Improvements to Stable Isotope and Label-Free Quantitative Proteomic Methods

By

Alexander S. Hebert

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

Doctor of Philosophy

(Biochemistry)

At the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 12/16/13

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

Joshua J. Coon, Professor, Biomolecular Chemistry and Chemistry
David J. Pagliarini, Assistant Professor, Biochemistry
Lingjun Li, Professor, Pharmaceutical Sciences and Chemistry
Julie C. Mitchell, Associate Professor, Biochemistry and Mathematics
Michael D. Sheets, Associate Professor, Biomolecular Chemistry

© Copyright by Alexander S. Hebert 2013

All Rights Reserved

IMPROVEMENTS TO STABLE ISOTOPE AND LABEL-FREE QUANTITATIVE PROTEOMIC METHODS

Alexander S. Hebert

Under the supervision of Professor Joshua J. Coon

At the University of Wisconsin-Madison

The following thesis presents strategies for improving throughput and accuracy of mass spectrometry based proteomic methods and the application of these methods to study biological systems. A brief description of qualitative and quantitative proteomics is presented in Chapter 1. In Chapter 2, I report the first description of using mass defects for stable isotope labeling with amino acids in cell culture (SILAC). Using amino acid isotopologues with sub dalton mass differences permits increased multiplexing without concomitant increase in spectral complexity that occurs with SILAC based methods. Chapter 3 introduces quantitative amine reactive mass defect labels. These labels offer quantitation similar to mass defect SILAC but are compatible with tissue sample analysis. Chapter 4 describes a gas phase purification method that improves the accuracy and dynamic range of isobaric labeling by reducing measurement interference derived from precursor co-isolatation. Gas phase purification is accomplished by reacting selected peptide precursors with a proton transfer reagent, followed by re-isolation, fragmentation, and analysis of the charge reduced peptide. This method is then applied to measure changes in the mouse liver mitochondrial acetylome in Chapters 5 and 6. Dramatic changes associated with SIRT3 knockout and calorie restriction are described in Chapter 5. More subtle effects of obesity and fasting are presented in Chapter 6. Finally, a high-throughput workflow to comprehensively sequence a yeast proteome is reported in Chapter 7. This workflow offers a path toward high-throughput label-free proteomics experiments.

Acknowledgements

I would like to thank Professor Coon for the guidance and insight he has given me in my graduate career. He is a machine when it comes to acquiring resources or recruiting talented personnel for his lab and he has coordinated an environment that breeds success. I fully appreciate the access to world class mass spectrometers and the leeway to use them for sometimes low probability of success experiments. His vision has generated a lab where people are typically specialists in informatics, instrumentation, or wet-lab techniques. Being able to specialize has certainly allowed me to achieve much more than if I had to be responsible for all aspects of proteomic research.

The work presented in this thesis would not have been possible without a number of past and present Coon lab members. Specifically, I would like to thank Derek Bailey, Anna Merrill, and Craig Wenger who were responsible for nearly all of the informatics that I used in each of these studies. Thanks to Danielle Swaney and Jason Russell for mentoring me when I joined the lab. Thanks to Alicia Richards, Arne Ulbrich, and Emma Coughlin for help with a number of aspects and studies. Thanks to Chris Rose for being an idea sounding board. Thanks to Mike Westphall for being the group psychologist and keeping everything running smoothly. Also, thanks to Graeme McAlister, Aaron Ledvina, Doug Phanstiel, M. Violet Lee, Justin Brumbaugh, and Paul Grimsrud for all their help and mentorship. I was able to step right into a highly functioning lab because of those that came before me.

I have had the privilege to work with a number of very talented collaborators from the Pagliarini, Denu, and Attie labs. Specifically, Amelia Still, Brendan Floyd, Jon Stefely and Kristin Dittenhafer-Reed provided experimental designs, materials, and/or biochemical perspectives and validation to proteomic data and biochemical validation. I would also like to acknowledge and thank my committee members Professor David Pagliarini, Lingjun Li, Michael D. Sheets and Julie C. Mitchell for their service on my thesis committee.

I initially learned the how to do science at Merck with the help of some excellent supervisors, Mary Retzlaff, John MacNair, and Chitrananda Abeygunawardana (Abey). I would never have attended the University of Wisconsin let alone graduate school if it weren't for Mary Retzlaff or Abey. I had always planned on spending a couple of years in industry and then going to graduate school. That dream had started to die when Abey called me and asked me in a not so friendly tone "what I was doing" and strongly suggested I "was wasting my time if did not go to graduate school." After visiting one school I was convinced that I never wanted to get a PhD. I only visited UW because Mary had written me a letter of recommendation and said that I owed it to her to visit the school and that I would like it. They were both right and I am better off for it.

I would like to thank my undergraduate supervisor Professor Alex Drohat at the University of Albany. I think you have to spend a considerable amount of time not having a clue what is going on before you have any scientific insight. Fortunately for me, I was allowed to perform a lot of different techniques and experiments over the course of my senior year. This experience really provided me with a strong scientific foundation. Additionally, Dr. Drohat set up the interview for me with Abey at Merck, which turned out to be a great experience and I will be forever grateful.

At any point in my career I could have veered towards underachieving (and still could) if it were not for the support of my friends and family. Thanks to my parents and brother for always providing positive reinforcement and sometimes financial reinforcement. Graduate school occasionally makes you cantankerous and Brenda Renaud deserves a medal for putting up with me throughout this process. She handles most of the responsibilities we should share, which has allowed me to be consumed with science for the past 4 years.

