 mM imidazole, 10 μg/mL leupeptin, 0.1 mM PMSF, and 5 μg/ml aprotinin. Insoluble material was pelleted at 15,000 x g for 30 minutes, then the soluble lysate was filter clarified and loaded onto an Ni chelating column on FPLC. The column was washed with lysis buffer without protease inhibitors, then eluted with a linear gradient with a step at 15% 300 mM imidazole for three column volumes. Fractions containing Sirt3 were dialyzed overnight with 50 mM Tris pH 7.6, 1 mM DTT, 10% glycerol.

3 CHAPTER THREE

Annotation of orphan mitochondrial proteins via interaction mapping

Brendan J. Floyd, Catie E. Minogue, Emily M. Wilkerson, Mike T. Veling, Katarzyna A. Gromek, Emily T. Beebe, Holly Cho, Brendan K. Dolan, Sarah L. Bohl, Kelly M. Werner, Adam Jochem, Russell L. Wrobel, Michael S. Westphall, Jung-Ja Kim, Joshua J. Coon, and David J. Pagliarini

3.1 Abstract

Mitochondrial function is essential for nearly all eukaryotic cells, and dysfunction is associated with hundreds of diseases. Over the past decade, the known mitochondrial proteome has been expanded to include hundreds of newly identified proteins. Despite the importance of mitochondria for health and disease, many mitochondrial proteins have no known function. This lack of knowledge serves as a bottleneck, impeding the progress of biomedical research. As such, a systematic approach to annotating these uncharacterized proteins is needed. Here, we assemble the current state of mitochondrial protein-centric knowledge through the integration of diverse datasets and the hand-annotation of each of the 1,143 proteins. We find that 209 mitochondrial proteins have no known function, which we term orphan mitochondrial proteins. To begin the systematic annotation of these proteins, we build a cell type- and condition-specific protein-protein interaction (PPI) network. Through the application of state-of-the-art methods of affinity enrichment-mass spectrometry (AE-MS) quantification and analysis, we identify high-confidence interactions involving 80 OMPs. We associate more than a dozen OMPs with known pathways, a key first step toward characterization. Our efforts form the first direct evidence of a human coenzyme Q biosynthetic network *in vivo* and define its dynamic nature as well as its molecular architecture. We also prioritize LYRM5 as a key novel member of the electron transfer flavoprotein (ETF) complex,

and we find that the interaction is direct. Our work forms the basis for the annotation of 21 OMPs, and will serve as an important resource for the field of mitochondrial biology.

3.2 Introduction

Mitochondria are key centers of metabolism for nearly all eukaryotic cells [3]. Once considered to be mere sites of ATP generation, it is now known that mitochondria participate in a wide range of essential functions related to cellular metabolism, signaling, and programmed cell death. Consistently, large-scale proteomics- and computation-based efforts during the past decade have revealed that the mitochondrial proteome is much more complicated than once thought, and dysfunction of these organelles is now associated with hundreds of inborn errors of metabolism and common diseases [3, 4].

Despite our advanced understanding of the mitochondrial proteome - estimated to be defined to approximately 90% completion - annotation of the biochemical functions of these proteins is far from complete. This gap in knowledge has thwarted our understanding of basic mitochondrial biology and has obscured the nature and cause of many mitochondrial diseases. For instance, many patients with biochemically-established mitochondrial disease do not harbor mutations in the suspected genes, implying the existence of unidentified proteins necessary for proper function of the affected process [18, 23]. Alternatively, many other diseases arise from mutations in proteins with no known function, making it difficult or impossible to understand the metabolic mechanism of the disease. As such, we sought to establish a robust approach for systematically annotating the functions

of uncharacterized mitochondrial proteins.

Previous large-scale efforts have proven effective in assigning functions to uncharacterized proteins. Efforts to define the human protein-protein interaction network, termed the interactome, have been particularly effective at functional annotation of uncharacterized proteins [48]. Even so, recent analysis of the known human interactome has revealed it to be woefully incomplete [57] and systematically biased against unstudied proteins [55]. This sparsity prohibits further systematic and computational annotation of uncharacterized proteins, as they exist as "islands," lacking connections to the greater mass of annotated proteins. However, recent advancements in the study of PPIs by mass spectrometry have assisted the characterization of several previously unstudied proteins [58, 65, 69]. In this approach, termed affinity enrichment-mass spectrometry (AE-MS [70]), a protein of interest is enriched from a sample and co-enriching proteins are identified by mass spectrometry (MS). As the majority of co-enriching proteins are not meaningful interactors of the bait protein, this approach requires the analysis of a large number of bait proteins, quantitative MS, and scoring algorithms to differentiate between meaningful interactions and the non-specific binding of the majority of co-enriching proteins.

In this study, we begin by curating a list of the proteins from the known mitochondrial proteome that still lack significant functional annotation, which we call orphan mitochondrial proteins (OMPs). We estimate there to be 209 OMPs in humans, including more than 32 that have already

been associated with human diseases. We then leverage cutting edge strategies for AE-MS to define an extensive cell type- and condition-specific interactome of 80 OMPs and disease-relevant mitochondrial proteins. This focused approach on high-yield bait proteins allows us to assess the role of cells and mitochondria-relevant conditions in modulating PPIs, thus affording a greater depth of characterization. From more than 130,000 observed interactions, we identify 978 new, high-confidence PPIs (0.8%), enabling us to define an initial annotation for a broad swath of OMPs and leading us to nominate several proteins as likely candidates in disease. Our findings include the identification of a human coenzyme Q biosynthetic complex that we validate using *in vitro* cell-free protein expression and purification and that is dynamic in response to nutrient state. Notably, our work also establishes that the OMP LYRM5, which we rename electron transferring flavoprotein associated factor (ETFAF), directly interacts with the ETF complex and likely facilitates its activity.

In summary, our work here has catalyzed the initial association of 21 OMPs with known mitochondrial pathways and processes, has defined the molecular architecture of several complexes using *in vitro* protein binding assays, and has nominated several proteins as likely candidates in mitochondrial disease. Our data will serve as a powerful resource for establishing the function and disease-association of orphan mitochondrial proteins.

3.3 Results

Twenty percent of the mitochondrial proteome lacks annotation

Previous analysis suggested that much of the mitochondrial proteome is uncharacterized [13, 44]. We, therefore, sought to capture an up-to-date assessment of mitochondrial proteins and their functions. First, we curated a high-confidence list of human mitochondrial proteins, integrating the MitoCarta compendium [13] with recent mass spectrometry-based analysis of the mitochondrial matrix and inter-membrane space [14, 15]. This list includes 1,143 human proteins with validated mitochondrial localization. Next, we annotated these proteins based on the integration of online databases in conjunction with careful hand-annotation of the literature. Our results suggest that at least 209 mitochondrial proteins have no known function, and a further 21 proteins with dual localization to the mitochondrion and another cellular compartment do not have clear roles within the mitochondrion. These orphan mitochondrial proteins (OMPs) constitute approximately 20% of the mitochondrial proteome. Despite their lack of characterization, 34% of OMPs are conserved in yeast, indicating their key functions in the cell.

A further assessment of the importance of these OMPs is their prior association with disease. Many mitochondrial proteins are associated with disease, often with no clear pathological mechanism [4]. To connect

mitochondrial proteins with their associated diseases, we mapped large-scale disease lists onto our list of mitochondrial proteins [17, 45, 46]. We found that hundreds of mitochondrial proteins, including at least 30 OMPs, are associated with more than 300 human diseases. Surprisingly, despite no mechanistic knowledge of their function, we observe that mutations in at least 20 OMPs are associated with disease phenotypes in OMIM.

Based on their conservation in lower organisms and role in disease-states, we therefore predict that OMPs play a key role in the cell, and therefore are encouraged to accelerate the rate of their characterization. We selected targets from among the 209 OMPs and disease-relevant proteins with unclear pathological roles. Among these proteins, 28% are associated with disease, and 34% have homologs in *S. cerevisiae*. We prioritized an initial set of 80 target baits with an emphasis on disease-relevance, conservation, and lack of transmembrane domains. To associate these OMPs with known pathways and processes, and because previous efforts have shown that PPI network analysis can efficiently connect uncharacterized proteins to known pathways, we elected to establish OMP-specific interactions via affinity enrichment-mass spectrometry (AE-MS).

Overall experimental strategy

Each of the selected OMPs was cloned into a mammalian expression vector with a C-terminal FLAG tag (Fig. 3.1A). To re-establish mitochondrial localization of each bait, and as overexpression and tags may prevent

proper protein folding and localization, we validated the localization of each construct in HEK-293 cells. Consistent with our use of a short, C-terminal tag, more than 90% of our OMP fusion constructs localized to the mitochondrion (Fig. 3.1B). Only these properly localized mitochondrial proteins were used for subsequent analyses.

Many PPIs are context- or cell type-specific, potentially mediated by differences in protein abundance, localization, or post-translational modifications. To increase the breadth of the potential interaction space, we employed two cell lines and two nutrient conditions for our interaction analyses. We chose HEK293 cells, a workhorse of cellular biology, and HepG2 hepatoma cells as liver cells have the greatest diversity of mitochondrial proteins [13]. For each cell line we optimized a transient transfection protocol for maximal expression of the OMPs.

To examine the effect of the modulation of mitochondrial activity on interactions, we employed a well-characterized nutrient-switching model, replacing glucose for galactose in the media of half of the replicates. Galactose is known to up-regulate oxygen consumption and increase dependence on mitochondrial function [163], and we observe this in our chosen cell lines (Fig. 3.1C). Previous work in our lab has defined clear proteomic, phosphoproteomic, and lipidomic differences between cells grown in glucose and galactose, indicating that we are interrogating diverse cellular states.

Finally, because of the stochastic natures of biology and MS, we col-

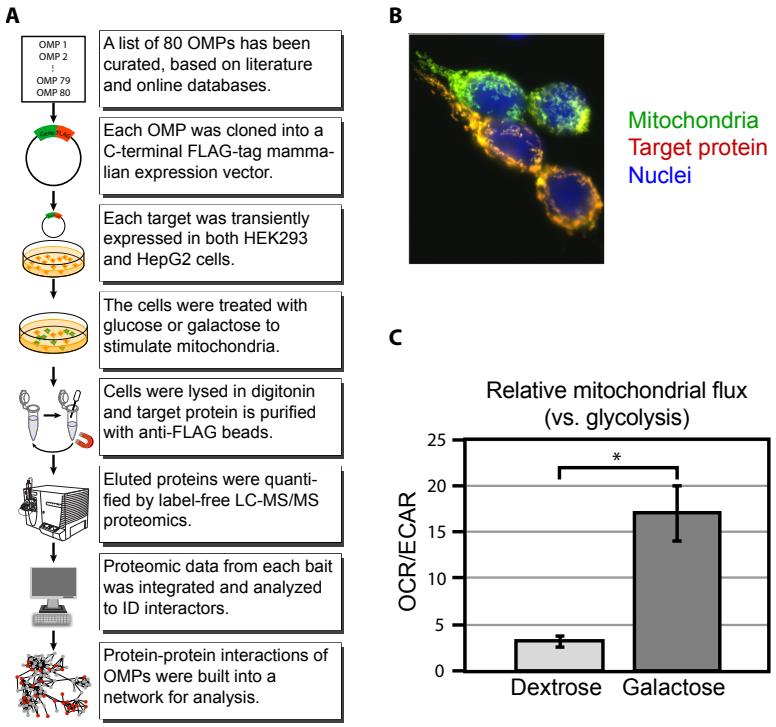


Figure 3.1: *AE-MS methodology.* A) Schematic workflow of the AE-MS method. OMP genes were selected, were cloned into a mammalian expression vector, and were validated for mitochondrial localization of the FLAG-fusion protein. Plasmids were transfected into six plates for each cell line, treated with galactose or glucose, and collected for anti-FLAG immunoprecipitation followed by LC-MS/MS analysis. Proteomic data from each bait enrichment was integrated and scored, with the top scoring interactions chosen for network mapping and follow-up analysis. B) Localization of FLAG-tagged constructs to the mitochondrion based on MLS-GFP and anti-FLAG fluorescence microscopy. C) Galactose induces mitochondrial respiration. Ratio of rates of oxygen consumption (OCR) and extracellular acidification (ECAR) in HepG2 cells grown in 10 mM dextrose or galactose for 24 hours prior to assay. OCR/ECAR is proportional to mitochondrial vs. glycolytic flux.

lected three replicates of each condition for a total of 12 plates per bait protein. Replicate analysis such as this is not standard in the field and this depth of analysis has never been attempted, to our knowledge. Our approach allows us to assess the differential nature of interactions, while affording us greater confidence in our results.

Previous attempts at enriching for membrane-associated proteins have shown that the nature of the approach, especially the buffer and detergent conditions, are key factors [164, 165]. To enhance detection of membrane-associated proteins, we employed a lysis buffer featuring glycerol and digitonin, modified from one previously shown to stabilize yeast mitochondrial protein interactions [166]. Each bait protein was isolated by anti-FLAG immunoaffinity enrichment followed by competitive elution. The bait and co-purifying proteins were then subjected to label-free quantitative proteomics by LC-MS/MS.

Interactome analysis identifies hundreds of new interactions

To help identify informative interactions above the background, we utilized CompPASS, a known and validated algorithmic approach to highlighting high-confidence interactions and removing nonspecific preys [60]. As not much is known about the interactions of our OMPs, we supplemented our list of target baits with mitochondrial proteins of known function.

Additionally, we included a variant of green fluorescent protein harboring an N-terminal mitochondrial localization sequence (MLS-GFP-FLAG) as a negative control.

To assess the performance of our approach and to determine an appropriate cutoff score, we began by focusing on our positive-control bait proteins with known binding partners. Based on the literature, we curated a list of high-confidence (HC) PPIs involving our positive-control baits, and compared their CompPASS scores to all observed preys or just mitochondrial preys. We observe that 96% of all observed CompPASS scores in our data are below 10, independent of the known subcellular localization of the prey (Fig. 3.2). In contrast, 40% of our HC-PPIs were scored at greater than 10. As such, we chose a cutoff of 10 for subsequent analyses.

Applying this threshold to the rest of our data, we identify 4,805 interactions out of an initial observed set of 138,863 ($\leqslant 4\%$). As we prioritize interactions between mitochondrial proteins, we limited the list to the 738 observed mitochondrial preys and identified 978 interactions involving 370 mitochondrial proteins (Fig. 3.3). As each bait was tested in two cell lines, under two conditions, and in triplicate, we analyzed nearly 1000 independent AE-MS experiments. This approach and scale afforded us great clarity in the elucidation of novel high-confidence interactions.

Given our success in identifying known PPIs, we next sought to explore pathway-specific interactions of OMPs.

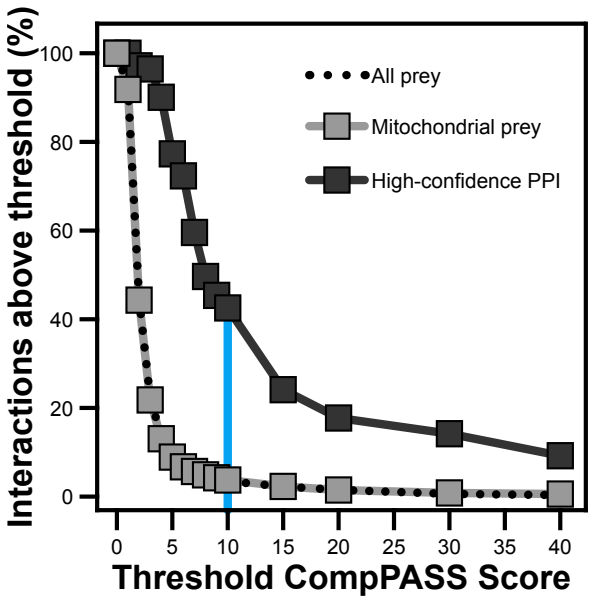


Figure 3.2: *CompPASS scoring accuracy.* CompPASS scores were calculated for all bait-prey interactions, including 138,863 total (black dotted), 75,252 involving only mitochondrial prey (gray), and 142 determined *a priori* to be high-confidence PPIs based on the literature (black). At each score threshold, the percent of remaining PPI per bin was calculated. The vertical blue line indicates the score threshold used in this study.

New insights into CoQ biosynthesis

Coenzyme Q (CoQ) is a key redox-active lipid-constituent of the electron transport chain. CoQ is synthesized within the mitochondrion in a process that involves at least 13 human proteins (PDSS1-2, COQ2-7, COQ9, COQ10A-B, and ADCK3-4), six of which are OMPs. Much of the CoQ biosynthetic process is conserved from humans to yeast and bacteria. Over many years and across a series of isolated studies, eight of the yeast proteins, Coq3p-9p and Coq11p, have been found to physically interact

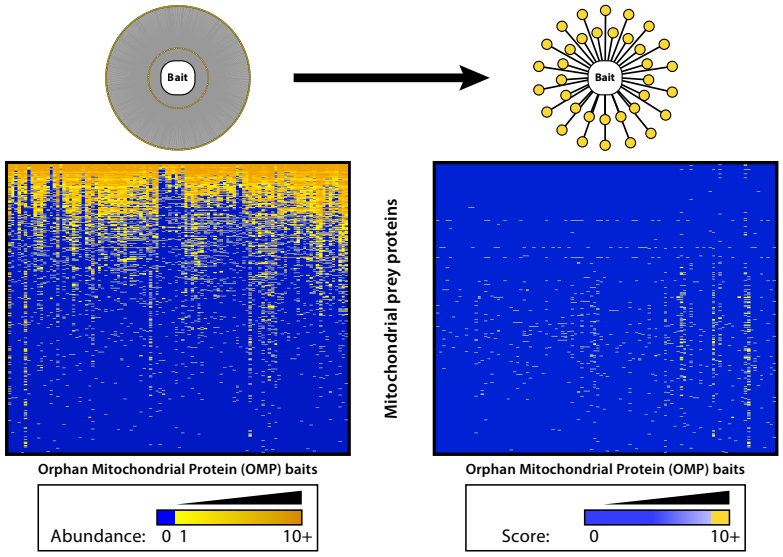


Figure 3.3: *Quantitative score enriches for high-confidence interactors* Top: Schematic of the results of CompPASS score filtering. Bottom left: Abundance (peptide spectral matches) of all mitochondrial prey proteins is indicated in a heat map where blue indicates a prey was not observed, and shades of yellow indicate greater abundance. Prey proteins are in rows, and FLAG-tagged baits are in columns. Data are averages from 6 replicates per cell line (3 dextrose, 3 galactose). Bottom right: Heat map showing scores above CompPASS threshold of 10. Preys are organized in the same order as in left heat map. Blue indicates a score that did not rise above the threshold.

and potentially form a biosynthetic complex [166, 167]. Our approach is, perhaps, more comprehensive and rigorous than this work.