Table of Contents

Table of Contents Continued…

Table of Contents Continued…

List of Figures

List of Figures Continued…

List of Tables

List of Abbreviations and acronyms

ACN	Acetonitrile
AGC	Automatic gain control
BCA	Bicinchoninic acid protein assay
BCAA	Branched chain amino acid
CAD	Collision-activated dissociation
CD	Control diet
CI	Chemical ionization
CID	Collision-induced dissociation
CR	Calorie restriction

Da	Dalton, a unit of mass (1 Da = 1 g/mol)
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetracetic acid
ER	Endoplasmic reticulum
ESI	Electrospray ionization
E-value	Expectation value
FAO	Fatty acid oxidation
FDR	False discovery rate
FT	Fourier transform
FTICR	Fourier transform ion cyclotron resonance
FWOM	Full width at one percent maximum peak height
FWTM	Full width at ten percent maximum peak height
GO	Gene ontology
HCD	Higher-energy collisional dissociation
i.d.	Inner diameter
IT	Ion trap
iTRAQ	Isobaric label for relative and absolute quantitation, commercial name from Sciex
LC	Liquid chromatography
Lys-C	Endoproteinase lysine C
MEF	Mouse embryonic fibroblast
MRM	multiple reaction monitoring
MS	Mass spectrometry
MS1	Survey scan of peptide precursors
MS2	tandem mass spectrometry synonymous with MS/MS

mTRAQ	Mass tag for relative and absolute quantitation
m/z	mass-to-charge ratio
nLC	Nanoflow liquid chromatography
nHPLC	Nanoflow high performance liquid chromatography
nUPLC	Nanoflow ultra high performance liquid chromatography
NeuCode	Neutron encoding
o.d.	Outer diameter
OMSSA	Open Mass Spectrometry Search Algorithm
ORF	Open reading frame
OT	Orbitrap
PAF	Post-acquisition filtering
PCA	Principle component analysis
PDB	Protein data bank
ppm	Part per million
PRM	Parallel reaction monitoring
PSM	Peptide spectrum match
PTM	Post-translational modification
PTR	Proton transfer reaction
Q	Quadrupole
qIT	Quadrupole linear ion trap, synonymous with QLT and LTQ
R	Resolving power, typically resolution @ 400 m/z
RP	Reversed phase
RTF	Real time filtering
SILAC	Stable isotope labeling with amino acids in cell culture
SCX	Strong cation exchange

S/N	Signal to noise ratio
TAP	Tandem affinity purification
TCA	Tricarboxylic acid cycle
TEA	Trietyhlamine
Th	Thomsson, synonymous with m/z
TFA	Trifluoroacetic acid
TIC	Total ion chromatogram
TMT	Tandem Mass Tag, commercial name for isobaric labels from Thermo
TOF	Time-of-flight
z	Charge

Chapter 1

Background and introduction

Transformative nucleic acid sequencing technologies mark the advent of modern day genetics in which complete genomic and transcriptomic data can be acquired for thousands of different organisms on a semi-routine basis[1-8]. Although expressed proteins are the genome's primary functional vessels responsible for phenotypic differences, changes in mRNA transcript levels measured using microarray or next-generation sequencing often serve as a proxy for protein expression owing to the power of these methods[9,10]. mRNA and protein expression, however, only have limited correlation (R2 = 0.4 to 0.7) and in many cases, proteomic dynamics trail those of corresponding transcripts [11,12]. Functional protein expression is further complicated by the presence or absence of post translational modifications (PTMs) which can alter activity, half-life, protein-protein interactions, or cellular location[13-16]. Mass spectrometry (MS)-based proteomics offers direct measurements of these functional genetic changes[17].

Over the last twenty years proteomics has matured from identifying ~100 proteins to sequencing comprehensive proteomes in a single experiment[18,19]. For example, reports by the Mann and Aebersold labs have described the sequencing of ~10,000 proteins from a single human cell line[20,21]. Proteomics now offers comparable data quality to nucleic acid sequencing, with similar or higher throughput, and will undoubtedly grant researchers more access to genetic expression information at the protein level[22,23].

Comprehensive global proteomic experiments typically employ a bottom up approach (**Figure 1**). In brief, proteins extracted from lysed cells are digested with a proteolytic enzyme to generate peptides. Peptide mixtures are often fractionated to reduce the sample complexity and increase proteomic coverage. Samples are further simplified for PTM analysis via PTM-specific chromatography or immunoaffinity enrichment steps. Peptides are loaded onto a reversed phase nano liquid chromatography (nLC) column and eluted over a gradient of increasing organic mobile phase with an acidic modifier. As peptides elute from the column and are ionized by electrospray, a survey scan (MS1) measures the *m/z* of each peptide precursor feature. Precursors are sequentially selected, isolated, fragmented, and the resulting fragments are re-analyzed (MS/MS). The fragmentation pattern observed in each MS/MS scan is compared against theoretical spectra from all the possible peptides with theoretical masses within a user-defined tolerance

of the experimentally measured peptide precursor mass (target matches). MS/MS spectra are also scored by a comparison to impossible theoretical peptides (decoy matches). A score cut-off at a selected false discovery rate is calculated by comparison of the target and decoy score distributions[24]. The identified peptides are then mapped back to the fewest proteins that can explain all of the peptide sequences[25].

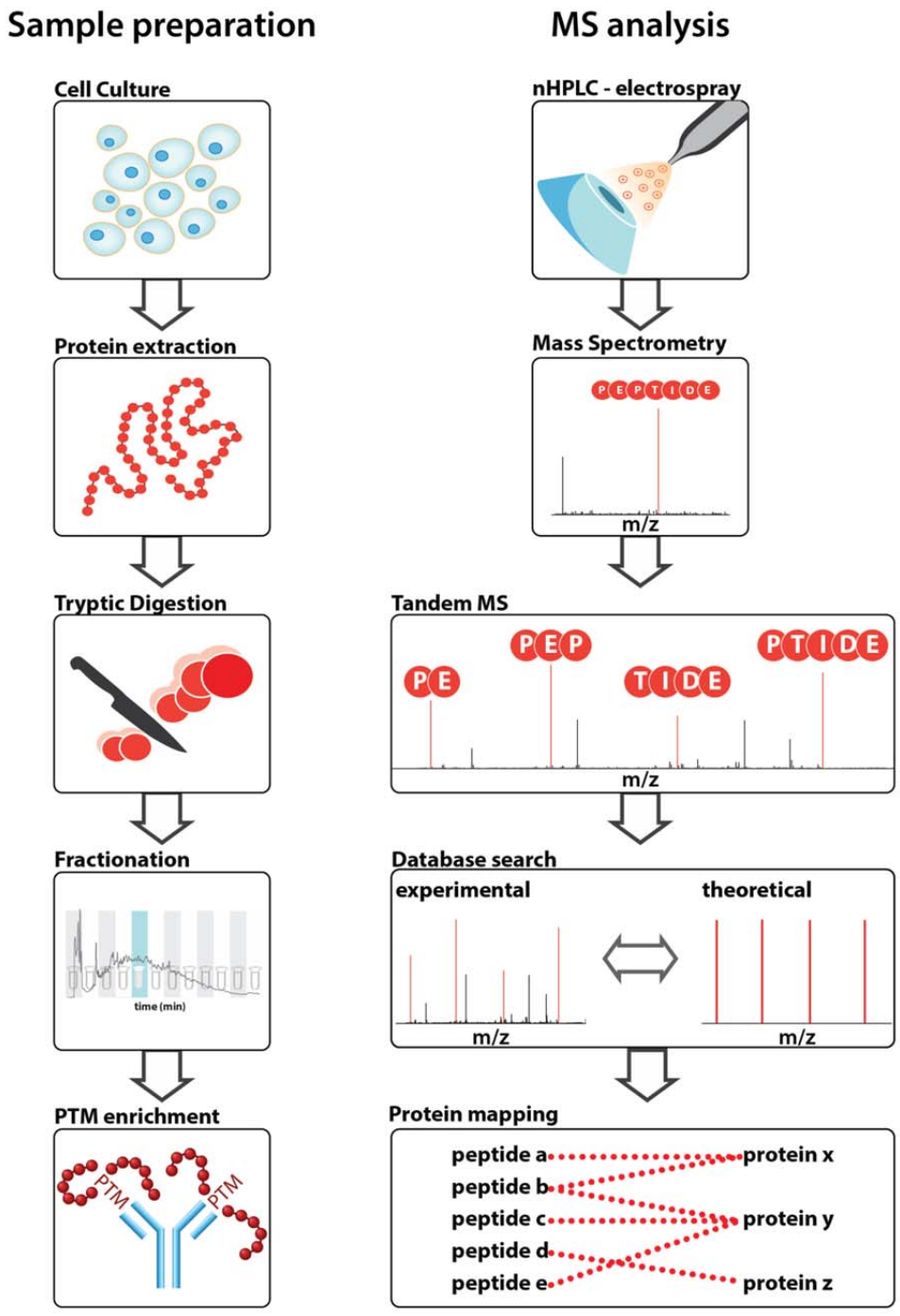
Qualitative MS-based proteomic sequence analysis has already found extraordinary utility in the life science fields, from confirming the identity of simple mixtures of purified proteins to mapping global subcellular protein location or determining the sites of PTMs. Contemporary studies routinely identify thousands of sites of phosphorylation, acetylation, ubiquitylation, methylation, etc. in the proteome[26-29]. New PTMs will likely be discovered and mapped to proteins with similar methods. However, MS based proteomics is not inherently quantitative due to incomplete, semi-random sampling. In data dependent analysis the top N most intense features at a point during the LC elution are sampled for MS/MS analysis. Each MS/MS must then be computationally identified by comparison to theoretical fragments of peptides in an organismal database. Due to the complexity of a sample generated by bottom up proteomic analysis, only a fraction of the peptidic features are fragmented and each MS/MS spectrum may or may not be converted into a peptide identification[30]. Mass spectral identification of a PTM or a protein, therefore, is evidence that that species likely exists in the investigated condition. On the other hand, little can be gleaned from the absence of identifications. Quantitative strategies must be adopted to answer questions regarding PTM functions and proteomic responses to differential conditions.

Obtaining quality quantitative proteomic data is challenging for two reasons. First, the same peptides must be compared between conditions to eliminate large variations in ionization that can occur between different peptide sequences. The stochastic nature of MS/MS sampling, however, leads to inconsistent sequence identification results between analyses. A single LC-MS/MS analysis generally cannot sample all of the observed features generated from a complex bottom-up proteomics sample. Furthermore, the same features may not even be observed because of variability in sample preparation, chromatography, ion current fluctuations, or if the peptide intensity is indistinguishable from noise in one

or more of the conditions. Second, quantitative proteomics experiments are resource-intensive. Typically, a quantitative proteomics comparison of 10-20 samples can take about a month, depending on desired proteomic and PTM coverage, limiting the number of samples that can feasibly be analyzed for a study.