No physical complex has yet been found for the human or bacterial proteins. To date, there are only 4 known human CoQ-related PPIs, one of which was identified recently by our group (Fig. 3.4B) [168–170]. Additionally, there are several key distinguishing features between the yeast and mammalian CoQ biosynthetic pathways. Most obviously, the hu-

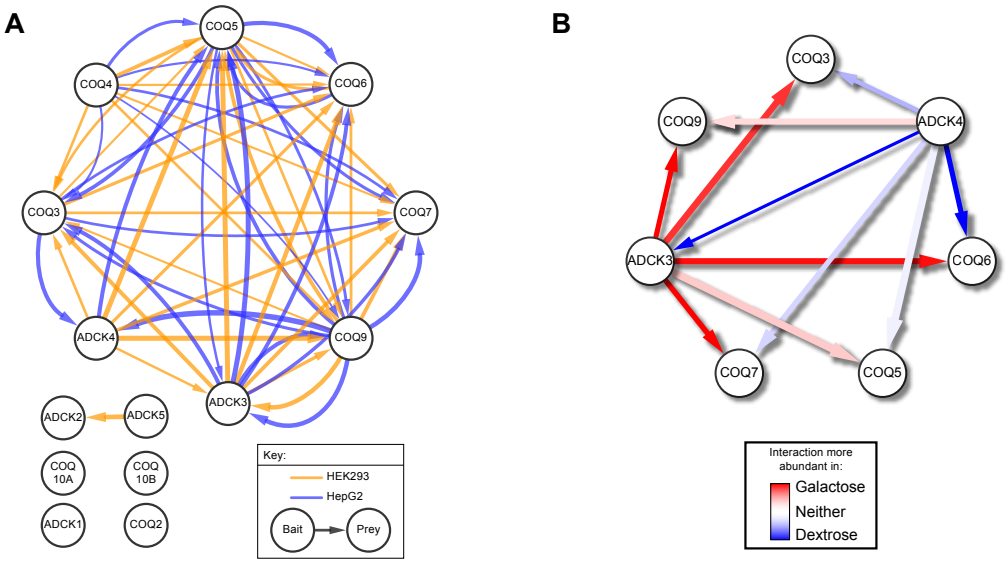


Figure 3.4: *AE-MS analysis reveals the human CoQ-related interactome and its dynamic nature.* A) All CoQ-related proteins used as baits and/or observed as preys in this study are shown as white nodes. Interactions rising above the score threshold are shown as arrows, colored gold for interactions from HEK293 cells, and blue from HepG2 cells. B) CoQ-related interactions involving ADCK3 or ADCK4 as bait were assessed for the effect of cellular metabolic status. Interactions more abundant in galactose or dextrose media are shown in red or blue, respectively, while those with no clear media effect are in white.

man system has been expanded to include paralogs of several proteins (PDSS1-2, ADCK3-4, COQ10A-B). Additionally, yeast can undergo certain chemical transformations in the CoQ biosynthetic process that are not possible in humans. Given the difficulty in observing PPIs involving the human COQ proteins, as well as the differences in the number and diversity of proteins and chemical transformations that can occur in the yeast and human systems, it has so far been unclear whether a physically-

interacting mammalian complex exists as a counterpart of the yeast system. The revelation of such a complex could facilitate the characterization of CoQ-related OMPs.

As described above, we optimized our AE-MS system to facilitate the identification of membrane-associated proteins. We did so with the elusive CoQ complex firmly in mind, and used COQ9 as one of our test baits. With this optimized system, we have established the first direct evidence of a mammalian CoQ biosynthetic complex (Fig. 3.4A). As in yeast, we find that the human complex is highly interconnected. Consistent with the essentiality of the CoQ complex, we observe 34 and 39 of these interactions in HEK293 and HepG2 cells, respectively.

A key question in the study of CoQ production is how, and even if, its rate of synthesis is regulated. Previous work in our lab has shown that CoQ abundance is elevated in respiring *S. cerevisiae*. As switching media from glucose to galactose stimulates mitochondrial respiration in our model (Fig. 3.4B), we expected to observe elevated CoQ abundance in the galactose condition. Indeed, previous work from our lab has shown that treatment of HEK293 cells with glucose for 4-24 hours leads to an increase in CoQ abundance. These data suggest that there may be a conserved mechanism by which CoQ biosynthesis is regulated.

We hypothesized that CoQ-related PPIs might differ across the transition from glucose to galactose and thereby facilitate increased CoQ production. Consistent with this hypothesis, the abundances of CoQ PPIs

involving ADCK3 are increased in galactose-treated cells, and those with ADCK4 are reciprocally decreased. ADCK3 and ADCK4 are the mammalian paralogs of the yeast Coq8p. While both of these proteins are OMPs, ADCK3 has an atypical protein kinase-like fold and both are essential for CoQ synthesis [169, 171]. These data suggest an important connection between CoQ production and interactions among CoQ biosynthetic complex members, and may indicate reciprocal regulation of the activities of the paralogs ADCK3 and ADCK4. These data also demonstrate the power of our method for identifying low-abundance, dynamically-regulated interactions.

Having established the existence and context-specific abundance of the mammalian CoQ biosynthetic complex, a key remaining question is how the complex is structured, that is, which PPIs are direct, and which are indirectly mediated by intervening proteins. While yeast CoQ-related PPIs have been observed for some time, the molecular architecture of this complex has not yet been elucidated in any organism. Illumination of the nature of interactions within the complex will reveal how proteins of known function transfer the hydrophobic CoQ intermediates between each subunit, and will suggest roles for the aforementioned CoQ-related OMPs.

As the molecular architecture of the CoQ biosynthetic complex has not yet been elucidated in any organism, we sought to apply cell-free protein expression and purification systems to ascertain the direct interactions

and subcomplexes among these proteins. Our lab recently generated the first purified mammalian CoQ protein subunits [168, 171]. We used these proteins to identify a direct interaction between COQ7 and COQ9, and showed that it is modulated by functionally relevant mutations on each protein [168]. We have now expanded this approach to the rest of the core complex subunits (Fig. 3.5).

Here, we show that almost all pairs among COQ3-7 and COQ9 are sufficient for direct interaction *in vitro*, with the exceptions of COQ9:COQ3 and COQ9:COQ4. This is consistent with *in vivo* data from yeast and mouse knockout models in which the loss of a single subunit of the complex significantly disturbs and destabilizes the rest of the proteins [168, 172].

After the successful identification of these binary interactions, we then attempted to rebuild the entire core CoQ complex *in vitro*. Similar to our binary PPI data, we observe that 40 of the 42 potential complexes involving up to six COQ proteins are stable *in vitro*. Interestingly, we observe that complexes involving COQ9 but lacking its direct binding partner COQ7 are weaker than other similar complexes, re-affirming the importance of this functional interaction. Notably, we establish the first *in vitro* CoQ complex containing all 6 of the known core members of the complex (Fig. 3.5). While it is likely that other CoQ-related proteins interact with the core subunits, such as ADCK3 and ADCK4 as in our AE-MS data, this work suggests that these core CoQ-related proteins are sufficient to structurally support complex stability. Our results reveal the nature of the complex at

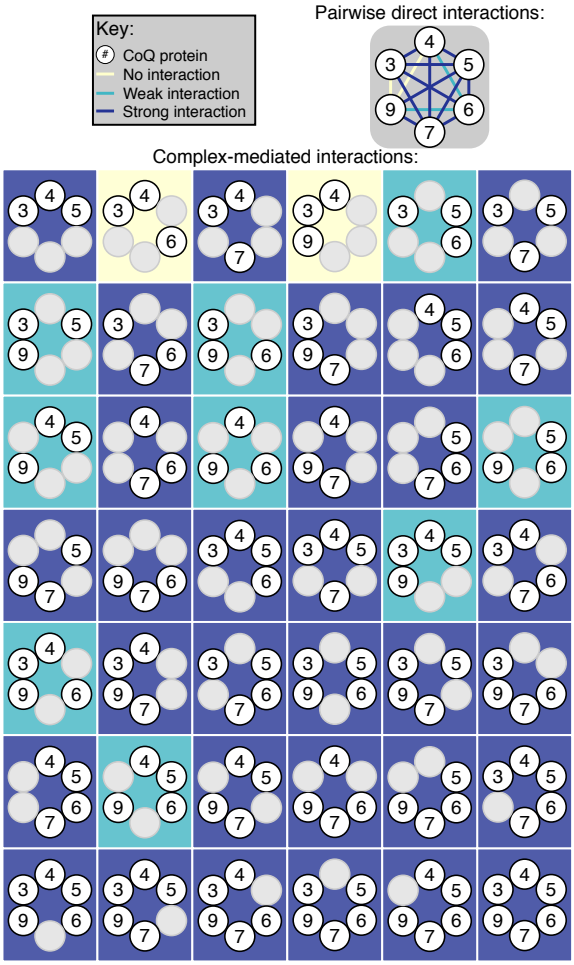


Figure 3.5: *CoQ-related proteins interact directly in vitro.* Schematic of CoQ network direct interactions established *in vitro*. Top: Each pair of core CoQ-related human proteins, one with a Strep(II)-tag and the other untagged, was independently transcribed and then co-translated. Translations were purified with StrepTactin resin, resolved by denaturing SDS-PAGE, and visualized with tryptophan fluorescence and subsequent Coomassie staining. Strong, weak, and nonexistent interactions were scored, and are represented as edges connecting COQ protein nodes. Bottom: Results of testing all 42 possible core CoQ network complex interactions. Interactions involving more than two proteins were tested with the same system. Proteins interrogated per square are shown in white nodes with black numbers, while proteins not included in that experiment are shown in gray. Node position is as above. Complex-mediated interactions that were observed to be strong, weak, or nonexistent are colored blue, green, and yellow, respectively.

molecular resolution and offer a platform for the future interrogation of the roles of each subunit, including the OMPs COQ4 and COQ9. In particular, with this core complex in hand, we can now attempt *in vitro* biochemical characterization of much of the biosynthetic pathway, addressing the roles of OMPs in this more tractable system.

Associating orphan proteins with established pathways

While the above helped to clarify how orphan, CoQ-related proteins fit into the CoQ pathway and complex, other OMPs have no prior association with any pathway or complex. Our work makes initial connections for many of these OMPs.

C2orf47

The OMP C2orf47 interacts quite strongly with AFG3L2 and SPG7, the two members of the human m-AAA protease complex (Fig. 3.6, right side). These proteins are known to form hetero-hexamers as well as homo-hexamers of AFG3L2 alone. The m-AAA protease is a trans-membrane complex responsible for the maturation of several membrane-associated proteins, and also plays an important role in mitochondrial protein quality control under stress conditions.

C2orf47 also interacts quite strongly with both prohibitins, PHB and PHB2, albeit just below our threshold. A previous AE-MS study found that C2orf47 is a prey of Phb2 in mouse embryonic fibroblasts [65]. Notably,

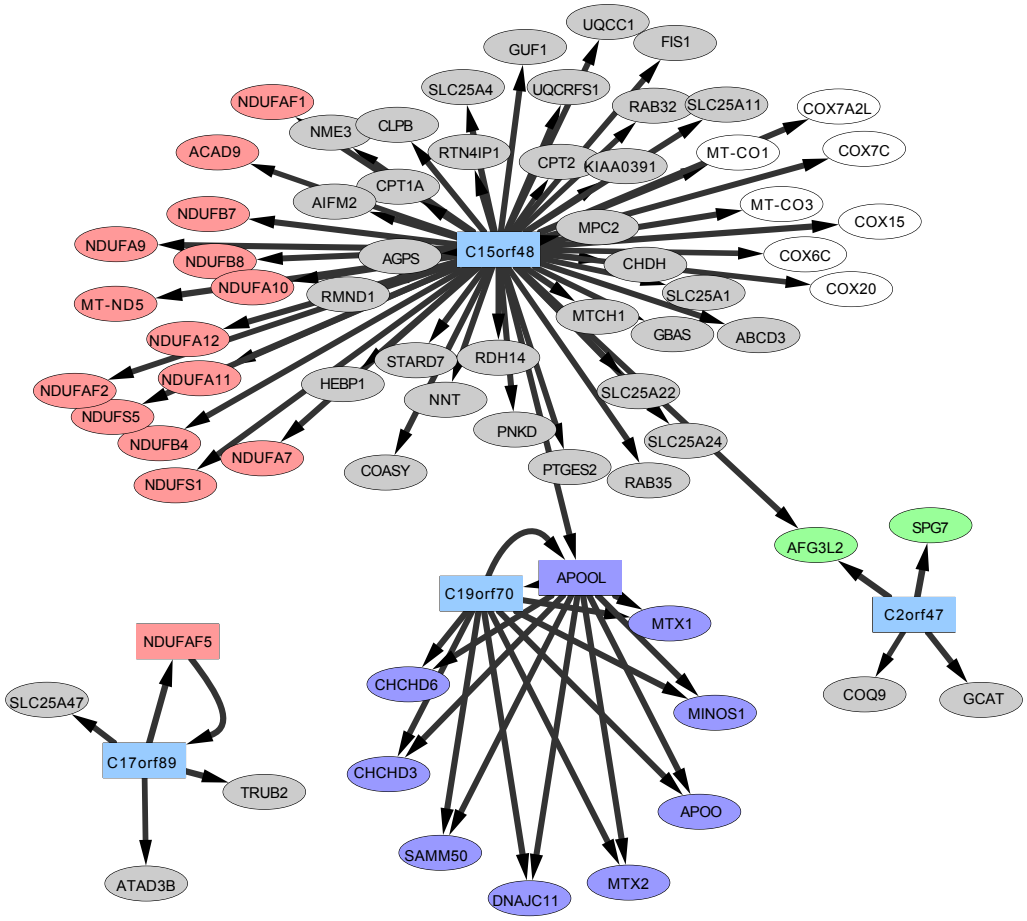


Figure 3.6: *OMP-related interaction network.* Mitochondrial PPIs involving selected OMPs are shown. Node color corresponds to pathway: CI in red, CIV in white, MICOS in purple, m-AAA protease in green, OMPs in blue, and all other mitochondrial preys in gray. Edge width is proportional to the score.

Phb2 also enriched for all members of the m-AAA and i-AAA protease in that study. In yeast, prohibitins are known to interact with and regulate the m-AAA protease [173]. Mouse knockout of Phb2 leads to dysfunctional mitochondrial membrane architecture [174] and both beta cell and neuronal dysfunction [175]. According to BioGPS, mouse C2orf47 has a strong mRNA correlation with Phb2, with many mitochondrial ribosomal proteins, and Dnajc19, a known interactor of the prohibitins and a modulator of cardiolipin metabolism.

There is little known about C2orf47. In the only publication that has examined the effect of modulation of the expression of C2orf47, it was found that siRNA-based knock down led to increased proliferation and prevention of apoptosis in SK-hep1 hepatic adenocarcinoma cells under serum- and glucose-starved conditions [176]. Overexpression of HA-tagged C2orf47 reduced proliferation rates of HEK293T cells [176]. Thus, C2orf47 may play a pro-apoptotic role. Cellular proliferation depends on prohibitins in the mitochondrion [177], and OPA1, a substrate of the m-AAA protease and regulator of mitochondrial structural dynamics, also has been linked to apoptosis [178].

C2orf47 is predicted by Phyre2 to have a cystatin-like fold with a Tim44-like domain. There are two yeast (Tim44p, Mba1p) and three human (TIMM44, MRPL45, C2orf47) Tim44-like domain containing proteins. Yeast with a multi-copy variant in the MBA1 gene can compensate for deletion of AFG3, the homolog of human AFG3L2. Mba1p is a receptor of

the mitochondrial ribosome along with Mdm38p. The human homologs of Mdm38p are leucine zipper-EF-hand containing transmembrane protein 1 and 2, LETM1 and LETM2. LETM1 is mitochondrial, associated with the inner membrane, and interacts with C2orf47 in our study, albeit at low abundance. Thus, there is a physical and functional relationship between Tim44-domain containing proteins (including C2orf47), the m-AAA protease, the mitochondrial ribosome, and calcium homeostasis that is conserved from yeast to humans.

C15orf48

C15orf48, also called NMES1 for "normal mucosa of esophagus specific gene-1" is downregulated in esophageal cancer. In our study, C15orf48 interacts strongly with several OxPhos complexes, enriching for both CI and CIV (Fig. 3.6, top). Its B12D domain is shared by NDUFA4 and NDUFA4L2. NDUFA4 was initially annotated as a subunit of CI, but more recent efforts have shown it regulates CIV activity [179, 180]. On the other hand, NDUFA4L2 is believed to be a negative regulator of CI under cases of hypoxic stress [181].

As C15orf48 interacts with both CI and CIV, it likely regulates the activity of one or both of these complexes. Following the example of NDUFA4L2, it is possible that its regulation is only under certain stress conditions. Consistent with this, the expression of C15orf48 is downregulated and in esophageal cancer [182] due to hyper-methylation [183],

downregulated in metastatic breast cancer [184], upregulated in a mouse model of resistance to TPA-induced cancer [185], yet also upregulated in mucinous ovarian cancer [186] and is diagnostic for squamous cell vs. pseudoepitheliomatous hyperplasia [187]. It may play a role in response to infection, as it is upregulated after vaccinia virus [188] or RSV infection [189], and highly upregulated by the herbicide atrazine [190], IL1-beta [191–193], and CD40 [194]. C15orf48 has been selected as a potential marker for effector dendritic cells [195].

Condition	Related process	C15orf48 expression	References
Esophageal cancer	Cancer	Reduced	[182, 183]
Metastatic breast cancer	Cancer	Reduced	[184]
Resistance to TPA-induced cancer	Cancer	Increased	[185]
Mucinous ovarian cancer	Cancer	Increased	[186]
Squamous cell cancer (vs pseudoepitheliomatous hyperplasia)	Cancer	Increased	[187]
Vaccinia virus infection	Infection	Increased	[188]
RSV infection	Infection	Increased	[189]
IL1-beta treatment	Immune	Increased	[191–193]
CD40 treatment	Immune	Increased	[194]
Effector dendritic cells	Immune	Increased	[195]
Atrazine treatment	Unclear	Increased	[190]

Table 3.1: *C15orf48 expression.* Studies showing significant conditional effects on C14orf58 mRNA or protein expression.

Like NDUFA4, it may interact with only low affinity, explaining why C15orf48 has not previously been observed interacting with these well-studied complexes. Given that it interacts with many members of each

complex, it is less likely to facilitate the assembly of an immature subcomplex, and more likely a supernumerary member of mature CI, CIV, or both in a supercomplex. Supercomplex formation has been shown to regulate ATP production and reactive oxygen species (ROS) generation [196], a process in which C15orf48 may play a role.

C17orf89

We observe a reciprocal enrichment between the baits C17orf89 and NDUFAF5 as each enriched for the other as prey (Fig. 3.6, lower left). While NDUFAF5 enriched for no other CI proteins, C17orf89 pulled out several CI members, albeit just below our stringent cutoff. Both C17orf89 and NDUFAF5 are OMPs, for while it has been established that NDUFAF5 is associated with CI assembly, it is unclear how. NDUFAF5 is predicted to be a methyltransferase, and the methyltransferase motif is necessary for its function in a *Dictyostelium* model [197], but no substrate has been found. Mouse C17orf89 mRNA expression is highly correlated with several CI proteins, according to BioGPS and genenetwork.org [198, 199]. Given that C17orf89 interacts with NDUFAF5, it is tempting to suggest that C17orf89 may facilitate methyltransferase activity of NDUFAF5, or even be the substrate of the transferase.