Two quantitative strategies have been developed to circumvent these limitations: stable isotope labeling and label-free quantitation. The former incorporates heavy isotopes into the proteome for one or more of the conditions of interest. The isotopically labeled and unlabeled versions of the peptide analytes are mixed during the sample preparation process. The resulting labeled and unlabeled peptides are generally chemically identical and equally susceptible to variations that occur during sample preparation and the LC-MS/MS analysis. The later approach, label-free quantitation, compares area under the curve for the same peptide or frequency counts of peptide identifications (spectral counting) between runs and employs complex algorithms to mitigate the reproducibility problems discussed above. Each quantitative method bears its own strengths and weaknesses, which are discussed below.

Figure 1. Bottom-up proteomics workflow. Cells from cell culture or tissue samples are lysed to release the protein content. Protein is digested with a proteolytic enzyme, typically LysC or trypsin, to generate peptides which can then be fractionated and enriched for a specific PTM. Peptides are analyzed via nLC coupled to a mass spectrometer through electrospray ionization. Intact peptide ions are detected by mass analysis, isolated, fragmented and re-analyzed to generate tandem MS measurements. Fragment ions are compared with *in silico* fragmentation of theoretical peptides to assign sequence identity to analyte peptides. Proteins are then identified by mapping peptide sequences back to the originating proteins according to the laws of parsimony.


Sample preparation
MS analysis
Cell Culture
Protein extraction
Tryptic Digestion
Fractionation
time (min)
PTM enrichment
PTM
PTM
nHPLC - electrospray
Mass Spectrometry
PEPTIDE
m/z
Tandem MS
PE
PEP
TIDE
PTIDE
m/z
Database search
experimental
theoretical
m/z
m/z
Protein mapping
peptide a
peptide b
peptide c
peptide d
peptide e
protein x
protein y
protein z

Stable Isotope Labeling with Amino acids in Cell culture

One of the first quantitative multiplexing techniques described for proteomics experiments was stable isotope labeling with amino acids in cell culture (SILAC, **Figure 2**)[31]. To compare two samples with this method, one cell culture is grown normally with naturally occurring lysine ("light") while the other culture is grown with an isotopically "heavy" lysine isotopologue in place of the naturally occurring lysine. After an empirically determined number of cell passages or doublings nearly all of the lysine content in the "heavy" cell culture will have been replaced with the lysine isotopologue. The "light" and "heavy" cultures are then mixed together, the protein extracted and digested with a proteolytic enzyme, and the subsequent peptides are analyzed by LC-MS/MS. A lysine-containing peptide will appear as two separate isotopic clusters: "light" and "heavy." The areas under the curve for each of these species are compared to give relative quantities of the peptide between the two samples.

SILAC is the de facto gold standard for quantitative measurements for three reasons. First, the peptides being compared are chemically indistinct; therefore, fluctuations that occur in chromatography and MS affect the peptides identically. Secondly, the ions used for quantitation are specific to the peptide being measured and, with a high resolution mass spectrometer, the interference from other ions is negligible. Thirdly, a peptide is measured across its entire elution profile, providing multiple measurements, which improves quantitative precision.

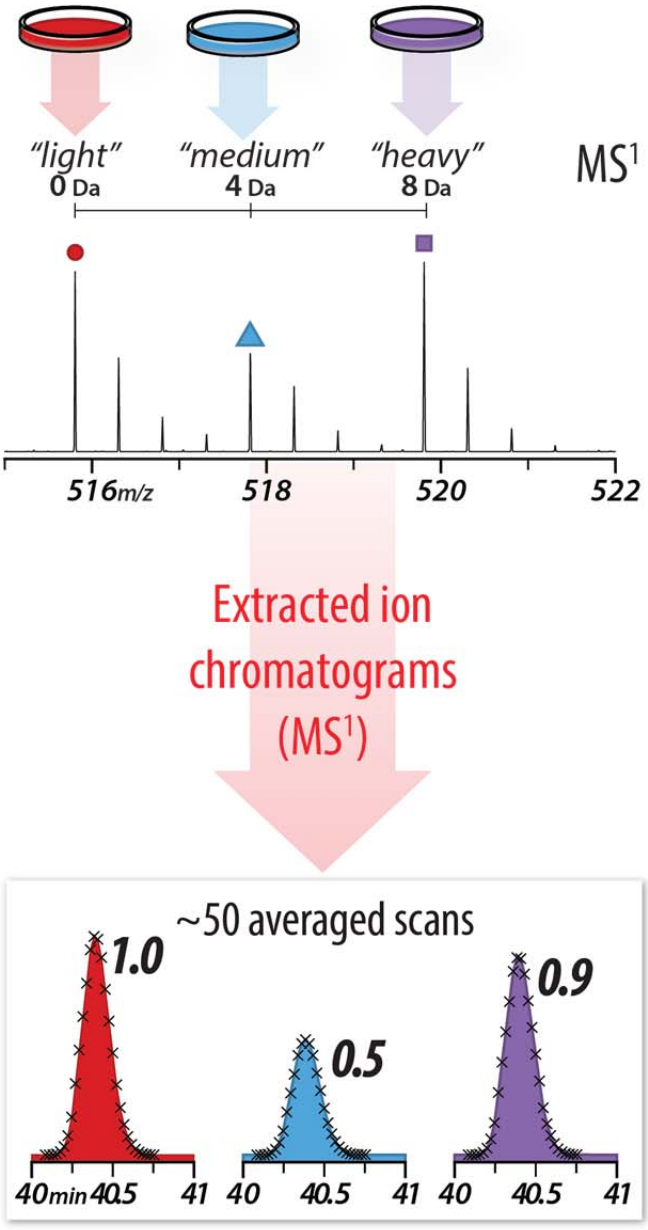
SILAC-type quantitation must compromise between higher multiplexing and increased spectral complexity. Every isotopologue must be spaced by least 4 Da to prevent interference; as a result, each plex adds an additional isotopic cluster for each peptide. Therefore, increasing SILAC multiplexing quickly clutters the MS1 survey scan[32]. The instrument will attempt to sample each cluster even if it produces redundant sequence information, reducing the effective sampling rate. Additionally, the presence of unnecessary features in MS spectra likely leads to more chimeric MS/MS spectra, diminished peptide identification rates, and reduced chances of observing lower abundance ions. These problems are

a direct result of relying on amino acids with differing numbers of total neutrons (e.g., +0 Da, +4 Da, +8 Da).