C19orf70

C19orf70, aka QIL1, interacts with nearly all known subunits of MICOS, including IMMT, DNAJC11, MINOS1, SAMM50, CHCHD6, APOO, APOOL, CHCHD3, MTX1 and MTX2, which is more than the positive control APOOL (Fig. 3.6, bottom center). C19orf70 was identified as a likely regulator of OxPhos as its mRNA expression is highly correlated with that of many OxPhos subunits [39]. According to BioGPS, mouse C19orf70 mRNA expression is highly correlated with transcripts from several OxPhos subunits, consistent with the association with OxPhos, but it is also highly correlated with the MICOS subunits Immt and Minos1 [198]. A different AP-MS study found that C19orf70, SAMM50, and IMMT were preys of CLN3 [200]. C19orf70, along with other MICOS members SAM50, IMMT, CHCHD3, CHCHD6, and APOO are carbonylated in adipocytes [201]. Similar to many members of MICOS, most transmembrane prediction software suggests that C19orf70 does not have a transmembrane domain, although Tmpred identifies a weak transmembrane from residues 17-33.

LYRM proteins

The mitochondrial family of LYR motif-containing proteins were first identified as supernumerary subunits of CI (NDUFA6 and NDUFB9), and have recently been found to act as assembly factors for CII (SDHAF1/LYRM8, SDHAF3/ACN9) [202, 203], CV (FMC1) [204], and iron-sulfur cluster biosynthetic enzymes (ISD11/LYRM4) [205, 206]. We observe many ex-

pected interactions between LYRM proteins and their known direct and indirect complex members (Fig. 3.7).

Several LYRM proteins are known to interact with the mitochondrial acyl carrier protein NDUFAB1 [207]. We find NDUFAB1 as a central hub of LYRM proteins and expand its interactions to include the LYRM4, LYRM7, ACN9, and the OMPs LYRM1, LYRM2, and LYRM5 (Fig. 3.7). The role of these NDUFAB1:LYRM protein interactions is still unclear, and future studies will be necessary to establish the role of NDUFAB1:LYRM interactions for other mitochondrial metabolic complexes.

LYRM5 is a specific interactor of ETF

LYRM1, LYRM2, LYRM5, and LYRM9 are OMPs. Of these, LYRM5 has high-scoring and highly-abundant interactions with both subunits of the electron transfer flavoprotein ETF. Of note, ETFA and ETFB are present in 77 and 84% of all our IPs, respectively. As such, non-quantitative approaches would likely consider these preys as interaction "contaminants," and eliminate them from further study. Instead, our quantitative approach is uniquely positioned to extract meaning from these common interactors. We identified a dramatic and significant enrichment of ETF in the LYRM5-FLAG purifications. Compared to all bait pull downs, the abundance of ETFA and ETFB was at least 5-fold higher with LYRM5 as bait compared to any other bait, including the known ETF substrate isovaleryl dehydrogenase (IVD), and more than 60-fold higher than the average.

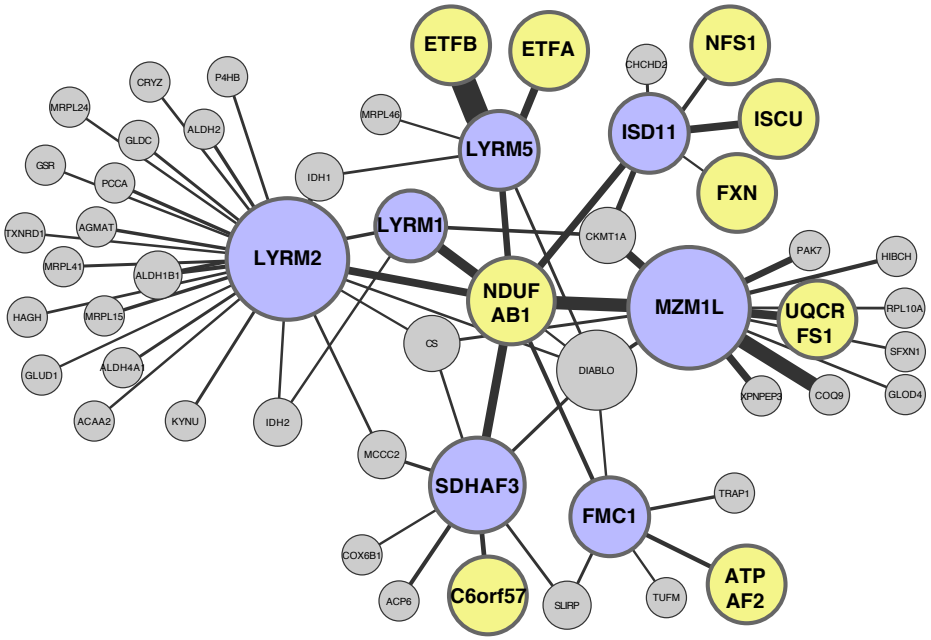


Figure 3.7: *LYRM OMP-related interaction network.* PPIs with LYRM proteins as bait are shown, with LYRM protein nodes in blue, validated binding partners (from the literature or this study) in yellow, and all other mitochondrial preys in gray. Edge width is proportional to the score.

To validate the specificity of the LYRM5:ETF interaction, as well as the accuracy of our MS-based quantification, we performed an anti-FLAG co-IP followed by anti-ETF immunoblot (Fig. 3.8A). While our negative controls of cytosolic GFP and mitochondrial NDUFA4 showed no co-purification of either ETF protein, LYRM5-FLAG efficiently enriched for both ETFA and ETFB. IVD, a known substrate of ETF [208], also enriches for both ETF subunits, but to a much weaker extent than does LYRM5.

Despite the relatively low-expression of LYRM5-FLAG compared to that of IVD-FLAG, more ETF co-purified with the former, showing greater efficiency of the interaction between LYRM5 and ETF. This is consistent with reports that the complex formed between ETF and oxidoreductases is transient in nature [209]. These data support the specific interaction between LYRM5 and ETF, as well as validate our label-free proteomic quantification.

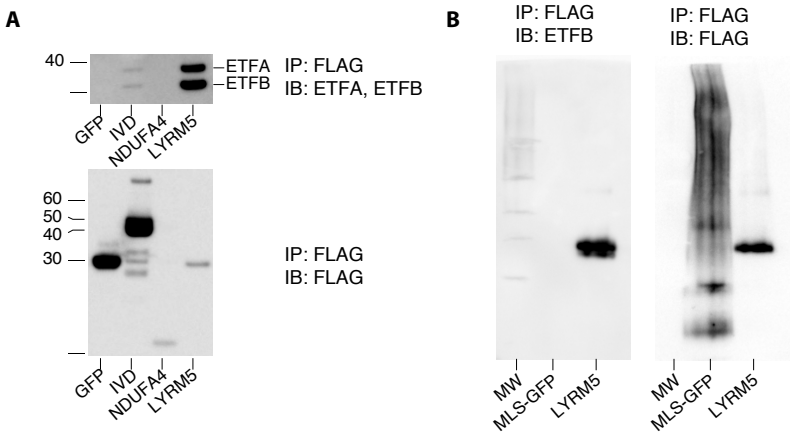


Figure 3.8: *LYRM5 interacts specifically with ETFA and ETFB.* A) C-terminally FLAG-tagged GFP, IVD, NDUFA4, and LYRM5 were expressed in HEK293 cells. After anti-FLAG enrichment, the elution was resolved by SDS-PAGE for immunoblotting. The membrane was probed with anti-ETFA and ETFB (top) or anti-FLAG (bottom) antibodies. LYRM5 efficiently enriched for both ETF proteins to a much greater extent than IVD. ETF was not observed with the negative controls GFP and NDUFA4. B) HEK cells expressing LYRM5-FLAG or MLS-GFP-FLAG were enriched by anti-FLAG purification, and the elution was subjected to Blue Native-PAGE analysis and immunoblotting. The same membrane was blotted for ETFB (left) and then FLAG (right). Abbreviations: IP, immunoprecipitation; IB, immunoblot.

The stable ETF heterodimer interacts in a series of mutually exclusive

complexes with at least 10 different oxidoreductases and ETFDH [209]. Interestingly, LYRM5 does not co-enrich for any of these oxidoreductases. All of the structurally diverse oxidoreductases are believed to bind the same or overlapping motifs [209] on the beta subunit of ETF [210]. As the interaction of LYRM5 with ETF *in vivo* appears to exclude the oxidoreductases, it is possible that LYRM5 binds this same motif. This is similar to the example of the LYRM proteins SDHAF1 and SDHAF3; binding of SDHAF1 and SDHAF3 to SDHB precludes SDHC and SDHD from binding SDHB, as these proteins likely bind the same motif.

To test whether LYRM5 interacts with both ETFA and ETFB simultaneously or in two separate complexes, we first immunoprecipitated LYRM5-FLAG from HEK293 cells, and then resolved the eluate by BN-PAGE followed by immunoblot (Fig. 3.8B). Across two experiments, we observed that the LYRM5 IP elution generated a single complex containing LYRM5, ETFA, and ETFB. MLS-GFP-FLAG did not enrich for either ETFA or ETFB. These data suggest that all three proteins interact specifically in one complex *in vivo*. Furthermore, we observed a single band in the LYRM5 IPs, indicating that the majority of LYRM5 is in complex with ETFA and ETFB.

Bioinformatic clues for LYRM5

The LYRM5 protein is highly conserved, with 87, 66, and 44% sequence identity in *M. musculus*, *C. elegans*, and *Y. lipolytica*, respectively. Based on

CLIME [211] and reciprocal protein BLAST analysis, we identify more than 60 organisms that harbor a LYRM5 homolog, beginning with *T. brucei*. Interestingly, no organisms contain a homolog of LYRM5 without a homolog of ETFA (Fig.3.9A). This is consistent with the role of LYRM5 in regulating ETF activity.

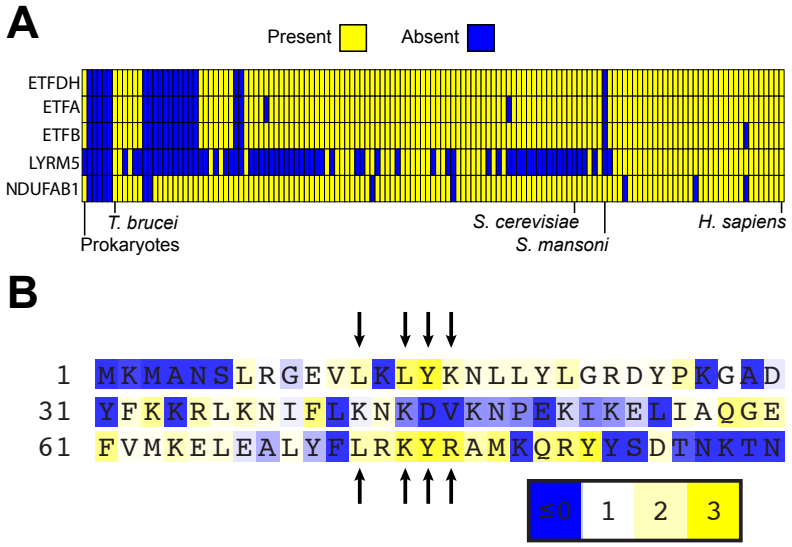


Figure 3.9: *LYRM5 protein and residues are highly conserved.* A) Presence/absence matrix analysis of ETF, LYRM5, and NDUFAB1. Results are the integration of CLIME and reciprocal protein BLAST analysis. Informative species are marked along the bottom. B) LYRM5 sequence conservation. Higher scores are shown in yellow, and are indicative of greater conservation. the two LYR motifs are identified with arrows.

Based on its primary sequence, LYRM5 harbors a pair of LYR motifs, one beginning at residue 12, and the other at 72 (Fig. 3.9B). The LYR motif of the Complex1_LYR-like superfamily is named for the three amino acids leucine-tyrosine-arginine, and the consensus sequence is

Lx(L/K)(Y/I)(R/K)xnF where x is any single residue, and n is a variable span. The two LYR motifs in LYRM5 are L_{12}KLYK_{16} and L_{72}RKYR_{76}, with F40 as the likely key phenylalanine. These motifs are highly conserved from humans through a variety of lower organisms (Fig. 3.9B).

Studies of other LYRM proteins have shown that the LYR motif can be important for facilitating PPIs [212, 213]. As such, we assessed the essentiality of these motifs by mutating the residues to alanine and assessing whether or not the mutant protein was able to interact with ETF. Notably, K14, the first highly-conserved residue of the first LYR motif, is dispensable, for the K14A mutant was still able to interact with ETF. Assessment of the other mutants is ongoing. This information suggests either that the LYR motifs are essential for correct protein folding and maintained stability, or that they are key to facilitate PPIs.

The interaction between LYRM5 and ETF is direct

While we hypothesized that the interaction between LYRM5 and ETF was direct, we sought to determine this experimentally. This is consistent with the LYRM proteins FMC1 and ISD11, as their top-scoring interactions are known to be direct in nature. As with our elucidation of the architecture of the CoQ biosynthetic complex, we decided to test the LYRM5 interactions *in vitro*.

To date, no LYR-motif proteins have yet been purified. Through the combined efforts of our structural-genomics consortium, we have had

success with LYRM5. LYRM5 was purified as a fusion protein with an N-terminal His-MBP (maltose-binding protein) tag to facilitate solubility. Upon incubating tagged LYRM5 with purified ETF *in vitro*, we observed a complex that migrated at a higher weight on BN-PAGE than any of the subunits (Fig. 3.10), consistent with a direct interaction. This complex was observed only when both LYRM5 and ETF were present, was not observed in the presence of the non-target His-MBP-ADCK3, (and was at greater abundance when a 4:1 molar ratio of LYRM5:ETF was used). These data suggest that the *in vivo* interaction between LYRM5 and ETF is direct, and may imply weak affinity between His-MBP-LYRM5 and ETF in our *in vitro* system.

As mentioned above, tagged LYRM5 protein can interact directly with ETF and forms a specific, stable complex *in vitro* and *in vivo*. To exclude the possibility that these interactions are tag mediated, we removed the His-MBP tag on the *E. coli* expressed LYRM5 with tobacco etch virus (TEV) cleavage, leaving only a single serine scar. Cleaved LYRM5 is competent to interact with ETF *in vitro* and forms a stable higher molecular weight complex as assessed by HPLC analysis (Fig. 3.11).

The direct, specific interaction between LYRM5 and ETF is the first functional characterization of LYRM5. As such, we suggest it be renamed ETFAF, for ETF associated factor.

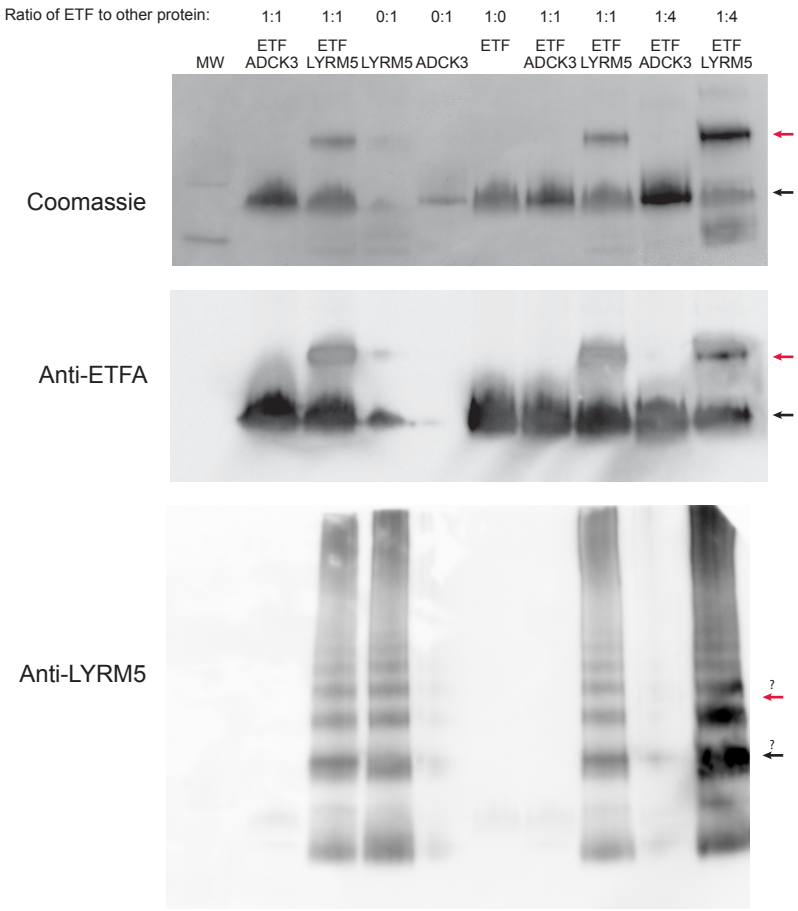


Figure 3.10: *LYRM5 interacts with ETF* in vitro. Proteins were mixed and, after room temperature incubation for 2 h, subjected to Blue Native-PAGE separation and transferred to PVDF membrane. The membrane was visualized with Coomassie stain for total protein (top) and then blotted for ETFA (middle) and LYRM5 (bottom). The position of the lower and upper ETF-containing bands are indicated with black and red arrows, respectively.

3.4 Conclusion

In summary, we have extended the annotation of the mitochondrial proteome, identifying proteins of unknown function and simultaneously

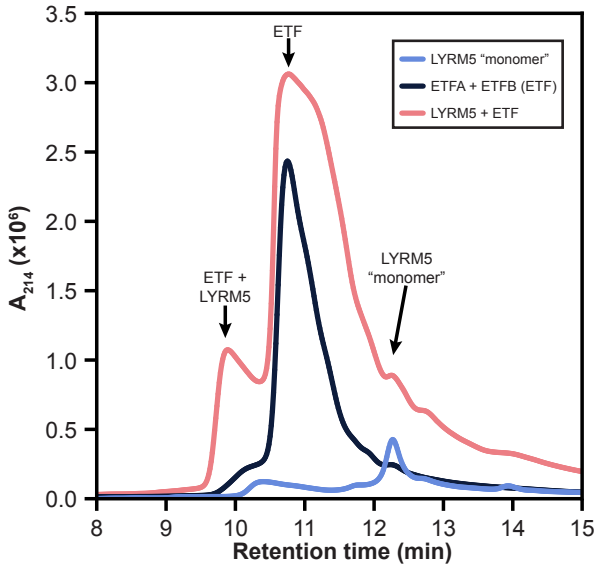


Figure 3.11: *LYRM5 and ETF form a stable complex.* LYRM5 and ETF were incubated alone or in combination at room temperature for one hour, and then subjected to HPLC. Protein abundance was monitored via absorbance at 214 nm.

associating more than 20 with well-characterized pathways. Our work established the presence and dynamic nature of the mammalian CoQ biosynthetic complex. We also establish a direct interaction between LYRM5 and ETF, which is the first known activity of this protein. This resource of OMP associations will continue to facilitate and catalyze the characterization of mitochondrial proteins, and widen the bottleneck of biomedical research.

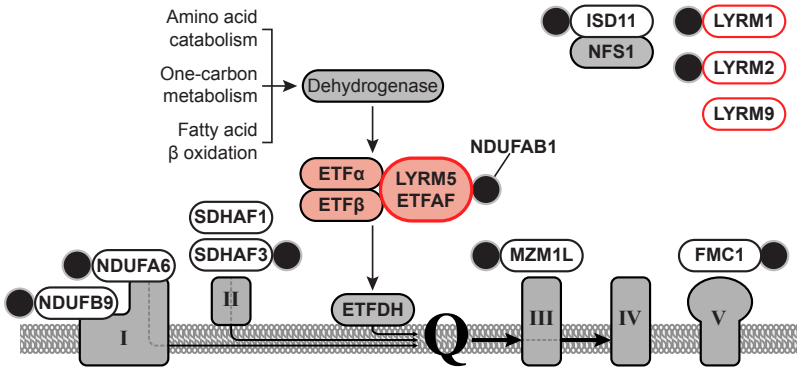


Figure 3.12: *Model of LYRM5/ETFAF interactions in the mitochondrion.* LYRM5/ETFAF interacts with ETFA and ETFB (pink fill) which are essential for electron transfer between several catabolic pathways and complex III of OxPhos. Many other LYRM proteins (white fill) also interact with and/or facilitate assembly of OxPhos complexes, while others (red outline) are OMPs. Additionally, the majority of LYRM proteins have now been found to interact with NDUFAB1, the mitochondrial acyl carrier protein (filled black circle).

3.5 Methods

Mitochondrial proteome compilation

The mouse MitoCarta (Mouse.MitoCarta.xls) was downloaded from The Broad Institute (`http://www.broadinstitute.org/pubs/MitoCarta/`) and Entrez GeneIDs were converted from mouse to human using HomoloGene (build 68). Data from the mitochondrial matrix [14] and IMS [15] proteomic studies were then integrated with the human MitoCarta. This list of 1,143 human mitochondrial proteins was used as the basis for subsequent determination of subcellular localization.

Identification of mitochondrial disease-related genes

Disease-gene annotations were downloaded from OMIM [17], Orphanet [45] (downloaded March 2015) and GeneCards/MalaCards [46, 47]. These were supplemented with mitochondria-specific disease-gene relationships [4].

Generation of tagged constructs

Mitochondrial open reading frames were obtained from The Broad Institute and DNASU [214]. Sequences were compared against UniProt [215] and dbSNP [216]. Any remaining mutations of question were assessed for evolutionary conservation using Homologene. Site-directed mutagenesis was performed by QuikChange PCR and primers were designed by PrimerX software: `http://www.bioinformatics.org/primerx/`.

Mammalian cell culture

HEK293 and HepG2 cells were grown in Dulbecco's modified Eagle's medium (DMEM, LifeTechnologies) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin (LifeTechnologies). Cells were subcultured by trypsinization. On day one, 7 million HEK HepG2 cells were plated in a 15 cm dish and allowed to grow overnight. On day two, cells were transiently transfected with a mix of 20 μg pcDNA3.1 gene-FLAG plasmid, 72 μg linear polyethylenimine (PEI, PolySciences), and 900 μL

Opti-MEM (LifeTechnologies). On day four, cells were washed with PBS and media was replaced with DMEM containing 10% FBS, 1% penicillin-streptomycin, and either 10 mM glucose or 10 mM galactose. After 24 hours (Day 5), cells were washed with and harvested into phosphate-buffered saline (PBS), were collected at 2,000 rcf, were snap frozen in liquid nitrogen, and were stored at −80°C.

Subcellular localization analysis

On the day prior to transfection, HEK293 cells were plated at a density of 75,000 cells/well onto poly-D-lysine-coated coverslips in 6-well dishes. Cells were transiently transfected with a mix of 1 μg pcDNA3.1 gene-FLAG, 0.5 μg plasmid encoding green fluorescent protein with an N-terminal mitochondrial localization sequence (MLS-GFP) [217], 7.5 μg PEI, and 200 μL Opti-MEM. After 24 hours, the cells were fixed (4% paraformaldehyde in PBS), permeabilized (0.2% Triton X-100 in PBS), blocked (1% BSA in PBS), and probed with mouse anti-FLAG M2 1° antibody (F1804, Sigma, 1:2000 (v/v) in 1% BSA in PBS) and Alexa Fluor 594-conjugated goat anti-mouse 2° antibody (LifeTechnologies, 1:2000 (v/v) in 1% BSA in PBS) in 1% BSA in PBS. Hoechst dye (1 μg/mL) was used to label nuclear DNA. Slides were placed in mounting medium (1:1, v/v, glycerol/PBS). Epifluorescent microscopy was performed on the Olympus IX81 microscope using 100X oil immersion optics. Confocal images were captured using the Nikon A1R system, Plan Apo VC 60X oil immersion optics, with sequential laser

excitation using 561 nm (Alexa Fluor), 488 nm (GFP), and 408 nm (Hoechst) lasers. Images were collected and assembled into a Z-stack using the NIS-Elements software.

Affinity enrichment

Digitonin-based approach

Cell pellets were lysed in 200 μL cold lysis buffer (20 mM HEPES, pH 7.40, 100 mM NaCl, 10% glycerol, 3% digitonin (Sigma), 1 mM DTT, protease inhibitors (10 μM benzamide HCl, 1 μg/mL 1,10-phenanthroline and 0.5 μg/mL each of pepstatin A, chymostatin, antipain, leupeptin, aprotinin; Sigma), phosphatase inhibitors (500 μM imidazole, 250 μM NaF, 300 μM sodium molybdate, 250 μM sodium orthovanadate, 1 mM sodium tartrate; Sigma), and deacetylase inhibitors (10 mM each sodium butyrate and nicotinamide; Sigma)), modified from Marbois et al [166]. After periodic vortexing on ice, insoluble materials were pelleted (16,000 g, 10 min, 4°C) and the supernatant was retained. The protein concentration was quantified by Bradford assay, and equal masses of cell supernatant were mixed with 30 μL pre-washed anti-FLAG magnetic beads (Sigma M8823) for 2-3 h at 4°C with end-over-end agitation. Following incubation, beads were washed four times in wash buffer (20 mM HEPES, pH 7.40, 100 mM NaCl, 0.05% digitonin, 10% glycerol) and once in final wash buffer (20 mM HEPES, pH 7.40, 100 mM NaCl). Proteins were eluted in 70 μL elution

buffer (final wash plus 0.2 mg/mL FLAG-peptide) for 30 min at room temperature with constant agitation.

NP-40-based approach

Cell pellets were lysed in 200 μL PBS containing 0.5% IGEPAL CA-630 (NP-40, Sigma), protease inhibitors (10 μM benzamide HCl, 1 μg/mL 1,10-phenanthroline and 0.5 μg/mL each of pepstatin A, chymostatin, antipain, leupeptin, aprotinin; Sigma), phosphatase inhibitors (500 μM imidazole, 250 μM NaF, 300 μM sodium molybdate, 250 μM sodium orthovanadate, 1 mM sodium tartrate; Sigma), and deacetylase inhibitors (10 mM each sodium butyrate and nicotinamide; Sigma). Insoluble materials were pelleted (16,000 g, 10 min, 4°C), the protein concentration was quantified by Bradford assay, and equal masses of the supernatant were mixed with 30 μL pre-washed anti-FLAG magnetic beads (Sigma M8823) for 2-3 h at 4°C. Following incubation, beads were washed four times in PBS containing 0.5% digitonin and glycerol (or 0.05% IGEPAL CA-630) and once in PBS. Proteins were eluted in 70 μL PBS containing 0.2 mg/mL FLAG-peptide for 30 min at room temperature.

Quantitative mass spectrometry

Protein digestion

Protein pull-downs (40 µL volume) were transferred from 96-well plates to separate 1.5 mL tubes (Eppendorf) containing 40 µL of 3 M Urea, 100 mM Tris (pH 8). Protein was reduced with 5 mM dithiothreitol (incubation at 37°C for 45 minutes) and alkylated with 15 mM iodoacetamide (incubation in the dark, at ambient temperature, for 45 minutes). Alkylation was quenched by adding an additional 5 mM dithiothreitol (incubation at ambient temperature for 15 minutes). Protein was enzymatically digested with 0.8 µg of sequencing-grade trypsin (Promega, Madison, WI) and incubated at ambient temperature overnight. An additional 0.4 µg of trypsin was added to each sample the next morning and the resulting mixtures were incubated at room temperature for an hour. Digests were quenched by bringing the pH 2 with trifluoroacetic acid and immediately desalted using C18 solid-phase extraction columns (SepPak, Waters, Milford, MA). Prior to washing peptides with 3 mL of 0.1% TFA on the C18 columns, peptides were washed with 1 mL of 5% acetonitrile/0.1% TFA solution to reduce the abundance of the 1x FLAG peptide from each mixture.

LC-MS/MS analysis

All experiments were performed using a NanoAcquity UPLC system (Waters, Milford, MA) coupled to an Orbitrap Elite mass spectrometer (Thermo

Fisher Scientific, San Jose, CA). Reverse-phase columns were made in-house by packing a fused silica capillary (75 μm i.d., 360 μm o.d., with a laser-pulled electrospray tip) with 3.5 μm diameter, 130 Å pore size Bridged Ethylene Hybrid C18 particles (Waters) to a final length of 30 cm. The column was heated to 55°C for all experiments. Samples were loaded onto the column for 12 minutes in 95:5 buffer A [water, 0.2% formic acid, and 5% DMSO] : buffer B [acetonitrile, 0.2% formic acid, and 5% DMSO] at a flow-rate of 0.40 μL/min. Peptides were eluted using the following gradient: an increase to 22% B over 32 min, followed by a 5 min linear gradient from 22% to 28% B, followed by a 3 min linear gradient from 28% to 70% B which was held for 3 minutes. The column was equilibrated with 5% buffer B for an additional 15 min. Precursor peptide cations were generated from the eluent through the utilization of a nanoESI source.

Mass spectrometry instrument methods consisted of MS^1 survey scans ($1x10^6$ target value; 60,000 resolution; 300 Th - 1500 Th) that were used to guide fifteen subsequent data-dependent MS/MS scans (2 Th isolation window, HCD fragmentation, normalized collision energy of 30; $5x10^4$ target value, 15,000 resolution). Dynamic exclusion duration was set to 45 s, with a maximum exclusion list of 500 and an exclusion width of $\pm$10 ppm around the selected average mass. Maximum injection times were set to 50 ms for all MS^1 scans and 200 ms for MS/MS scans.

Data Analysis

Data was processed using the in-house software suite COMPASS [153]. OMSSA [159] (version 2.1.8) searches were performed against a target-decoy database (Uniprot (human) database, `www.uniprot.org`, April 4th, 2014, containing the 1X FLAG and MFP-GFP-FLAG sequences). Searches were conducted using a 150 ppm precursor mass tolerance and a 0.015 Da product mass tolerance. A maximum of 3 missed tryptic cleavages were allowed. The fixed modification specified was carbamidomethylation of cysteine residues. The variable modification specified was oxidation of methionine. Results were filtered to 1% FDR at the unique peptide level using COMPASS. Peptides identified within each bait set (6 replicates total - 3 grown in dextrose, 3 grown in galactose) were grouped into proteins according to previously reported rules [161] using COMPASS.

Data was also processed using the MaxQuant software suite.[218, 219] Searches were performed against a target-decoy database (Uniprot (human) database, `www.uniprot.org`, April 4th, 2014, containing the 1X FLAG and MFP-GFP-FLAG sequences) using the default settings for high-resolution mass spectra. A maximum of 2 missed tryptic cleavages were allowed. The fixed modification specified was carbamidomethylation of cysteine residues. The variable modification specified was oxidation of methionine. Results were filtered to 1% FDR at the unique peptide level and grouped into proteins within MaxQuant. Proteins were quantified across all replicates within each bait set using MaxLFQ [71].

Cell-free expression and purification.

Plasmid DNA was treated with 0.05 μg/μL proteinase K to remove trace amounts of RNase, purified, and used as individual transcription templates with SP6 RNase polymerase. Transcription and translation methods are as previously described [220]. Briefly, transcription reactions included 0.2 mg/mL DNA, 20 mM magnesium acetate, 2 mM spermidine trihydrochloride, 10 mM DTT, 80 mM Hepes-KOH, pH 7.8, 4 mM each NTP, pH 7.0, 1.6 U/μL SP5 RNA polymerase (Promega), and 1 U/μL RNasin (Promega). Transcriptions were incubated for 4 h at 37°C. Nonpurified transcription reactions were used singly or mixed 1:1 for pairwise interaction monitoring and added to wheat germ extract (WEPRO2240; CellFree Sciences) in a standard dialysis cup reaction [220]. Each 25 μL reaction contained 60 O.D. wheat germ extract (6.25 μL), 24 mM Hepes-KOH, pH 7.8, 100 mM potassium acetate, 6.25 mM magnesium acetate, 0.4 mM spermidine trihydrochloride, 4 mM DTT, 1.2 mM ATP, 0.25 mM GTP, 16 mM creatine phosphate, 0.0005% sodium azide, 0.04 mg/mL creatine kinase, 0.3 mM each amino acid (amino acid mixture adjusted to pH 7.0 with KOH), and 5 μL of nonpurified RNA. Thirty-two-fold excess dialysis buffer was used for each reaction, containing all reaction components except for wheat germ extract, creatine kinase, and RNA. After assembly of dialysis cups and 18-h incubation at 22°C, duplicate translations (50 μL) of each combination were pooled and centrifuged for 5 min at 20,000 rcf at 10°C. The soluble fraction was removed and added to 20 μL of StrepTactin

resin equilibrated in 25 mM Hepes, pH 7.8, 150 mM NaCl, and 1 mM DTT in a 96-well filter plate (HTS Multiscreen; Millipore). The resin with bound protein was washed three times with 150 μL of binding buffer. Strep(II)-tagged protein and protein complexes were eluted in 2.5 mM desthiobiotin in the same buffer. After sampling for SDS/PAGE, the eluate was applied to Ni-Sepharose (GE Healthcare) in 50 mM Hepes, pH 8.0, 300 mM NaCl, 50 mM imidazole, and 1 mM DTT and subjected to IMAC. Bound samples were eluted by increasing the imidazole concentration to 500 mM. Samples of both the StrepTactin and IMAC elutions were loaded without heating on 4-20% Stain-free TGX Criterion gels (Bio-Rad) and subjected to denaturing SDS/PAGE. Gels were imaged by tryptophan fluorescence using a Bio-Rad Stain Free Imager, followed by staining in Coomassie Brilliant Blue R-250.

Protein gels

SDS-PAGE

After competitive elution with FLAG peptide, the eluate was mixed with protein loading buffer and run on a 10% gel for 90 min at 150 V.

Blue Native-PAGE

NativePAGE Novex 4-16% Bis-Tris protein 10-well gels (LifeTechnologies) were prepared with anode buffer (50 mM Bis-Tris) and dark cathode buffer

(15 mM Bis-Tris/50 mM tricine, 0.02% Serva Blue G-250). Samples were mixed with NativePAGE sample buffer (LifeTechnologies, BN2003) to 1x final concentration, and were loaded onto gels alongside NativeMARK standard (LifeTechnologies). Gels were run at 150 V for 60 min at 4°C. The dark cathode buffer was then replaced with light cathode buffer (0.002% Serva Blue G-250), and the gels were run for another 60 min at 250 V.

For immunoblotting, gels were transferred to PVDF membranes as normal (100 V, 1 h, 4°C). The membrane was then incubated in 8% glacial acetic acid in water (to fix proteins) for 15 min to fix the proteins. After rinsing with water, methanol was added to the membrane to remove the majority of the Serva Blue stain. Immunoblot proceeded as normal.

Immunoblotting

Protein gels (SDS-PAGE or BN-PAGE) were transferred to activated PVDF membranes at 100 V for 1 h at 4°C. Membranes were washed, visualized with Ponceau S (SDS) or incubated with methanol (native). Membranes were blocked with 2% milk in tris-buffered saline with Tween 20 (0.5%), then incubated overnight with antibody at 4°C.

Recombinant proteins

Recombinant ETF was expressed in *E. coli* and stored in ETF storage buffer (10 mM K phosphate, pH 7.6, 150 mM KCl, 20% glycerol). His-MBP-LYRM5 was stored in storage buffer L (10 mM MES, pH 6, 150 mM KCl,

10 mM MgCl2, 0.018% DDM, 0.3 mM TCEP). His-MBP-Nd254-ADCK3 (N-terminal truncation) was expressed in *E. coli*.

In vitro interactions

Purified proteins (His-MBP-LYRM5, His-MBP-ADCK3, ETFA:ETFB) were aliquoted and mixed in Interaction Buffer (20 mM HEPES, pH 7.40, 100 mM KCl, 5% glycerol). Each tube contained 3.64 μM ETF, and LYRM5 or ADCK3 at 3.64 or 14.56 μM, in a total volume of 10 μL. Proteins were allowed to interact for two hours at room temperature or 4°C with periodic mixing before addition of NativePAGE loading dye and BN-PAGE analysis at 4°C.

Measurement of oxygen consumption

All respiration assays were performed using an XF-96 Extracellular Flux Analyzer (Seahorse Bioscience) with a 4-port injection system. On the day prior to the assay, Seahorse 96-well plates were coated with poly-D-lysine, and trypsinized HEK-293 or HepG2 cells were plated at 25,000 cells per well. After calibration of the assay plate using the XF96 software, cells were loaded into the instrument and subjected to a standard mitochondrial stress test protocol. All data analysis was performed using the XF96 software (version 1.4.1.4), and the "Level(Direct)Akos" algorithm that is a built-in factor of the software package.

Sequence analysis

Protein homologs were defined by CLIME [211], Homologene (release 68), and reciprocal protein BLAST searches [221]. LYRM5 protein sequence alignment was computed with CLUSTAL Omega for all 68 known homologs. Using this alignment, amino acid conservation was computed at `http://compbio.cs.princeton.edu/conservation/` [222] with the following options: Scoring method: Sum of pairs, Window size: 3, Sequence weighting: True, Background: SwissProt, Matrix: BLOSUM62.

4 CHAPTER FOUR

Future directions and impact

Brendan J. Floyd

4.1 Abstract

Together, my work has touched upon three major themes: protein modifications, interactions, and functions. In this concluding chapter, I will discuss ideas and future directions related to the identification and analysis of protein-protein interactions and the annotation of mitochondrial function via the example of the OMP LYRM5.

4.2 Future directions for PPI network analysis

Interaction mapping via quantitative mass spectrometry

As the quality of mass spectrometry instrumentation and analysis has increased, so too has our knowledge of the inner workings of the cell. Related to the assessment of protein modifications and interactions, we can conceive of a future in which not only can we identify all cellular proteins, but also most, or even all, peptides [29]. First, this will facilitate the discovery of a greater diversity of PTMs and modified residues. Second, this will allow for more sophisticated AE-MS studies. On this latter point, it is common in the field to consider the enrichment stage as playing a key role in the process. If, however, we could identify every peptide in a cell, all one would need to do is to subtly enrich for a protein of interest, and then correlate the expression of all peptides to your target protein. This would overcome issues related to low abundance proteins, and may

facilitate the identification of more membrane protein interactions and on more acute time scales. This view is in line with recent work from the Mann Group [70].

While early attempts at using mass spectrometry to identify protein-protein interactions focused on the biochemical purity of the target complex, recent efforts have recognized that this can be to the detriment of the study [223]. As such, many investigators have transitioned from approaches, such as tandem affinity purification (TAP), which were designed to be amenable to strict purification conditions. Instead, the Mann Group has advocated for single-step purifications and the term *affinity enrichment* (AE-MS) rather than AP-MS. This appears to be more an adjustment of thought than anything else, but it applies both to the experimental setup as well as the data analysis. In AE-MS theory, the entire cellular proteome is present in the enriched sample, albeit at vastly different concentrations compared to the *in vivo* environment. PPIs will cause prey proteins to correlate with the bait and increase relative abundance compared to the entire cell. In some ways, it is a combination of correlome profiling PCP with affinity enrichment as the initial separation stage, and the computational framework underlying the analysis reflects this.