One promising solution to reduce SILAC spectral complexity is to use amino acids having the same number of total neutrons but located on different elements[33]. NeuCode SILAC (**Chapter 2**) utilizes mass defects to closely space isotopic clusters of peptide isotopologues. Every isotope of every element releases a different amount of energy during nucleus formation. In consequence, every isotope of every element has a distinct mass and the difference in exact mass compared to the number of nuclides is the mass defect. For example, ^{14}C has a mass defect of +3.24 mDa and is 0.17 mDa heavier than ^{14}N. Further, a lysine with $^{2}H_8$ is 36 mDa heavier than a lysine with $^{13}C_6{}^{15}N_2$. Peptides with these mass differences are indistinguishable during peptide precursor sampling selection, when scanning with a medium resolving power scan (resolution = 30,000 at 400 m/z). Thus, oversampling is prevented and, because the two isotopic clusters exist at roughly the same *m/z* space, the chance of chimeric spectra is reduced. The quantitative information is then subsequently revealed during a high resolving power scan (resolution = 240,000 at 400 m/z). We show that this method provides quantitative information equivalent to traditional SILAC but typically produces 30% to 50% more identifications.

Another drawback to SILAC is that incorporation of amino acid isotopologues into tissue proteomes is expensive and time consuming[34]. SILAC-type quantitation can be achieved for these samples by chemically linking the primary amines of the lysines and peptide n-terminii to “light” and “heavy” isotopically labeled moieties, such as methyl groups or pyrazines[35,36]. This strategy, however, is hindered by the same SILAC pitfalls described above. We synthesized novel amine-reactive NeuCode labels to show that chemically labeling with mass defects provides a tissue-compatible strategy for achieving the highest quality data (**Chapter 3**)[37]. We believe that the development of a high-performance, amine-reactive NeuCode label presents an optimal strategy for delivering proteomic quantitation.

Figure 2. 3-plex Lysine SILAC workflow. Cells are cultured with naturally occurring lysine ("light"), a +4 Da lysine isotopologue ("medium"), or a +8 Da lysine isotopologue ("heavy"). Protein from these cells are combined, digested with LysC, and analyzed by nLC-MS/MS. Any peptide bearing a lysine will appear as 3 isotopic clusters in MS1 analysis. Relative abundance measurements are made by comparing the areas under the curve of the peptide elution profile for each cluster.



Isobaric labeling

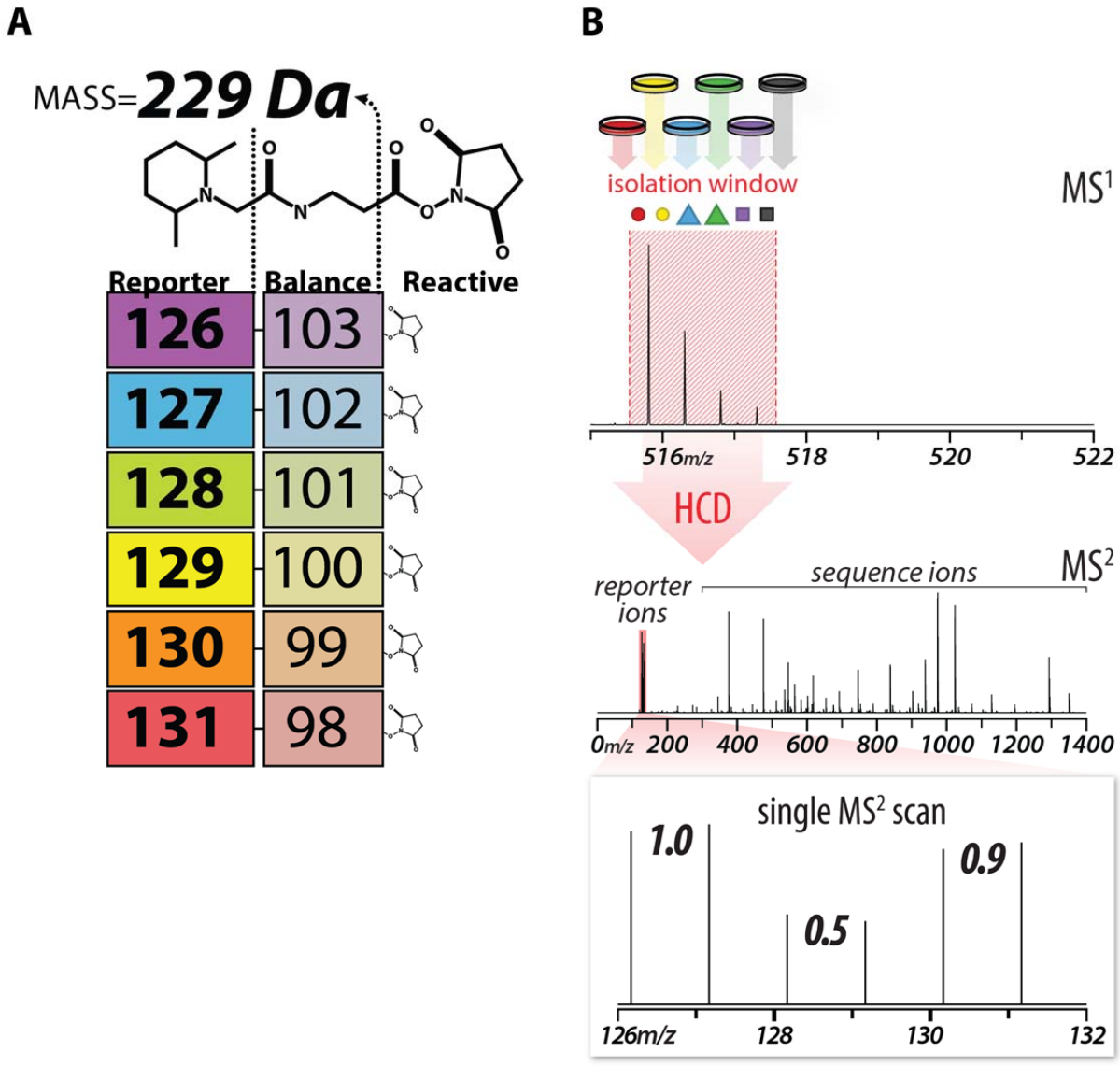
Isobaric labeling methods are capable of the highest degrees of MS multiplexing[38,39]. Currently up to 10 samples can be simultaneously analyzed with commercially available tags. Similar to the NeuCode approach, isobaric labeling relies on the location of additional neutrons. Sets of isobaric labels all have the same total mass yet differ in the masses of the reporter ion and balance regions (**Figure 3**). Because of this elegant design, the intact peptide mass and sequences fragment ion masses will be identical between peptides labeled with different isobaric labels. Only upon fragmentation of peptide precursors are isobaric label-specific ions observed. Reporter ion intensities are compared between the samples for each spectrum and used to make peptide and protein relative abundance measurements.