High-throughput assessment of direct interactions

Another major issue to be tackled is the systematic analysis of direct interactions. While AE-MS can prioritize pathways and processes by associating

a protein of unknown function with one or several members of the pathway, it cannot yield information as to the topology of the complex. Often, this information is key, as it can suggest specific mechanisms by which one protein affects the function of another.

Several methods have been used to elucidate direct PPIs. The most common is the yeast two hybrid (Y2H). Over the past decade, the Vidal Group has published a series of updated Y2H-identified interactions involving the "human ORFeome" or a set of open reading frames (ORFs) [55, 224, 225]. They have now tested 82,000,000 pairwise interactions, out of a possible 200,000,000 given one isoform for each of 20,000 genes. One key benefit of this technology is that interactions observed are likely to be direct, and it is clear what proteins are involved. There are, however, a number of caveats: the interactions are tested in yeast, the proteins often must translocate to the nucleus (this is not the case in some newer variants of Y2H), expression levels of the proteins is not taken into consideration, and proteins that directly bind only under certain conditions or in the presence of other stabilizing factors will not be observed.

Given the above caveats of Y2H, other technologies have been established in an attempt to measure direct interactions *in vivo* in human cells. Of these, bioluminescence resonance energy transfer (BRET and FRET) and bimolecular fluorescence complementation (BiFC) are two of the next most widely-used. BRET, is reversible and measures very short distances between proteins, but requires the addition of fairly large (30 kDa) fusion

tags. BiFC, can utilize slightly shorter tags, especially with the new tripartite variant, but it is irreversible with a half life of 22 years for split GFP, and it still requires genetic manipulation of the system. Both BRET and BiFC must be optimized for the protein pair, as tagging certain regions of the protein may cause unfolding or mislocalization as with all tags, but the topological orientation of the fluorescent proteins may generate a false negative if the orientation or distance is outside their reach.

Differential network approaches

To date, the vast majority of work on physical protein-protein interactions has focused simply on building out the network. As this process is still in its relatively early stages [57], this is a key first stage. That said, this in some ways reduces a highly-dimensional aspect of the cell to a fairly flat and limited understanding. For instance, what is the role of cell type, differentiation status, nutritional and metabolic state, various post-translational modifications, phamacologic treatments, or dozens more cellular modulations on each of these identified PPIs? Which PPIs are overlooked because the cell is not interrogated under the appropriate conditions? What is the temporal nature or affinity of these interactions?

The aforementioned questions form the basis of the growing trend toward differential network biology [226]. Differential approaches are a natural extension of network approaches, and are indispensable for other high-throughput technologies like transcriptomics, metabolomics, and

proteomics. In our analysis of mitochondrial acetylation, for example, not only was it important that we observed acetylation events, but we found that the regulatory sites on the protein Acat1 were those that were differentially regulated in at least one metabolic state [33]. The mere presence of an anlyte or PPI, while informative, yields only part of the story.

Although global analyses of the modulation of PPIs due to cellular events have not been undertaken to any great extent, their presence in signalling pathways forms the core of entire fields of molecular and cellular biology. For example, receptor tyrosine kinase activation by ligand often leads to auto-phosphorylation as both an activating signal as well as a key substrate for other protein binding. Innumerable other examples exist, but have not been studied via any systematic approaches. With the human interactome now in some state of development, it will be easier for other investigators to apply their favorite perturbation and interrogate whether interactions persist or are modulated in abundance.

Additionally, this network analysis can be extended to include mutant forms of proteins, whether they reflect disease-relevant mutations or biochemically-informative mutations such as those for predicted interface or otherwise key residues. Indeed, several groups are undertaking analyses of the role of disease-associated missense mutations on PPIs, termed "edgeotype" analysis [227, 228]. Some of these mutations form "edgetic" alleles, that is, the mutation that affects only some of the binding partners

of a given bait, changing the number and diversity of the edges connecting the bait and prey nodes. Much of this work has been done in the setting of Y2H, but the techniques could easily be applied to AE-MS, complexome profiling, or nearly any other method of PPI-detection.

Edgeotypes may explain a great number of disease-relevant mutations that have so far eluded mechanistic understanding. Given that many binding sites are nucleated by just a few residues, mutation of one or several could increase or decrease affinity of the bait:prey interaction, or even yield novel preys for a given bait. As the only known function of many proteins, including assembly factors or other scaffolding proteins, is to form PPIs, such small differences in edgeotype may be enough to start or catalyze a disease process.

Furthermore, if edgeotype is expanded to include various means of pathology, it is rational that such perturbations would be observable in patient samples. In the same way that the mapping of the human genome has lowered the activation energy necessary to map causal mutations, so, too, will the establishment of a human interactome in any state facilitate the identification of pathological deviations from the norm. Clinical interactome mapping is currently out of the realm of analysis for the majority of the commonly used PPI-mapping methods, Y2H and AE-MS, as each requires genetic manipulation of a model system. There is some promise, however, in newer methods like complexome profiling or soluble complex analysis. As these generally only can identify the proteins of

higher abundance, we may expect to see that edgeotype differences in a protein of low abundance might affect complex formation (as with the CI assembly factors), or such methods will have to be extended to facilitate the identification of low abundance proteins as with targeted MS analysis.

4.3 The OMP LYRM5/ETFAF

LYRM5 and ETF

LYRM5 was first named Ghiso as it was found to be a growth hormone-inducible soluble factor [229]. Early studies implicated this gene as a member of a chromosomal region responsible for increased cancer progression in mice, called *Pulmonary Adenoma Susceptibility 1* locus (Pas1) [229–231]. It was later determined that the Kras gene, which is 160 bp away from Lyrm5 in a tail-to-tail orientation, is likely the key factor for this disease. While no allelic-effects have been found related to the Lyrm5 protein in mice [229, 230], transient expression of a C-terminally V5-His tagged form of mouse Lyrm5 led to 3-fold reduced cell growth in NCI-H520 cells, but no effects in A549 cells, both of which are human non-small cell lung carcinoma cell lines.

Our data show that LYRM5 interacts with ETFA, ETFB, and NDUFAB1. ETFA and ETFB form the heterodimeric ETF complex, necessary for transfer of electrons from at least 11 dehydrogenases to ETFDH (aka ETF:QO) and then to coenzyme Q (CoQ). The cofactors for ETFA and ETFB are FAD

and AMP, respectively.

NDUFAB1

NDUFAB1 (mitochondrial acyl carrier protein, CI 8 kDa protein) is known to bind nearly all LYRM proteins [207]. It was first characterized as an essential supernumerary subunit of complex I (CI), where it binds both NDUFA6 (LYRM6) and NDUFB9 (LYRM3). There are likely two copies of NDUFAB1 for every active CI. Our study finds it an abundant and specific interactor of LYRM1, LYRM2, LYRM4/ISD11, LYRM5, LYRM7/MZM1, FMC1, and ACN9/SDHAF3.

Despite its wide number of binding partners, the specific role of NDUFAB1 remains unclear. It is required for CI activity [212] and the lipoylation of several mitochondrial enzymes including PDC and BCKDC [232]. Additionally, knockdown of NDUFAB1 causes cell death within 72 hours [232]. The *S. cerevisiae* homolog of NDUFAB1 is Acp1p, whose gene knockout is inviable. Interestingly, despite interacting with at least 9 LYRM proteins, each of which has specific binding partners, it does not appear to generate a supercomplex of all LYRM proteins and their substrates. This suggests a model whereby each LYRM protein acts as a bridge connecting NDUFAB1 with each substrate.

Several other proteins are also known to interact with NDUFAB1. Alpha-1-microglobulin (A1M), a redox scavenging molecule, has been shown to interact with NDUFAB1 and bind to mitochondria, reducing

redox stress and maintaining the structural integrity of this organelle [233, 234]. This is particularly interesting because there are reactive oxygen species (ROS) phenotypes associated with modulation of each of several LYRM proteins.

MRPL46

Both LYRM5 (this work) and ETFA (unpublished results from Huttlin et. al.) bind MRPL46 as prey, while MRPL46 as bait enriches for NDUFAB1 as prey. MRPL46 is a subunit of the large mitochondrial ribosome with a modified nudix domain and no similarity to any cytosolic ribosomal protein. Nudix proteins are phosphohydrolases and appear to bind nucleoside diphosphates but can act on other substrates. Acyl- or free-CoA, when hydrolyzed by a Nudix protein (Y87G2A.14 in *C. elegans* and Nudt7 in *M. musculus*), generates ADP and phosphopantetheine [235, 236]. Phosphopantetheine is the essential NDUFAB1 cofactor. It is tempting to suggest that MRPL46 could moonlight as a phosphopantetheine hydrolase to faciliate generation or even addition of cofactors for both NDUFAB1 and ETFB. The yeast homolog of MRPL46 is Mrpl17p. Both Mrpl17p and Aim45p, the homolog of ETFA, are necessary for mitochondrial genome maintenance.

Regulation of LYRM-related proteins

It has been shown that siRNA-mediated knock down of NDUFAB1 affects CI and CII activity as well as ROS in HEK293 cells after 48-96 hours [232], consistent with its PPIs with NDUFA6, NDUFB9, and SDHAF3. Knock down of SDHAF3 also affects mitochondrial ROS. We speculate that the octanoyl/lipoyl group on NDUFAB1 is an essential ROS scavenger at each of these ROS-forming enzymes. ETF has also been shown to produce substantial amounts of both superoxide and hydrogen peroxide, most effectively when partnered with ACADM [237]. Interestingly, a recent study has shown that ethanol, likely via ROS, induces increased 4-HNE adduction of ETFA as one of the main mitochondrial proteins modified in this way [238].

We speculate that LYRM5 is necessary for and mediates the interaction between ETF and NDUFAB1. This interaction may also be essential for appropriate NDUFAB1 abundance and, likely, CI activity. This chain may explain the prediction of LYRM5 as a CI assembly factor, and that knock down of LYRM5 leads to a modest reduction in CI activity [13].

Further questions related to LYRM proteins

Many essential questions remain regarding the LYRM family of proteins, including 1) what is their specific mechanism of action 2) what is the role of NDUFAB1, and 3) how do their binding partners differentiate between

one LYRM protein and the rest?

These latter two points are particularly interesting, as each LYRM protein has one common (NDUFAB1) and one or more unique direct interactors. As the family is defined by the LYR motif, it seems likely that this motif is responsible for the interactions with NDUFAB1. Consistent with this, several studies have noted the LYR motif is necessary and sufficient for PPIs [212, 213]. Alternatively, the LYR motif may simply nucleate the structure of these proteins and not play a direct role in forming the PPI interface.

The precise mechanism by which LYR motif-containing proteins facilitate the function and/or assembly of their client complexes has only been suggested for a small number of LYRM proteins. NDUFA6 is now known to tether NDUFAB1 to CI [212], which is necessary for CI function, albeit for an unclear reason. Interestingly, loss of these proteins is associated with reactive oxygen species (ROS) stress.

4.4 Impact

This work holds great promise for the study of uncharacterized proteins, particularly orphan mitochondrial proteins. As many scientists work on pathways and processes that still have gaps in knowledge of what proteins carry out certain transformations, the techniques outlined here may serve as examples for the elucidation of such pathways.

Our PPI network has identified a number of proteins that are only one degree away from genes with disease-causing mutations. As such, they are prioritized as other possible causes of similar diseases without molecular diagnoses. This problem of patients with a well-characterized disease but having no mutations in the commonly identified genes is a growing issue despite the advancements in genome sequencing.

LYRM5/ETFAF is associated with ETF, a cause of glutaric acidemia type II (GAII, multiple acyl-CoA dehydrogenase deficiency, MADD). We have now identified nearly a dozen patients from three continents who have GAII without pathogenic mutations in ETFA, ETFB, or ETFDH. It is possible that their disease is caused by some modulation of ETFAF activity.

Several other OMPs have PPIs with disease-causing genes. C2orf47 interacts with AFG3L2, SPG7, PHB, and PHB2, all of which cause neurological disease including spastic paraplegia. C15orf48 and C17orf89 are both associated with OxPhos complexes I and or IV, and as many as 50% of patients with loss of CI or CIV function have no known causal mutation. C19orf70/QIL1 and OCIAD1 interact with the MICOS complex, many of the subunits of which are associated with cardiomyopathy or lipid dysregulation. Thus, further characterization of these OMPs may substantiate their important role in both health and disease, and should therefore be a priority of such research.

BIBLIOGRAPHY

[1] Claude, A. Fractionation of mammalian liver cells by differential centrifugation: I. Problems, methods, and preparation of extract. *The Journal of experimental medicine* **84,** 51–59 (June 1946).

[2] Friedkin, M & Lehninger, A. L. Esterification of inorganic phosphate coupled to electron transport between dihydrodiphosphopyridine nucleotide and oxygen. *Journal of Biological Chemistry* **178,** 611–644 (Apr. 1949).

[3] Nunnari, J. & Suomalainen, A. Mitochondria: in sickness and in health. *Cell* **148,** 1145–1159 (Mar. 2012).

[4] Koopman, W. J. H., Willems, P. H. G. M. & Smeitink, J. A. M. Monogenic Mitochondrial Disorders. *New England Journal of Medicine* **366,** 1132–1141 (Mar. 2012).

[5] Vafai, S. B. & Mootha, V. K. Mitochondrial disorders as windows into an ancient organelle. *Nature* (2012).

[6] Perocchi, F., Gohil, V. M., Girgis, H. S., Bao, X. R., McCombs, J. E., Palmer, A. E. & Mootha, V. K. MICU1 encodes a mitochondrial EF hand protein required for Ca(2+) uptake. *Nature* **467,** 291–296 (Sept. 2010).

[7] Two-dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry: toward a human mitochondrial proteome. *Electrophoresis* **19,** 1006–1014 (May 1998).

[8] High-throughput profiling of the mitochondrial proteome using affinity fractionation and automation. *Electrophoresis* **21,** 3427–3440 (Oct. 2000).

[9] MITOP: database for mitochondria-related proteins, genes and diseases. *Nucleic Acids Research* **27,** 153–155 (Jan. 1999).

[10] Scharfe, C, Zaccaria, P, Hoertnagel, K, Jaksch, M, Klopstock, T, Dembowski, M, Lill, R, Prokisch, H, Gerbitz, K. D., Neupert, W, Mewes, H. W. & Meitinger, T. MITOP, the mitochondrial proteome database: 2000 update. *Nucleic Acids Research* **28,** 155–158 (Jan. 2000).

[11] Mootha, V. K., Bunkenborg, J, Olsen, J. V. & Hjerrild, M. Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* (2003).

[12] Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid chromatography coupled with tandem mass spectrometry. *Journal of proteome research* **3,** 495–505 (May 2004).

[13] Pagliarini, D. J., Calvo, S., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S.-E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A. & Mootha, V. K.

A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134,** 112–123 (July 2008).

[14] Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A. & Ting, A. Y. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science (New York, N.Y.)* **339,** 1328–1331 (Mar. 2013).

[15] Hung, V., Zou, P., Rhee, H.-W., Udeshi, N. D., Cracan, V., Svinkina, T., Carr, S. A., Mootha, V. K. & Ting, A. Y. Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Molecular cell* (July 2014).

[16] Taylor, S. W., Fahy, E., Zhang, B., Glenn, G. M., Warnock, D. E., Wiley, S., Murphy, A. N., Gaucher, S. P., Capaldi, R. A., Gibson, B. W. & Ghosh, S. S. Characterization of the human heart mitochondrial proteome. *Nature biotechnology* **21,** 281–286 (Mar. 2003).

[17] Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research* **43,** D789–98 (Jan. 2015).

[18] Calvo, S., Tucker, E. J., Compton, A. G., Kirby, D. M., Crawford, G., Burtt, N. P., Rivas, M., Guiducci, C., Bruno, D. L., Goldberger, O. A., Redman, M. C., Wiltshire, E., Wilson, C. J., Altshuler, D., Gabriel, S. B., Daly, M. J., Thorburn, D. R. & Mootha, V. K. High-throughput,

pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nature Genetics* **42,** 851–858 (Oct. 2010).

[19] Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H. E., Schönfisch, B., Perschil, I., Chacinska, A., Guiard, B., Rehling, P., Pfanner, N. & Meisinger, C. The proteome of Saccharomyces cerevisiae mitochondria. *Proceedings of the National Academy of Sciences* **100,** 13207–13212 (Nov. 2003).

[20] Foster, L. J., de Hoog, C. L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V. K. & Mann, M. A mammalian organelle map by protein correlation profiling. *Cell* **125,** 187–199 (Apr. 2006).

[21] Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A. & Mann, M. Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426,** 570–574 (Dec. 2003).

[22] Anderson, S, Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J, Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F, Schreier, P. H., Smith, A. J., Staden, R & Young, I. G. Sequence and organization of the human mitochondrial genome. *Nature* **290,** 457–465 (Apr. 1981).

[23] Haack, T. B., Madignier, F., Herzer, M., Lamantea, E., Danhauser, K., Invernizzi, F., Koch, J., Freitag, M., Drost, R., Hillier, I., Haberberger, B., Mayr, J. A., Ahting, U., Tiranti, V., Rotig, A., Iuso, A., Horvath, R.,

Tesarova, M., Baric, I., Uziel, G., Rolinski, B., Sperl, W., Meitinger, T., Zeviani, M., Freisinger, P. & Prokisch, H. Mutation screening of 75 candidate genes in 152 complex I deficiency cases identifies pathogenic variants in 16 genes including NDUFB9. *Journal of medical genetics* **49,** jmedgenet–2011–100577–89 (Dec. 2011).

[24] Calvo, S., Compton, A. G., Hershman, S. G., Lim, S. C., Lieber, D. S., Tucker, E. J., Laskowski, A., Garone, C., Liu, S., Jaffe, D. B., Christodoulou, J., Fletcher, J. M., Bruno, D. L., Goldblatt, J., Dimauro, S., Thorburn, D. R. & Mootha, V. K. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Science translational medicine* **4,** 118ra10–118ra10 (Jan. 2012).

[25] Lieber, D. S., Calvo, S., Shanahan, K., Slate, N. G., Liu, S., Hershman, S. G., Gold, N. B., Chapman, B. A., Thorburn, D. R., Berry, G. T., Schmahmann, J. D., Borowsky, M. L., Mueller, D. M., Sims, K. B. & Mootha, V. K. Targeted exome sequencing of suspected mitochondrial disorders. *Neurology* **80,** 1762–1770 (May 2013).

[26] Takata, A., Kato, M., Nakamura, M., Yoshikawa, T., Kanba, S., Sano, A. & Kato, T. Exome sequencing identifies a novel missense variant in RRM2B associated with autosomal recessive progressive external ophthalmoplegia. *Genome biology* **12,** R92 (2011).

[27] Marcadier, J. L., Smith, A. M., Pohl, D., Schwartzentruber, J., Al-Dirbashi, O. Y., FORGE Canada Consortium, Majewski, J., Ferdinandusse, S., Wanders, R. J. A., Bulman, D. E., Boycott, K. M., Chakraborty, P. & Geraghty, M. T. Mutations in ALDH6A1 encoding methylmalonate semialdehyde dehydrogenase are associated with dysmyelination and transient methylmalonic aciduria. *Orphanet journal of rare diseases* **8,** 98 (2013).

[28] Falk, M. J., Pierce, E. A., Consugar, M., Xie, M. H., Guadalupe, M., Hardy, O., Rappaport, E. F., Wallace, D. C., LeProust, E. & Gai, X. Mitochondrial disease genetic diagnostics: optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. *Discovery medicine* **14,** 389–399 (Dec. 2012).

[29] Mann, M., Kulak, N. A., Nagaraj, N. & Cox, J. The coming age of complete, accurate, and ubiquitous proteomes. *Molecular cell* **49,** 583–590 (Feb. 2013).

[30] Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S. & Coon, J. J. The one hour yeast proteome. *Molecular & cellular proteomics: MCP* **13,** 339–347 (Jan. 2014).

[31] Grimsrud, P. A., Carson, J. J., Hebert, A. S., Hubler, S. L., Niemi, N. M., Bailey, D. J., Jochem, A., Stapleton, D. S., Keller, M. P., Westphall, M. S., Yandell, B. S., Attie, A. D., Coon, J. J. & Pagliarini, D. J. A quantitative map of the liver mitochondrial phosphoproteome

reveals posttranslational control of ketogenesis. *Cell metabolism* **16,** 672–683 (Nov. 2012).

[32] Hebert, A. S., Dittenhafer-Reed, K. E., Yu, W., Bailey, D. J., Selen, E. S., Boersma, M. D., Carson, J. J., Tonelli, M., Balloon, A. J., Higbee, A. J., Westphall, M. S., Pagliarini, D. J., Prolla, T. A., Assadi-Porter, F., Roy, S., Denu, J. M. & Coon, J. J. Calorie restriction and SIRT3 trigger global reprogramming of the mitochondrial protein acetylome. *Molecular cell* **49,** 186–199 (Jan. 2013).

[33] Still, A. J., Floyd, B. J., Hebert, A. S., Bingman, C. A., Carson, J. J., Gunderson, D. R., Dolan, B. K., Grimsrud, P. A., Dittenhafer-Reed, K. E., Stapleton, D. S., Keller, M. P., Westphall, M. S., Denu, J. M., Attie, A. D., Coon, J. J. & Pagliarini, D. J. Quantification of mitochondrial acetylation dynamics highlights prominent sites of metabolic regulation. *Journal of Biological Chemistry* **288,** 26209–26219 (Sept. 2013).

[34] Overmyer, K. A., Evans, C. R., Qi, N. R., Minogue, C. E., Carson, J. J., Chermside-Scabbo, C. J., Koch, L. G., Britton, S. L., Pagliarini, D. J., Coon, J. J. & Burant, C. F. Maximal Oxidative Capacity during Exercise Is Associated with Skeletal Muscle Fuel Selection and Dynamic Changes in Mitochondrial Protein Acetylation. *Cell metabolism* **21,** 468–478 (Mar. 2015).

[35] Park, J., Chen, Y., Tishkoff, D. X., Peng, C., Tan, M., Dai, L., Xie, Z., Zhang, Y., Zwaans, B. M. M., Skinner, M. E., Lombard, D. B. & Zhao, Y. SIRT5-mediated lysine desuccinylation impacts diverse metabolic pathways. *Molecular cell* **50,** 919–930 (June 2013).

[36] Zhao, X., León, I. R., Bak, S., Mogensen, M., Wrzesinski, K., Højlund, K. & Jensen, O. N. Phosphoproteome analysis of functional mitochondria isolated from resting human muscle reveals extensive phosphorylation of inner membrane protein complexes and enzymes. *Molecular & cellular proteomics: MCP* **10,** M110.000299–M110.000299 (Jan. 2011).

[37] The mitochondrial deubiquitinase USP30 opposes parkin-mediated mitophagy. *Nature* **510,** 370–375 (June 2014).

[38] Klomp, J. A. & Furge, K. A. Genome-wide matching of genes to cellular roles using guilt-by-association models derived from single sample analysis. *BMC research notes* **5,** 370 (2012).

[39] Baughman, J. M., Nilsson, R., Gohil, V. M., Arlow, D. H., Gauhar, Z. & Mootha, V. K. A computational screen for regulators of oxidative phosphorylation implicates SLIRP in mitochondrial RNA homeostasis. *PLoS genetics* **5,** e1000590 (Aug. 2009).

[40] Jovanovic, M., Rooney, M. S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., Rodriguez, E. H., Fields, A. P., Schwartz, S., Raychowdhury, R., Mumbach, M. R., Eisenhaure, T., Rabani, M., Gennert, D.,

Lu, D., Delorey, T., Weissman, J. S., Carr, S. A., Hacohen, N. & Regev, A. Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science (New York, N.Y.)* **347,** 1259038–1259038 (Mar. 2015).

[41] Baughman, J. M., Perocchi, F., Girgis, H. S., Plovanich, M., Belcher-Timme, C. A., Sancak, Y., Bao, X. R., Strittmatter, L., Goldberger, O., Bogorad, R. L., Koteliansky, V. & Mootha, V. K. Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter. *Nature* **476,** 341–345 (Aug. 2011).

[42] De Stefani, D., Raffaello, A., Teardo, E., Szabò, I. & Rizzuto, R. A forty-kilodalton protein of the inner membrane is the mitochondrial calcium uniporter. *Nature* **476,** 336–340 (Aug. 2011).

[43] Lanning, N. J., Looyenga, B. D., Kauffman, A. L., Niemi, N. M., Sudderth, J., DeBerardinis, R. J. & MacKeigan, J. P. A mitochondrial RNAi screen defines cellular bioenergetic determinants and identifies an adenylate kinase as a key regulator of ATP levels. *Cell reports* **7,** 907–917 (May 2014).

[44] Pagliarini, D. J. & Rutter, J. Hallmarks of a new era in mitochondrial biochemistry. *Genes & development* **27,** 2615–2627 (Dec. 2013).

[45] Rath, A., Olry, A., Dhombres, F., Brandt, M. M., Urbero, B. & Ayme, S. Representation of rare diseases in health information systems:

the Orphanet approach to serve a wide range of end users. *Human mutation* **33,** 803–808 (May 2012).

[46] Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., Sirota-Madi, A., Olender, T., Golan, Y., Stelzer, G., Harel, A. & Lancet, D. GeneCards Version 3: the human gene integrator. *Database* **2010,** baq020–baq020 (2010).

[47] Rappaport, N., Twik, M., Nativ, N., Stelzer, G., Bahir, I., Stein, T. I., Safran, M. & Lancet, D. MalaCards: A Comprehensive Automatically-Mined Database of Human Diseases. *Current protocols in bioinformatics* **47,** 1.24.1–1.24.19 (2014).

[48] Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H. & Wanker, E. E. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122,** 957–968 (Sept. 2005).

[49] Oliver, S. Guilt-by-association goes global. *Nature* **403,** 601–603 (Feb. 2000).

[50] A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403,** 623–627 (Feb. 2000).

[51] Jin, K., Musso, G., Vlasblom, J., Jessulat, M., Deineko, V., Negroni, J., Mosca, R., Malty, R. H., Nguyen-Tran, D.-H., Aoki, H., Minic, Z., Freywald, T., Phanse, S., Xiang, Q., Freywald, A., Aloy, P., Zhang, Z. & Babu, M. Yeast Mitochondrial Protein-Protein Interactions Reveal Diverse Complexes and Disease-Relevant Functional Relationships. *Journal of proteome research,* 141229170901006 (Dec. 2014).

[52] Sanadi, D. R., Littlefield, J. W. & Bock, R. M. Studies on alpha-ketoglutaric oxidase. II. Purification and properties. *The Journal of biological chemistry* **197,** 851–862 (May 1952).

[53] Schägger, H & von Jagow, G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Analytical biochemistry* **199,** 223–231 (Dec. 1991).

[54] Schägger, H & Pfeiffer, K. Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *The EMBO Journal* **19,** 1777–1783 (Apr. 2000).

[55] Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., Kamburov, A., Ghiassian, S. D., Yang, X., Ghamsari, L., Balcha, D., Begg, B. E., Braun, P., Brehme, M., Broly, M. P., Carvunis, A.-R., Convery-Zupan, D., Corominas, R., Coulombe-Huntington, J., Dann, E., Dreze, M., Dricot, A., Fan, C., Franzosa, E., Gebreab, F., Gutierrez, B. J., Hardy, M. F., Jin, M., Kang, S., Kiros, R., Lin, G. N., Luck, K., MacWilliams,

A., Menche, J., Murray, R. R., Palagi, A., Poulin, M. M., Rambout, X., Rasla, J., Reichert, P., Romero, V., Ruyssinck, E., Sahalie, J. M., Scholz, A., Shah, A. A., Sharma, A., Shen, Y., Spirohn, K., Tam, S., Tejeda, A. O., Trigg, S. A., Twizere, J.-C., Vega, K., Walsh, J., Cusick, M. E., Xia, Y., Barabási, A.-L., Iakoucheva, L. M., Aloy, P., De Las Rivas, J., Tavernier, J., Calderwood, M. A., Hill, D. E., Hao, T., Roth, F. P. & Vidal, M. A proteome-scale map of the human interactome network. *Cell* **159,** 1212–1226 (Nov. 2014).

[56] Yang, J.-S., Kim, J., Park, S., Jeon, J., Shin, Y.-E. & Kim, S. Spatial and functional organization of mitochondrial protein network. *Scientific Reports* **3,** 1403 (2013).

[57] Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J. & Barabási, A.-L. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science (New York, N.Y.)* **347,** 1257601 (Feb. 2015).

[58] Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., Shales, M., Mercenne, G., Pache, L., Li, K., Hernandez, H., Jang, G. M., Roth, S. L., Akiva, E., Marlett, J., Stephens, M., D'Orso, I., Fernandes, J., Fahey, M., Mahon, C., O'Donoghue, A. J., Todorovic, A., Morris, J. H., Maltby, D. A., Alber, T., Cagney, G., Bushman, F. D., Young, J. A., Chanda, S. K., Sundquist, W. I., Kortemme, T., Hernandez, R. D., Craik, C. S., Burlingame, A., Sali,

A., Frankel, A. D. & Krogan, N. J. Global landscape of HIV-human protein complexes. *Nature* **481,** 365–370 (Jan. 2012).

[59] Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. Network organization of the human autophagy system. *Nature* **466,** 68–76 (July 2010).

[60] Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138,** 389–403 (July 2009).

[61] Braun, P. Interactome mapping for analysis of complex phenotypes: insights from benchmarking binary interaction assays. *Proteomics* **12,** 1499–1518 (May 2012).

[62] Sun, X., Hong, P., Kulkarni, M, Kwon, Y. & Perrimon, N. An advanced method for identifying protein-protein interaction by analyzing TAP/MS data. *2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM),* 1–6 (2012).

[63] Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sørensen,

B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W. V., Figeys, D. & Tyers, M. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415,** 180–183 (Jan. 2002).

[64] Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A.-M., Cruciat, C.-M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. & Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147 (Jan. 2002).

[65] Richter-Dennerlein, R., Korwitz, A., Haag, M., Tatsuta, T., Dargazanli, S., Baker, M., Decker, T., Lamkemeyer, T., Rugarli, E. I. & Langer, T. DNAJC19, a Mitochondrial Cochaperone Associated with Cardiomyopathy, Forms a Complex with Prohibitins to Regulate Cardiolipin Remodeling. *Cell metabolism* (May 2014).

[66] Weber, T. A., Koob, S., Heide, H., Wittig, I., Head, B., van der Bliek, A., Brandt, U., Mittelbronn, M. & Reichert, A. S. APOOL is a

cardiolipin-binding constituent of the Mitofilin/MINOS protein complex determining cristae morphology in mammalian mitochondria. *PLoS one* **8,** e63683 (2013).

[67] Gan, X., Kitakawa, M., Yoshino, K.-I., Oshiro, N., Yonezawa, K. & Isono, K. Tag-mediated isolation of yeast mitochondrial ribosome and mass spectrometric identification of its new components. *European journal of biochemistry / FEBS* **269,** 5203–5214 (Nov. 2002).

[68] Richter, R., Rorbach, J., Pajak, A., Smith, P. M., Wessels, H. J., Huynen, M. A., Smeitink, J. A., Lightowlers, R. N. & Chrzanowska-Lightowlers, Z. M. A functional peptidyl-tRNA hydrolase, ICT1, has been recruited into the human mitochondrial ribosome. *The EMBO Journal* **29,** 1116–1125 (Mar. 2010).

[69] Guarani, V., Paulo, J., Zhai, B., Huttlin, E. L., Gygi, S. P. & Harper, J. W. TIMMDC1/C3orf1 functions as a membrane-embedded mitochondrial complex I assembly factor through association with the MCIA complex. *Molecular and Cellular Biology* **34,** 847–861 (Mar. 2014).

[70] Keilhauer, E. C., Hein, M. Y. & Mann, M. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Molecular & cellular proteomics: MCP* **14,** 120–135 (Jan. 2015).

[71] Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics: MCP* **13,** 2513–2526 (Sept. 2014).

[72] Morris, J. H., Knudsen, G. M., Verschueren, E., Johnson, J. R., Cimermancic, P., Greninger, A. L. & Pico, A. R. Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions. *Nature Protocols* **9,** 2539–2554 (Nov. 2014).

[73] Choi, H. Computational detection of protein complexes in AP-MS experiments. *Proteomics* **12,** 1663–1668 (May 2012).

[74] Nesvizhskii, A. I. Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics* **12,** 1639–1655 (May 2012).

[75] Teng, B., Zhao, C., Liu, X. & He, Z. Network inference from AP-MS data: computational challenges and solutions. *Briefings in bioinformatics* (Nov. 2014).

[76] Armean, I. M., Lilley, K. S. & Trotter, M. W. B. Popular computational methods to assess multiprotein complexes derived from label-free affinity purification and mass spectrometry (AP-MS) experiments. *Molecular & cellular proteomics: MCP* **12,** 1–13 (Jan. 2013).

[77] Pu, S., Vlasblom, J., Turinsky, A., Marcon, E., Phanse, S., Trimble, S. S., Olsen, J., Greenblatt, J., Emili, A. & Wodak, S. J. Extracting

high confidence protein interactions from affinity purification data: At the crossroads. *Journal of proteomics* (Mar. 2015).

[78] Davis, Z. H., Verschueren, E., Jang, G. M., Kleffman, K., Johnson, J. R., Park, J., Von Dollen, J., Maher, M. C., Johnson, T., Newton, W., Jäger, S., Shales, M., Horner, J., Hernandez, R. D., Krogan, N. J. & Glaunsinger, B. A. Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes. *Molecular cell* (Dec. 2014).

[79] Sarraf, S. A., Raman, M., Guarani-Pereira, V., Sowa, M. E., Huttlin, E. L., Gygi, S. P. & Harper, J. W. Landscape of the PARKIN-dependent ubiquitylome in response to mitochondrial depolarization. *Nature* **496,** 372–376 (Apr. 2013).

[80] Bennett, E. J., Rush, J., Gygi, S. P. & Harper, J. W. Dynamics of cullin-RING ubiquitin ligase network revealed by systematic quantitative proteomics. *Cell* **143,** 951–965 (Dec. 2010).

[81] Choi, H., Glatter, T., Gstaiger, M. & Nesvizhskii, A. I. SAINT-MS1: protein-protein interaction scoring using label-free intensity data in affinity purification-mass spectrometry experiments. *Journal of proteome research* **11,** 2619–2624 (Apr. 2012).

[82] Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z. S., Tyers, M., Gingras, A.-C. & Nesvizhskii, A. I.

SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature Methods* **8,** 70–73 (Jan. 2011).

[83] Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., Gingras, A.-C. & Choi, H. SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *Journal of proteomics* **100,** 37–43 (Apr. 2014).

[84] Lambert, J.-P., Tucholska, M., Go, C., Knight, J. D. R. & Gingras, A.-C. Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *Journal of proteomics* (Oct. 2014).

[85] Martell, J. D., Deerinck, T. J., Sancak, Y., Poulos, T. L., Mootha, V. K., Sosinsky, G. E., Ellisman, M. H. & Ting, A. Y. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nature biotechnology* **30,** 1143–1148 (Nov. 2012).

[86] Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of cell biology* **196,** 801–810 (Mar. 2012).

[87] Williams, C. C., Jan, C. H. & Weissman, J. S. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science (New York, N.Y.)* **346,** 748–751 (Nov. 2014).

[88] Heide, H., Bleier, L., Steger, M., Ackermann, J., Dröse, S., Schwamb, B., Zörnig, M., Reichert, A. S., Koch, I., Wittig, I. & Brandt, U. Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex. *Cell metabolism* **16,** 538–549 (Oct. 2012).

[89] Wessels, H. J. C. T., Vogel, R. O., Lightowlers, R. N., Spelbrink, J. N., Rodenburg, R. J., van den Heuvel, L. P., van Gool, A. J., Gloerich, J., Smeitink, J. A. M. & Nijtmans, L. G. Analysis of 953 human proteins from a mitochondrial HEK293 fraction by complexome profiling. *PLoS one* **8,** e68340 (2013).

[90] Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P., Wan, C., Vlasblom, J., Dar, V.-u.-N., Bezginov, A., Clark, G. W., Wu, G. C., Wodak, S. J., Tillier, E. R. M., Paccanaro, A., Marcotte, E. M. & Emili, A. A census of human soluble protein complexes. *Cell* **150,** 1068–1081 (Aug. 2012).

[91] Allfrey, V. G., Faulkner, R & Mirsky, A. E. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences* **51,** 786–794 (May 1964).

[92] Phillips, D. M. The presence of acetyl groups of histones. *The Biochemical journal* **87,** 258–263 (May 1963).

[93] Gnad, F., Forner, F., Zielinska, D. F., Birney, E., Gunawardena, J. & Mann, M. Evolutionary constraints of phosphorylation in eukaryotes, prokaryotes, and mitochondria. *Molecular & cellular proteomics: MCP* **9,** 2642–2653 (Dec. 2010).

[94] Norvell, A. & McMahon, S. B. Cell biology. Rise of the rival. *Science (New York, N.Y.)* **327,** 964–965 (Feb. 2010).

[95] Guan, K.-L. & Xiong, Y. Regulation of intermediary metabolism by protein acetylation. *Trends in biochemical sciences* **36,** 108–116 (Feb. 2011).

[96] Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., Grishin, N. V., White, M., Yang, X.-J. & Zhao, Y. Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Molecular cell* **23,** 607–618 (Aug. 2006).

[97] Lundby, A., Lage, K., Weinert, B. T., Bekker-Jensen, D. B., Secher, A., Skovgaard, T., Kelstrup, C. D., Dmytriyev, A., Choudhary, C., Lundby, C. & Olsen, J. V. Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns. *Cell reports* **2,** 419–431 (Aug. 2012).

[98] Anderson, K. A. & Hirschey, M. D. Mitochondrial protein acetylation regulates metabolism. *Essays in biochemistry* **52,** 23–35 (2012).

[99] Vadvalkar, S. S., Baily, C. N., Matsuzaki, S., West, M., Tesiram, Y. A. & Humphries, K. M. Metabolic inflexibility and protein lysine acetylation in heart mitochondria of a chronic model of type 1 diabetes. *The Biochemical journal* **449,** 253–261 (Jan. 2013).

[100] Lombard, D. B., Tishkoff, D. X. & Bao, J. Mitochondrial sirtuins in the regulation of mitochondrial activity and metabolic adaptation. *Handbook of experimental pharmacology* **206,** 163–188 (2011).