Multiplexed, isobaric-labeled samples produce MS spectra that are essentially no more complex than that of non-multiplexed samples. The reliance of isobaric labeling on reporter ions, however, and not on sequence specific-ions makes this technique exceptionally susceptible to interference from co-isolated peptides[40]. Co-isolated interfering peptides fragment and produce reporter ions identical to the reporter ions from the peptide selected for MS/MS sampling. During the analysis of two similar conditions (e.g., a knockout experiment), most peptides are not changing in abundance between the samples and interference causes a suppression of dynamic range. The result is a high rate of false negatives. On the other hand, during the analysis of two disparate conditions (e.g., comparing proteomes across different tissues), interference could actually cause two peptides with identical expression to appear different, producing false positives. The interference problem is further exacerbated by the differential reporter ion intensities of peptides with different charge states. Thus, an interfering peptide observed in the MS1 may have low intensity compared to the selected peptide, but may produce a substantial portion of the reporter ion intensity. This aspect prevents accurate modeling of interference based on MS1 spectra.

Chapter 4 describes a novel gas phase purification method (QuantMode) that decreases the impact of interference in isobaric labeling experiments[41]. With this method, peptide precursors selected for MS/MS sampling are isolated and reacted with a proton transfer reagent. The charge reduced

precursor is then isolated, fragmented and analyzed. Peptides with different initial charge states will be observed at disparate *m/z* values upon charge state reduction. Additionally, the reduced charge state separates peptides in *m/z* space, further purifying the isolated precursor from all peptides with the same initial charge state. This method was applied to characterize differences in the mouse liver mitochondrial acetyl proteome of SIRT3 knockout with and without calorie restriction (**Chapter 5**), as well as alterations in fasted/re-fed and obese/lean mouse models (**Chapter 6**)[29,42]. The improved dynamic range afforded by the QuantMode method enabled the measurement of both drastic -- up to ~ 100 fold differences in acetylation in the SIRT3 experiments and subtle, but reproducible, differences in the fasted/re-fed and obese/lean comparisons.

Figure 3. Isobaric label structure and workflow. (A) The 6-plex tandem mass tag structure is displayed as an example of isobaric label design. As the reporter ion mass increases the balance region mass decreases, such that the mass sum of these two sections is constant. (B) Peptides from 6 samples are each independently conjugated to an isobaric label, mixed, and analyzed. The labels are indistinguishable during intact peptide mass analysis but tandem MS analysis generates reporter ions. Relative abundance measurements are performed by comparing reporter ion intensities.