[101] Vaitheesvaran, B., Yang, L., Hartil, K., Glaser, S., Yazulla, S., Bruce, J. E. & Kurland, I. J. Peripheral effects of FAAH deficiency on fuel and energy homeostasis: role of dysregulated lysine acetylation. *PLoS one* **7,** e33717 (2012).

[102] Kim, H.-S., Patel, K., Muldoon-Jacobs, K., Bisht, K. S., Aykin-Burns, N., Pennington, J. D., van der Meer, R., Nguyen, P., Savage, J., Owens, K. M., Vassilopoulos, A., Ozden, O., Park, S.-H., Singh, K. K., Abdulkadir, S. A., Spitz, D. R., Deng, C.-X. & Gius, D. SIRT3 is a mitochondria-localized tumor suppressor required for maintenance of mitochondrial integrity and metabolism during stress. *Cancer cell* **17,** 41–52 (Jan. 2010).

[103] Hirschey, M. D., Shimazu, T., Jing, E., Grueter, C. A., Collins, A. M., Aouizerat, B., Stančáková, A., Goetzman, E., Lam, M. M., Schwer, B., Stevens, R. D., Muehlbauer, M. J., Kakar, S., Bass, N. M., Kuusisto, J., Laakso, M., Alt, F. W., Newgard, C. B., Farese, R. V., Kahn, C. R.

& Verdin, E. SIRT3 deficiency and mitochondrial protein hyperacetylation accelerate the development of the metabolic syndrome. *Molecular cell* **44,** 177–190 (Oct. 2011).

[104] Pereira, C. V., Lebiedzinska, M., Wieckowski, M. R. & Oliveira, P. J. Regulation and protection of mitochondrial physiology by sirtuins. *Mitochondrion* **12,** 66–76 (Jan. 2012).

[105] Rardin, M. J., Newman, J. C., Held, J. M., Cusack, M. P., Sorensen, D. J., Li, B., Schilling, B., Mooney, S. D., Kahn, C. R., Verdin, E. & Gibson, B. W. Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 6601–6606 (Apr. 2013).

[106] Beltrao, P., Albanèse, V., Kenner, L. R., Swaney, D. L., Burlingame, A., Villén, J., Lim, W. A., Fraser, J. S., Frydman, J. & Krogan, N. J. Systematic functional prioritization of protein posttranslational modifications. *Cell* **150,** 413–425 (July 2012).

[107] He, W., Newman, J. C., Wang, M. Z., Ho, L. & Verdin, E. Mitochondrial sirtuins: regulators of protein acylation and metabolism. *Trends in endocrinology and metabolism: TEM* **23,** 467–476 (Sept. 2012).

[108] Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. & Coon, J. J. Gas-phase purification en-

ables accurate, multiplexed proteome quantification with isobaric tagging. *Nature Methods* **8,** 933–935 (Nov. 2011).

[109] De Hoon, M. J. L., Imoto, S, Nolan, J & Miyano, S. Open source clustering software. *Bioinformatics* **20,** 1453–1454 (June 2004).

[110] Saldanha, A. J. Java Treeview–extensible visualization of microarray data. *Bioinformatics* **20,** 3246–3248 (Nov. 2004).

[111] Locasale, J. W. & Cantley, L. C. Metabolic flux and the regulation of mammalian cell growth. *Cell metabolism* **14,** 443–451 (Oct. 2011).

[112] Baez-Ruiz, A. Metabolic adaptations of liver mitochondria during restricted feeding schedules. *AJP: Gastrointestinal and Liver Physiology* **289,** G1015–G1023 (Aug. 2005).

[113] Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., Yao, J., Zhou, L., Zeng, Y., Li, H., Li, Y., Shi, J., An, W., Hancock, S. M., He, F., Qin, L., Chin, J., Yang, P., Chen, X., Lei, Q., Xiong, Y. & Guan, K.-L. Regulation of cellular metabolism by protein lysine acetylation. *Science (New York, N.Y.)* **327,** 1000–1004 (Feb. 2010).

[114] Longo, N., Amat di San Filippo, C. & Pasquali, M. Disorders of carnitine transport and the carnitine cycle. *American journal of medical genetics. Part C, Seminars in medical genetics* **142C,** 77–85 (May 2006).

[115] Adeva, M. M., Calviño, J., Souto, G. & Donapetry, C. Insulin resistance and the metabolism of branched-chain amino acids in humans. *Amino acids* **43,** 171–181 (July 2012).

[116] Rogers, G. W., Brand, M. D., Petrosyan, S., Ashok, D., Elorza, A. A., Ferrick, D. A. & Murphy, A. N. High throughput microplate respiratory measurements using minimal quantities of isolated mitochondria. *PLoS one* **6,** e21746 (2011).

[117] Kendrick, A. A., Choudhury, M., Rahman, S. M., McCurdy, C. E., Friederich, M., Van Hove, J. L. K., Watson, P. A., Birdsey, N., Bao, J., Gius, D., Sack, M. N., Jing, E., Kahn, C. R., Friedman, J. E. & Jonscher, K. R. Fatty liver is associated with reduced SIRT3 activity and mitochondrial protein hyperacetylation. *The Biochemical journal* **433,** 505–514 (Feb. 2011).

[118] Keller, M. P., Choi, Y., Wang, P., Davis, D. B., Rabaglia, M. E., Oler, A. T., Stapleton, D. S., Argmann, C., Schueler, K. L., Edwards, S., Steinberg, H. A., Chaibub Neto, E., Kleinhanz, R., Turner, S., Hellerstein, M. K., Schadt, E. E., Yandell, B. S., Kendziorski, C. & Attie, A. D. A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome Research* **18,** 706–716 (May 2008).

[119] Buchner, D. A., Yazbek, S. N., Solinas, P., Burrage, L. C., Morgan, M. G., Hoppel, C. L. & Nadeau, J. H. Increased mitochondrial oxidative phosphorylation in the liver is associated with obesity and insulin resistance. *Obesity (Silver Spring, Md.)* **19,** 917–924 (May 2011).

[120] Finley, L. W. S., Haas, W., Desquiret-Dumas, V., Wallace, D. C., Procaccio, V., Gygi, S. P. & Haigis, M. C. Succinate dehydrogenase is a direct target of sirtuin 3 deacetylase activity. *PLoS one* **6,** e23295 (2011).

[121] Ahn, B.-H., Kim, H.-S., Song, S., Lee, I. H., Liu, J., Vassilopoulos, A., Deng, C.-X. & Finkel, T. A role for the mitochondrial deacetylase Sirt3 in regulating energy homeostasis. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 14447–14452 (Sept. 2008).

[122] Bao, J., Scott, I., Lu, Z., Pang, L., Dimond, C. C., Gius, D. & Sack, M. N. SIRT3 is regulated by nutrient excess and modulates hepatic susceptibility to lipotoxicity. *Free radical biology & medicine* **49,** 1230–1237 (Oct. 2010).

[123] Cimen, H., Han, M.-J., Yang, Y., Tong, Q., Koc, H. & Koc, E. C. Regulation of succinate dehydrogenase activity by SIRT3 in mammalian mitochondria. *Biochemistry* **49,** 304–311 (Jan. 2010).

[124] Hirschey, M. D., Shimazu, T., Goetzman, E., Jing, E., Schwer, B., Lombard, D. B., Grueter, C. A., Harris, C., Biddinger, S., Ilkayeva, O. R., Stevens, R. D., Li, Y., Saha, A. K., Ruderman, N. B., Bain, J. R., Newgard, C. B., Farese, R. V., Alt, F. W., Kahn, C. R. & Verdin, E. SIRT3 regulates mitochondrial fatty-acid oxidation by reversible enzyme deacetylation. *Nature* **464,** 121–125 (Mar. 2010).

[125] Han, Q., Cai, T., Tagle, D. A., Robinson, H. & Li, J. Substrate specificity and structure of human aminoadipate aminotransferase/kynurenine aminotransferase II. *Bioscience reports* **28,** 205–215 (Aug. 2008).

[126] Rossi, F., Garavaglia, S., Montalbano, V., Walsh, M. A. & Rizzi, M. Crystal structure of human kynurenine aminotransferase II, a drug target for the treatment of schizophrenia. *The Journal of biological chemistry* **283,** 3559–3566 (Feb. 2008).

[127] Han, Q., Robinson, H. & Li, J. Crystal structure of human kynurenine aminotransferase II. *The Journal of biological chemistry* **283,** 3567–3573 (Feb. 2008).

[128] Oxenkrug, G. F. Metabolic syndrome, age-associated neuroendocrine disorders, and dysregulation of tryptophan-kynurenine metabolism. *Annals of the New York Academy of Sciences* **1199,** 1–14 (June 2010).

[129] Noland, R. C., Woodlief, T. L., Whitfield, B. R., Manning, S. M., Evans, J. R., Dudek, R. W., Lust, R. M. & Cortright, R. N. Peroxisomal-mitochondrial oxidation in a rodent model of obesity-associated insulin resistance. *American journal of physiology. Endocrinology and metabolism* **293,** E986–E1001 (Oct. 2007).

[130] Stein, L. R. & Imai, S.-i. The dynamic regulation of NAD metabolism in mitochondria. *Trends in endocrinology and metabolism: TEM* **23,** 420–428 (Sept. 2012).

[131] Feldman, J. L., Dittenhafer-Reed, K. E. & Denu, J. M. Sirtuin catalysis and regulation. *The Journal of biological chemistry* **287,** 42419–42427 (Dec. 2012).

[132] Williamson, D. H., Lund, P & Krebs, H. A. The redox state of free nicotinamide-adenine dinucleotide in the cytoplasm and mitochondria of rat liver. *The Biochemical journal* **103,** 514–527 (May 1967).

[133] McGarry, J. D. & Foster, D. W. Regulation of hepatic fatty acid oxidation and ketone body production. *Annual review of biochemistry* **49,** 395–420 (1980).

[134] Korman, S. H. Inborn errors of isoleucine degradation: a review. *Molecular genetics and metabolism* **89,** 289–299 (Dec. 2006).

[135] Megee, P. C., Morgan, B. A., Mittman, B. A. & Smith, M. M. Genetic analysis of histone H4: essential role of lysines subject to reversible acetylation. *Science (New York, N.Y.)* **247,** 841–845 (Feb. 1990).

[136] Haapalainen, A. M., Meriläinen, G., Pirilä, P. L., Kondo, N., Fukao, T. & Wierenga, R. K. Crystallographic and kinetic studies of human mitochondrial acetoacetyl-CoA thiolase: the importance of potassium and chloride ions for its structure and function. *Biochemistry* **46,** 4305–4321 (Apr. 2007).

[137] Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding Nε-acetyllysine in recombinant proteins. *Nature Chemical Biology* **4,** 232–234 (Feb. 2008).

[138] Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D. & Chin, J. W. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Molecular cell* **36,** 153–163 (Oct. 2009).

[139] Herrera, E & Freinkel, N. Internal standards in the estimation of acetyl-CoA in liver extracts. *Journal of lipid research* **8,** 515–518 (Sept. 1967).

[140] Boekhorst, J., van Breukelen, B., Heck, A. & Snel, B. Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome biology* **9,** R144 (2008).

[141] Benovic, J, Tillman, T, Cudd, A & Fridovich, I. Electrostatic facilitation of the reaction catalyzed by the manganese-containing and the iron-containing superoxide dismutases. *Archives of biochemistry and biophysics* **221,** 329–332 (Mar. 1983).

[142] Tao, R., Coleman, M. C., Pennington, J. D., Ozden, O., Park, S.-H., Jiang, H., Kim, H.-S., Flynn, C. R., Hill, S., Hayes McDonald, W, Olivier, A. K., Spitz, D. R. & Gius, D. Sirt3-mediated deacetylation of evolutionarily conserved lysine 122 regulates MnSOD activity in response to stress. *Molecular cell* **40,** 893–904 (Dec. 2010).

[143] Scott, I., Webster, B. R., Li, J. H. & Sack, M. N. Identification of a molecular component of the mitochondrial acetyltransferase programme: a novel role for GCN5L1. *The Biochemical journal* **443,** 655–661 (May 2012).

[144] Paik, W. K., Pearson, D, Lee, H. W. & Kim, S. Nonenzymatic acetylation of histones with acetyl-CoA. *Biochimica et biophysica acta* **213,** 513–522 (Aug. 1970).

[145] Tanner, K. G., Trievel, R. C., Kuo, M. H., Howard, R. M., Berger, S. L., Allis, C. D., Marmorstein, R & Denu, J. M. Catalytic mechanism and function of invariant glutamic acid 173 from the histone acetyltransferase GCN5 transcriptional coactivator. *The Journal of biological chemistry* **274,** 18157–18160 (June 1999).

[146] Newman, J. C., He, W. & Verdin, E. Mitochondrial protein acylation and intermediary metabolism: regulation by sirtuins and implications for metabolic disease. *Journal of Biological Chemistry* **287,** 42436–42443 (Dec. 2012).

[147] Covian, R. & Balaban, R. S. Cardiac mitochondrial matrix and respiratory complex protein phosphorylation. *American journal of physiology. Heart and circulatory physiology* **303,** H940–66 (Oct. 2012).

[148] Xiong, Y. & Guan, K.-L. Mechanistic insights into the regulation of metabolic enzymes by acetylation. *The Journal of cell biology* **198,** 155–164 (July 2012).

[149] Popov, K. M., Zhao, Y, Shimomura, Y, Kuntz, M. J. & Harris, R. A. Branched-chain alpha-ketoacid dehydrogenase kinase. Molecular cloning, expression, and sequence similarity with histidine protein kinases. *The Journal of biological chemistry* **267,** 13127–13130 (July 1992).

[150] Shimazu, T., Hirschey, M. D., Hua, L., Dittenhafer-Reed, K. E., Schwer, B., Lombard, D. B., Li, Y., Bunkenborg, J., Alt, F. W., Denu, J. M., Jacobson, M. P. & Verdin, E. SIRT3 deacetylates mitochondrial 3-hydroxy-3-methylglutaryl CoA synthase 2 and regulates ketone body production. *Cell metabolism* **12,** 654–661 (Dec. 2010).

[151] Newgard, C. B. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell metabolism* **15,** 606–614 (May 2012).

[152] Zhao, E., Keller, M. P., Rabaglia, M. E., Oler, A. T., Stapleton, D. S., Schueler, K. L., Neto, E. C., Moon, J. Y., Wang, P., Wang, I.-M., Lum, P. Y., Ivanovska, I., Cleary, M., Greenawalt, D., Tsang, J., Choi, Y. J., Kleinhanz, R., Shang, J., Zhou, Y.-P., Howard, A. D., Zhang, B. B., Kendziorski, C., Thornberry, N. A., Yandell, B. S., Schadt, E. E. & Attie, A. D. Obesity and genetics regulate microRNAs in islets, liver, and adipose of diabetic mice. *Mammalian genome : official journal of the International Mammalian Genome Society* **20,** 476–485 (Aug. 2009).

[153] Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* **11,** 1064–1074 (Mar. 2011).

[154] Blommel, P. G., Becker, K. J., Duvnjak, P. & Fox, B. G. Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnology progress* **23,** 585–598 (May 2007).

[155] Yu, W., Dittenhafer-Reed, K. E. & Denu, J. M. SIRT3 protein deacetylates isocitrate dehydrogenase 2 (IDH2) and regulates mitochondrial redox status. *Journal of Biological Chemistry* **287,** 14078–14086 (Apr. 2012).

[156] Hallows, W. C., Lee, S. & Denu, J. M. Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. *Proceedings of the National Academy of Sciences* **103,** 10230–10235 (July 2006).

[157] Middleton, B. The acetoacetyl-coenzyme A thiolases of rat brain and their relative activities during postnatal development. *The Biochemical journal* **132,** 731–737 (Apr. 1973).

[158] Baker, N. A., Sept, D, Joseph, S, Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* **98,** 10037–10041 (Aug. 2001).

[159] Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W. & Bryant, S. H. Open mass spectrometry search algorithm. *Journal of proteome research* **3,** 958–964 (Sept. 2004).

[160] Phanstiel, D. H., Brumbaugh, J., Wenger, C. D., Tian, S., Probasco, M. D., Bailey, D. J., Swaney, D. L., Tervo, M. A., Bolin, J. M., Ruotti, V., Stewart, R., Thomson, J. A. & Coon, J. J. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nature Methods* **8,** 821–827 (2011).

[161] Nesvizhskii, A. I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Molecular & cellular proteomics: MCP* **4,** 1419–1440 (Oct. 2005).

[162] Storey, J. D. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 479–498 (Aug. 2002).

[163] Gohil, V. M., Sheth, S. A., Nilsson, R., Wojtovich, A. P., Lee, J. H., Perocchi, F., Chen, W., Clish, C. B., Ayata, C., Brookes, P. S. & Mootha, V. K. Nutrient-sensitized screening for drugs that shift energy metabolism from mitochondrial respiration to glycolysis. *Nature biotechnology* **28,** 249–255 (Mar. 2010).

[164] Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et biophysica acta* **1666,** 105–117 (Nov. 2004).

[165] Skieterska, K., Duchou, J., Lintermans, B. & Van Craenenbroeck, K. Detection of G protein-coupled receptor (GPCR) dimerization by coimmunoprecipitation. *Methods in cell biology* **117,** 323–340 (2013).

[166] Marbois, B., Gin, P., Faull, K. F., Poon, W. W., Lee, P. T., Strahan, J., Shepherd, J. N. & Clarke, C. F. Coq3 and Coq4 define a polypeptide complex in yeast mitochondria for the biosynthesis of coenzyme Q. *The Journal of biological chemistry* **280,** 20231–20238 (May 2005).

[167] Allan, C. M., Awad, A. M., Johnson, J. S., Shirasaki, D. I., Wang, C., Blaby-Haas, C. E., Merchant, S. S., Loo, J. A. & Clarke, C. F. Identification of Coq11, a new coenzyme Q biosynthetic protein in the CoQ-synthome in Saccharomyces cerevisiae. *Journal of Biological Chemistry,* jbc.M114.633131 (Jan. 2015).

[168] Lohman, D. C., Forouhar, F., Beebe, E. T., Stefely, M. S., Minogue, C. E., Ulbrich, A., Stefely, J. A., Sukumar, S., Luna-Sánchez, M., Jochem, A., Lew, S., Seetharaman, J., Xiao, R., Wang, H., Westphall, M. S., Wrobel, R. L., Everett, J. K., Mitchell, J. C., López, L. C., Coon, J. J., Tong, L. & Pagliarini, D. J. Mitochondrial COQ9 is a lipid-binding protein that associates with COQ7 to enable coenzyme

Q biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **111,** E4697–705 (Nov. 2014).

[169] Ashraf, S., Gee, H. Y., Woerner, S., Xie, L. X., Vega-Warner, V., Lovric, S., Fang, H., Song, X., Cattran, D. C., Avila-Casado, C., Paterson, A. D., Nitschké, P., Bole-Feysot, C., Cochat, P., Esteve-Rudd, J., Haberberger, B., Allen, S. J., Zhou, W., Airik, R., Otto, E. A., Barua, M., Al-Hamed, M. H., Kari, J. A., Evans, J., Bierzynska, A., Saleem, M. A., Bockenhauer, D., Kleta, R., El Desoky, S., Hacihamdioglu, D. O., Gok, F., Washburn, J., Wiggins, R. C., Choi, M., Lifton, R. P., Levy, S., Han, Z., Salviati, L., Prokisch, H., Williams, D. S., Pollak, M., Clarke, C. F., Pei, Y., Antignac, C. & Hildebrandt, F. ADCK4 mutations promote steroid-resistant nephrotic syndrome through CoQ10 biosynthesis disruption. *The Journal of clinical investigation* **123,** 5179–5189 (Dec. 2013).

[170] Nguyen, T. P. T., Casarin, A., Desbats, M. A., Doimo, M., Trevisson, E., Santos-Ocaña, C., Navas, P., Clarke, C. F. & Salviati, L. Molecular characterization of the human COQ5 C-methyltransferase in coenzyme Q10 biosynthesis. *Biochimica et biophysica acta* **1841,** 1628–1638 (Nov. 2014).

[171] Stefely, J. A., Reidenbach, A. G., Ulbrich, A., Oruganty, K., Floyd, B. J., Jochem, A., Saunders, J. M., Johnson, I. E., Minogue, C. E., Wrobel, R. L., Barber, G. E., Lee, D., Li, S., Kannan, N., Coon, J.

J., Bingman, C. A. & Pagliarini, D. J. Mitochondrial ADCK3 Employs an Atypical Protein Kinase-like Fold to Enable Coenzyme Q Biosynthesis. *Molecular cell* (Dec. 2014).

[172] Baba, S. W., Belogrudov, G. I., Lee, J. C., Lee, P. T., Strahan, J., Shepherd, J. N. & Clarke, C. F. Yeast Coq5 C-methyltransferase is required for stability of other polypeptides involved in coenzyme Q biosynthesis. *The Journal of biological chemistry* **279,** 10052–10059 (Mar. 2004).

[173] Steglich, G, Neupert, W & Langer, T. Prohibitins regulate membrane protein degradation by the m-AAA protease in mitochondria. *Molecular and Cellular Biology* **19,** 3435–3442 (May 1999).

[174] Prohibitins and the functional compartmentalization of mitochondrial membranes. *Journal of cell science* **122,** 3823–3830 (Nov. 2009).

[175] Loss of prohibitin induces mitochondrial damages altering β-cell function and survival and is responsible for gradual diabetes development. *Diabetes* **62,** 3488–3499 (Oct. 2013).

[176] Liu, G.-q., Huang, H.-x. & Han, D.-d. A preliminary study on human mitochondrial protein coding gene C2orf47 in cell proliferation and apoptosis. *å¤Žæ—¦å-¦æŠ¥ï¼šè‡ªç„¶ç§‘å-¦ç‰ˆ*, 328–333 (2013).

[177] Prohibitins control cell proliferation and apoptosis by regulating OPA1-dependent cristae morphogenesis in mitochondria. *Genes & development* **22,** 476–488 (Feb. 2008).

[178] Zanna, C., Ghelli, A., Porcelli, A. M., Karbowski, M., Youle, R. J., Schimpf, S., Wissinger, B., Pinti, M., Cossarizza, A., Vidoni, S., Valentino, M. L., Rugolo, M. & Carelli, V. OPA1 mutations associated with dominant optic atrophy impair oxidative phosphorylation and mitochondrial fusion. *Brain : a journal of neurology* **131,** 352–367 (Feb. 2008).

[179] Balsa, E., Marco, R., Perales-Clemente, E., Szklarczyk, R., Calvo, E., Landázuri, M. O. & Enríquez, J. A. NDUFA4 is a subunit of complex IV of the mammalian electron transport chain. *Cell metabolism* **16,** 378–386 (Sept. 2012).

[180] Pitceathly, R. D. S., Rahman, S., Wedatilake, Y., Polke, J. M., Cirak, S., Foley, A. R., Sailer, A., Hurles, M. E., Stalker, J., Hargreaves, I., Woodward, C. E., Sweeney, M. G., Muntoni, F., Houlden, H., Taanman, J.-W., Hanna, M. G. & UK10K Consortium. NDUFA4 mutations underlie dysfunction of a cytochrome c oxidase subunit linked to human neurological disease. *Cell reports* **3,** 1795–1805 (June 2013).

[181] Tello, D., Balsa, E., Acosta-Iborra, B., Fuertes-Yebra, E., Elorza, A., Ordóñez, Á., Corral-Escariz, M., Soro, I., López-Bernardo, E., Perales-Clemente, E., Martínez-Ruiz, A., Enríquez, J. A., Aragonés, J., Cadenas, S. & Landázuri, M. O. Induction of the mitochondrial ND-

UFA4L2 protein by HIF-1α decreases oxygen consumption by inhibiting Complex I activity. *Cell metabolism* **14,** 768–779 (Dec. 2011).

[182] Zhou, J., Wang, H., Lu, A., Hu, G., Luo, A., Ding, F., Zhang, J., Wang, X., Wu, M. & Liu, Z. A novel gene, NMES1, downregulated in human esophageal squamous cell carcinoma. *International Journal of Cancer* **101,** 311–316 (2002).

[183] Feng, Q., Yu, M. & Kiviat, N. B. Molecular biomarkers for cancer detection in blood and bodily fluids. *Critical reviews in clinical laboratory sciences* **43,** 497–560 (2006).

[184] Xu, X., Qiao, M., Zhang, Y., Jiang, Y., Wei, P., Yao, J., Gu, B., Wang, Y., Lu, J., Wang, Z., Tang, Z., Sun, Y., Wu, W. & Shi, Q. Quantitative proteomics study of breast cancer cell lines isolated from a single patient: discovery of TIMM17A as a marker for breast cancer. *Proteomics* **10,** 1374–1390 (Apr. 2010).

[185] Riggs, P. K., Angel, J. M., Abel, E. L. & Digiovanni, J. Differential gene expression in epidermis of mice sensitive and resistant to phorbol ester skin tumor promotion. *Molecular carcinogenesis* **44,** 122–136 (Oct. 2005).

[186] Heinzelmann-Schwarz, V. A., Gardiner-Garden, M, Henshall, S. M., Scurry, J. P., Scolyer, R. A., Smith, A. N., Bali, A, Vanden Bergh, P, Baron-Hay, S, Scott, C, Fink, D, Hacker, N. F., Sutherland, R. L. & O'Brien, P. M. A distinct molecular profile associated with

mucinous epithelial ovarian cancer. *British journal of cancer* **94,** 904–913 (Mar. 2006).

[187] Su, A., Ra, S., Li, X., Zhou, J. & Binder, S. Differentiating cutaneous squamous cell carcinoma and pseudoepitheliomatous hyperplasia by multiplex qRT-PCR. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **26,** 1433–1437 (Nov. 2013).

[188] Guerra, S., López-Fernández, L. A., Pascual-Montano, A., Muñoz, M., Harshman, K. & Esteban, M. Cellular gene expression survey of vaccinia virus infection of human HeLa cells. *Journal of virology* **77,** 6493–6506 (June 2003).

[189] Malykhina, O. A Respiratory Syncytial Virus Replicon That Is Non-Cytotoxic and Capable of Long-Term Foreign Gene Expression (2011).

[190] Suzawa, M. & Ingraham, H. A. The herbicide atrazine activates endocrine gene networks via non-steroidal NR5A nuclear receptors in fish and mammalian cells. *PLoS one* **3,** e2117 (2008).

[191] Lee, Y., Sooranna, S. R., Terzidou, V., Christian, M., Brosens, J., Huhtinen, K., Poutanen, M., Barton, G., Johnson, M. R. & Bennett, P. R. Interactions between inflammatory signals and the progesterone receptor in regulating gene expression in pregnant human uterine

myocytes. *Journal of cellular and molecular medicine* **16,** 2487–2503 (Oct. 2012).

[192] King, E. M., Chivers, J. E., Rider, C. F., Minnich, A., Giembycz, M. A. & Newton, R. Glucocorticoid repression of inflammatory gene expression shows differential responsiveness by transactivation- and transrepression-dependent mechanisms. *PLoS one* **8,** e53936 (2013).

[193] Okabe, H., Makino, S., Kato, K., Matsuoka, K., Seki, H. & Takeda, S. The effect of progesterone on genes involved in preterm labor. *Journal of reproductive immunology* **104-105,** 80–91 (Oct. 2014).

[194] Tummers, B., Goedemans, R., Jha, V., Meyers, C., Melief, C. J. M., van der Burg, S. H. & Boer, J. M. CD40-Mediated Amplification of Local Immunity by Epithelial Cells Is Impaired by HPV. *The Journal of investigative dermatology* (June 2014).

[195] Zimmer, A., Bouley, J., Le Mignon, M., Pliquet, E., Horiot, S., Turfkruyer, M., Baron-Bodo, V., Horak, F., Nony, E., Louise, A., Moussu, H., Mascarell, L. & Moingeon, P. A regulatory dendritic cell signature correlates with the clinical efficacy of allergen-specific sublingual immunotherapy. *The Journal of allergy and clinical immunology* **129,** 1020–1030 (Apr. 2012).

[196] Mitochondrial respiratory supercomplex association limits production of reactive oxygen species from complex I. *Antioxidants & redox signaling* **19,** 1469–1480 (Nov. 2013).

[197] Carilla-Latorre, S., Annesley, S. J., Muñoz-Braceras, S., Fisher, P. R. & Escalante, R. Ndufaf5 deficiency in the Dictyostelium model: new roles in autophagy and development. *Molecular biology of the cell* **24,** 1519–1528 (May 2013).

[198] WebQTL: web-based complex trait analysis. **1,** 299–308 (2003).

[199] BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome biology* **10,** R130 (2009).

[200] Scifo, E, Szwajda, A & Debski, J. Drafting the CLN3 Protein Interactome in SH-SY5Y Human Neuroblastoma Cells: A Label-free Quantitative Proteomics Approach. *Journal of proteome …* (2013).

[201] Curtis, J. M. Adipocyte protein carbonylation and oxidative stress in obesity-linked mitochondrial dysfunction and insulin resistance. (Sept. 2011).

[202] Sánchez, E., Lobo, T., Fox, J. L., Zeviani, M., Winge, D. R. & Fernandez-Vizarra, E. LYRM7/MZM1L is a UQCRFS1 chaperone involved in the last steps of mitochondrial Complex III assembly in human cells. *Biochimica et biophysica acta* **1827,** 285–293 (Mar. 2013).

[203] Atkinson, A., Smith, P., Fox, J. L., Cui, T.-Z., Khalimonchuk, O. & Winge, D. R. The LYR protein Mzm1 functions in the insertion of the Rieske Fe/S protein in yeast mitochondria. *Molecular and Cellular Biology* **31,** 3988–3996 (Oct. 2011).

[204] Lefebvre-Legendre, L, Vaillier, J, Benabdelhak, H, Velours, J, Slonimski, P. P. & di Rago, J. P. Identification of a nuclear gene (FMC1) required for the assembly/stability of yeast mitochondrial F(1)-ATPase in heat stress conditions. *The Journal of biological chemistry* **276,** 6789–6796 (Mar. 2001).

[205] Shan, Y., Napoli, E. & Cortopassi, G. Mitochondrial frataxin interacts with ISD11 of the NFS1/ISCU complex and multiple mitochondrial chaperones. *Human molecular genetics* **16,** 929–941 (Apr. 2007).

[206] Adam, A. C., Bornhövd, C., Prokisch, H., Neupert, W. & Hell, K. The Nfs1 interacting protein Isd11 has an essential role in Fe/S cluster biogenesis in mitochondria. *The EMBO Journal* **25,** 174–183 (Jan. 2006).

[207] Angerer, H. Eukaryotic LYR Proteins Interact with Mitochondrial Protein Complexes. *Biology* **4,** 133–150 (2015).

[208] Toogood, H. S., van Thiel, A., Basran, J., Sutcliffe, M. J., Scrutton, N. S. & Leys, D. Extensive domain motion and electron transfer in the human electron transferring flavoprotein.medium chain Acyl-

CoA dehydrogenase complex. *The Journal of biological chemistry* **279,** 32904–32912 (July 2004).

[209] Hoard-Fruchey, H. M., Goetzman, E., Benson, L., Naylor, S. & Vockley, J. Mammalian electron transferring flavoprotein.flavoprotein dehydrogenase complexes observed by microelectrospray ionization-mass spectrometry and surface plasmon resonance. *The Journal of biological chemistry* **279,** 13786–13791 (Apr. 2004).

[210] Toogood, H. S., Leys, D. & Scrutton, N. S. Dynamics driving function: new insights from electron transferring flavoproteins and partner complexes. *FEBS Journal* **274,** 5481–5504 (Nov. 2007).

[211] Li, Y., Calvo, S., Gutman, R., Liu, J. S. & Mootha, V. K. Expansion of biological pathways based on evolutionary inference. *Cell* **158,** 213–225 (July 2014).

[212] Angerer, H., Radermacher, M., Małkowska, M., Steger, M., Zwicker, K., Heide, H., Wittig, I., Brandt, U. & Zickermann, V. The LYR protein subunit NB4M/NDUFA6 of mitochondrial complex I anchors an acyl carrier protein and is essential for catalytic activity. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 5207–5212 (Apr. 2014).

[213] Maio, N., Singh, A., Uhrigshardt, H., Saxena, N., Tong, W.-H. & Rouault, T. A. Cochaperone Binding to LYR Motifs Confers Speci-

ficity of Iron Sulfur Cluster Delivery. *Cell metabolism* **19,** 445–457 (Mar. 2014).

[214] Seiler, C. Y., Park, J. G., Sharma, A., Hunter, P., Surapaneni, P., Sedillo, C., Field, J., Algar, R., Price, A., Steel, J., Throop, A., Fiacco, M. & LaBaer, J. DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research. *Nucleic Acids Research* **42,** D1253–60 (Jan. 2014).

[215] UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Research* **43,** D204–12 (Jan. 2015).

[216] Sherry, S. T., Ward, M. H., Kholodov, M, Baker, J, Phan, L, Smigielski, E. M. & Sirotkin, K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* **29,** 308–311 (Jan. 2001).

[217] Hanson, G. T., Aggeler, R., Oglesbee, D., Cannon, M., Capaldi, R. A., Tsien, R. Y. & Remington, S. J. Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators. *The Journal of biological chemistry* **279,** 13044–13053 (Mar. 2004).

[218] Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26,** 1367–1372 (Dec. 2008).

[219] Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. & Mann, M. Andromeda: a peptide search engine integrated into the

MaxQuant environment. *Journal of proteome research* **10,** 1794–1805 (Apr. 2011).

[220] Makino, S.-i., Beebe, E. T., Markley, J. L. & Fox, B. G. in *Structural Genomics* 161–178 (Humana Press, Totowa, NJ, Oct. 2013).

[221] Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **43,** D6–17 (Jan. 2015).

[222] Capra, J. A. & Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics* **23,** 1875–1882 (Aug. 2007).

[223] Gavin, A.-C., Maeda, K. & Kühner, S. Recent advances in charting protein-protein interaction: mass spectrometry-based approaches. *Current opinion in biotechnology* **22,** 42–49 (Feb. 2011).

[224] Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P. & Vidal, M. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437,** 1173–1178 (Oct. 2005).

[225] Tavernier, J, Wanker, E. E., Barabási, A. L. & Vidal, M. An empirical framework for binary interactome mapping. *Nature* (2009).

[226] Ideker, T. & Krogan, N. J. Differential network biology. *Molecular systems biology* **8,** 565 (2012).

[227] Sahni, N., Yi, S., Taipale, M., Fuxman Bass, J. I., Coulombe-Huntington, J., Yang, F., Peng, J., Weile, J., Karras, G. I., Wang, Y., Kovács, I. A., Kamburov, A., Krykbaeva, I., Lam, M. H., Tucker, G., Khurana, V., Sharma, A., Liu, Y.-Y., Yachie, N., Zhong, Q., Shen, Y., Palagi, A., San-Miguel, A., Fan, C., Balcha, D., Dricot, A., Jordan, D. M., Walsh, J. M., Shah, A. A., Yang, X., Stoyanova, A. K., Leighton, A., Calderwood, M. A., Jacob, Y., Cusick, M. E., Salehi-Ashtiani, K., Whitesell, L. J., Sunyaev, S., Berger, B., Barabási, A.-L., Charloteaux, B., Hill, D. E., Hao, T., Roth, F. P., Xia, Y., Walhout, A. J. M., Lindquist, S. & Vidal, M. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161,** 647–660 (Apr. 2015).

[228] Mosca, R., Tenorio-Laranga, J., Olivella, R., Alcalde, V., Ceol, A., Soler-López, M. & Aloy, P. dSysMap: exploring the edgetic role of disease mutations. *Nature Methods* **12,** 167–168 (Feb. 2015).

[229] Manenti, G., Galbiati, F., Giannì-Barrera, R., Pettinicchio, A., Acevedo, A. & Dragani, T. A. Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma

susceptibility 1 (Pas1) locus activity in mice. *Oncogene* **23,** 4495–4504 (May 2004).

[230] Galbiati, F., Pettinicchio, A., Dragani, T. A. & Manenti, G. Allelic effects of mouse Pas1 candidate genes in human lung cancer cell lines. *Cancer letters* **244,** 176–181 (Dec. 2006).

[231] Andrea, B., Antonella, G., Jos, J., Wafa, C., Orlando, R., Nancy, S., Marcelo, D. F., Olga, I. e., Giacomo, M. & Tommaso, D. Cis-acting genetic elements at or near the Pas1 locus control Kras2 mutations and gene expression in lung tumors from mice selected for acute inflammatory response. *Frontiers in Immunology* **4** (2013).

[232] Feng, D., Witkowski, A. & Smith, S. Down-regulation of mitochondrial acyl carrier protein in mammalian cells compromises protein lipoylation and respiratory complex I and results in cell death. *The Journal of biological chemistry* **284,** 11436–11445 (Apr. 2009).

[233] Olsson, M. G., Rosenlöf, L. W., Kotarsky, H., Olofsson, T., Leanderson, T., Mörgelin, M., Fellman, V. & Åkerström, B. The radical-binding lipocalin A1M binds to a Complex I subunit and protects mitochondrial structure and function. *Antioxidants & redox signaling* **18,** 2017–2028 (June 2013).

[234] Åkerström, B. & Gram, M. A1M, an extravascular tissue cleaning and housekeeping protein. *Free radical biology & medicine* **74,** 274–282 (Sept. 2014).

[235] AbdelRaheim, S. R. & McLennan, A. G. The Caenorhabditis elegans Y87G2A.14 Nudix hydrolase is a peroxisomal coenzyme A diphosphatase. *BMC biochemistry* **3,** 5 (Mar. 2002).

[236] Reilly, S.-J., Tillander, V., Ofman, R., Alexson, S. E. H. & Hunt, M. C. The nudix hydrolase 7 is an Acyl-CoA diphosphatase involved in regulating peroxisomal coenzyme A homeostasis. *Journal of biochemistry* **144,** 655–663 (Nov. 2008).

[237] Rodrigues, J. V. & Gomes, C. M. Mechanism of superoxide and hydrogen peroxide generation by human electron-transfer flavoprotein and pathological variants. *Free radical biology & medicine* **53,** 12–19 (July 2012).

[238] Andringa, K. K., Andringa, K. K., Udoh, U. S., Udoh, U. S., Landar, A., Landar, A., Bailey, S. M. & Bailey, S. M. Proteomic analysis of 4-hydroxynonenal (4-HNE) modified proteins in liver mitochondria from chronic ethanol-fed rats. *Redox biology* **2C,** 1038–1047 (Oct. 2014).