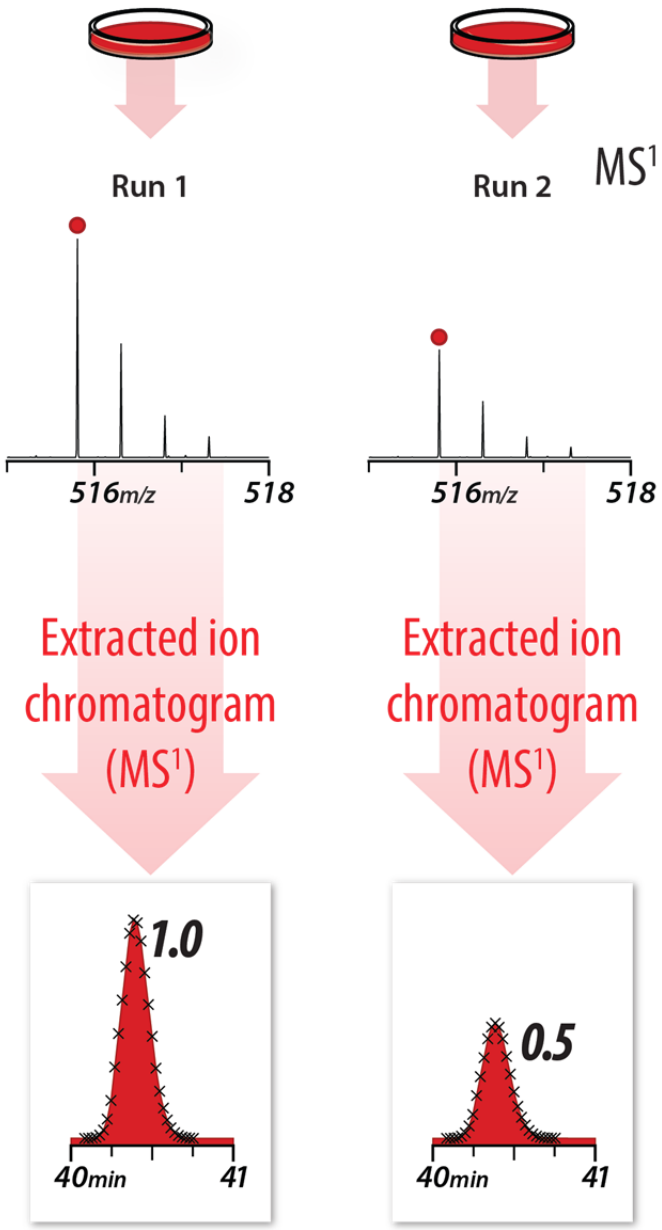


Label-free quantitation

Label-free quantitation is typically performed by comparing the area under the curve (**Figure 4**) or frequency counts (spectral counting) for each peptide or protein, respectively, between runs and conditions[11,34,43]. These approaches have very straightforward sample preparation methods and produce the most coverage per unit of analysis time. However, label-free methods are inherently incapable of multiplexing sample analysis, demanding considerable amounts of analyst and instrument time. Analyst and instrument time, thus, scale with the number of samples analyzed and can become overly cumbersome if fractionation is required. For example, an analysis of 3 replicates of 2 conditions, separated into ~20 fractions, would be prepared and analyzed in ~4 days (2 days of sample preparation + 2 days instrument time) with isobaric labeling, but would require ~16 days (4 days sample preparation + 12 days instrument time) with label-free quantitation.

An obvious strategy to limit analysis time is to maximize the sequence coverage of non-fractionated samples. In 2012 Mann and colleagues published a paper reporting the acquisition of comprehensive sequence coverage of the yeast proteome in a 4-hr LC-MS/MS analysis[23]. This strategy presented the possibility of eliminating fractionation from the workflow and substantially increasing throughput for label-free proteomic experiments. We further optimized lysis, digestion, and chromatographic conditions and used a novel quadrupole-orbitrap-linear ion trap instrument to obtain comparable sequence coverage in a 70 min LC-MS/MS analysis (**Chapter 7**)[22]. This 3.5X improvement in throughput opens up the possibility of analyzing very large sample sets, such as knockout libraries or quantitative trait loci mapping, that once would have been thought too cumbersome to attempt.

Figure 4. Area under the curve label-free quantitation workflow. Peptides from cells to be compared are analyzed in separate runs. The extracted ion chromatograms of a specific peptide are compared between runs to measure relative peptide abundances.



An optimal quantitative proteomic strategy

There are certain attributes that are globally desirable for any quantitative proteomics technique. First, the method should be compatible with samples from any source. The SILAC methods listed above are technically compatible with tissue analysis but either requires a long and expensive labeling time to fully incorporate the isotopically heavy amino acid or comparison of an unlabeled tissue to that of a labeled reference tissue, effectively eliminating the ability to multiplex. Amine-reactive labeling strategies are typically tissue-compatible, but any label used should have minimal effects on chromatography, ionization, and fragmentation. Secondly, the method should be high throughput to allow the analysis of more replicates and/or conditions that enables more impactful studies. Increased throughput can be generated by either increasing multiplexing or reducing the time required to analyze each sample. Thirdly, the method should measure sequence specific ions or involve and approach to marginalize the effects of measurement interference. This thesis presents strategies to advance SILAC, isobaric labeling, and label-free quantitation towards these common goals.

References

1 (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:http://www.nature.com/nature/journal/v409/n6822/suppinfo/409860a0_S1.html

2 Augenlicht, L. H., Wahrman, M. Z., Halsey, H., Anderson, L., Taylor, J. & Lipkin, M. (1987) Expression of Cloned Sequences in Biopsies of Human Colonic Tissue and in Colonic Carcinoma Cells Induced to Differentiate in Vitro. *Cancer Research* **47**, 6017-6021

3 Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G. *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59, doi:http://www.nature.com/nature/journal/v456/n7218/suppinfo/nature07517_S1.html

4 Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science* **270**, 467-470, doi:10.1126/science.270.5235.467

5 Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. *et al.* (2001) The Sequence of the Human Genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040

6 Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A. *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872-876, doi:http://www.nature.com/nature/journal/v452/n7189/suppinfo/nature06884_S1.html

7 Alkan, C., Coe, B. P. & Eichler, E. E. (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* **12**, 363-376

8 Metzker, M. L. (2010) Sequencing technologies [mdash] the next generation. *Nat Rev Genet* **11**, 31-46

9 Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S. *et al.* (2010) Integrative Genomic Profiling of Human Prostate Cancer. *Cancer Cell* **18**, 11-22, doi:http://dx.doi.org/10.1016/j.ccr.2010.05.026

10 Lanford, R. E., Hildebrandt-Eriksen, E. S., Petri, A., Persson, R., Lindow, M., Munk, M. E. *et al.* (2010) Therapeutic Silencing of MicroRNA-122 in Primates with Chronic Hepatitis C Virus Infection. *Science* **327**, 198-201, doi:10.1126/science.1178178

11 Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J. *et al.* (2011) Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:http://www.nature.com/nature/journal/v473/n7347/abs/10.1038-nature10098-unlocked.html#supplementary-information

12 Gry, M., Rimini, R., Strömberg, S., Asplund, A., Pontén, F., Uhlén, M. *et al.* (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics* **10**, 1-14, doi:10.1186/1471-2164-10-365

13 Proost, P., Struyf, S., Couvreur, M., Lenaerts, J.-P., Conings, R., Menten, P. *et al.* (1998) Posttranslational Modifications Affect the Activity of the Human Monocyte Chemotactic Proteins MCP-1 and MCP-2: Identification of MCP-2(6–76) as a Natural Chemokine Inhibitor. *The Journal of Immunology* **160**, 4034-4041

14 Ciechanover, A. (1998) The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO J* **17**, 7151-7160

15 van Dieck, J., Teufel, D. P., Jaulent, A. M., Fernandez-Fernandez, M. R., Rutherford, T. J., Wyslouch-Cieszynska, A. *et al.* (2009) Posttranslational Modifications Affect the Interaction of S100 Proteins with Tumor Suppressor p53. *Journal of Molecular Biology* **394**, 922-930, doi:http://dx.doi.org/10.1016/j.jmb.2009.10.002

16 Huang, G., Zhang, M. & Erdman, S. E. (2003) Posttranslational Modifications Required for Cell Surface Localization and Function of the Fungal Adhesin Aga1p. *Eukaryotic Cell* **2**, 1099-1114, doi:10.1128/ec.2.5.1099-1114.2003

17 Laurent, J. M., Vogel, C., Kwon, T., Craig, S. A., Boutz, D. R., Huse, H. K. *et al.* (2010) Protein abundances are more conserved than mRNA abundances across diverse taxa. *PROTEOMICS* **10**, 4209-4212, doi:10.1002/pmic.201000327

18 Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O. *et al.* (1996) Linking genome and proteome by mass spectrometry: Large-scale identification of yeast

proteins from two dimensional gels. *Proceedings of the National Academy of Sciences* **93**, 14440-14445

19 Figeys, D., Ducret, A., Yates, J. R. & Aebersold, R. (1996) Protein identification by solid phase microextraction[mdash]capillary zone electrophoresis[mdash]microelectrospray[mdash]tandem mass spectrometry. *Nat Biotech* **14**, 1579-1583

20 Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A. *et al.* (2011) The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, doi:http://www.nature.com/msb/journal/v7/n1/suppinfo/msb201182_S1.html

21 Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J. *et al.* (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**, doi:http://www.nature.com/msb/journal/v7/n1/suppinfo/msb201181_S1.html

22 Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S. *et al.* (2013) The One Hour Yeast Proteome. *Molecular & Cellular Proteomics*, doi:10.1074/mcp.M113.034769

23 Nagaraj, N., Alexander Kulak, N., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O. *et al.* (2012) System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Molecular & Cellular Proteomics* **11**, doi:10.1074/mcp.M111.013722

24 Elias, J. E. & Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207-214, doi:10.1038/nmeth1019

25 Nesvizhskii, A. I. & Aebersold, R. (2005) Interpretation of shotgun proteomic data - The protein inference problem. *Molecular & Cellular Proteomics* **4**, 1419-1440, doi:10.1074/mcp.R500012-MCP200

26 Bremang, M., Cuomo, A., Agresta, A. M., Stugiewicz, M., Spadotto, V. & Bonaldi, T. (2013) Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Molecular BioSystems* **9**, 2231-2247, doi:10.1039/c3mb00009e

27 Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J. *et al.* (2010) Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. *Sci. Signal.* **3**, ra3-, doi:10.1126/scisignal.2000475

28 Kim, W., Bennett, Eric J., Huttlin, Edward L., Guo, A., Li, J., Possemato, A. *et al.* (2011) Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Molecular Cell* **44**, 325-340, doi:http://dx.doi.org/10.1016/j.molcel.2011.08.025

29 Hebert, Alexander S., Dittenhafer-Reed, Kristin E., Yu, W., Bailey, Derek J., Selen, Ebru S., Boersma, Melissa D. *et al.* (2013) Calorie Restriction and SIRT3 Trigger Global Reprogramming of the Mitochondrial Protein Acetylome. *Molecular Cell* **49**, 186-199, doi:http://dx.doi.org/10.1016/j.molcel.2012.10.024

30 Michalski, A., Cox, J. & Mann, M. (2011) More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC−MS/MS. *J. Proteome Res.* **10**, 1785-1793, doi:10.1021/pr101060v

31 Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **1**, 376-386, doi:10.1074/mcp.M200025-MCP200

32 Molina, H., Yang, Y., Ruch, T., Kim, J. W., Mortensen, P., Otto, T. *et al.* (2009) Temporal Profiling of the Adipocyte Proteome during Differentiation Using a Five-Plex SILAC Based Strategy. *J. Proteome Res.* **8**, 48-58, doi:10.1021/pr800650r

33 Hebert, A. S., Merrill, A. E., Bailey, D. J., Still, A. J., Westphall, M. S., Strieter, E. R. *et al.* (2013) Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat. Methods* **10**, 332-+, doi:10.1038/nmeth.2378

34 Krüger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C. A., Forner, F. *et al.* (2008) SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell* **134**, 353-364, doi:http://dx.doi.org/10.1016/j.cell.2008.05.033

35 Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protocols* **4**, 484-494

36 Mertins, P., Udeshi, N. D., Clauser, K. R., Mani, D., Patel, J., Ong, S.-e. *et al.* (2012) iTRAQ Labeling is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics. *Molecular & Cellular Proteomics* **11**, doi:10.1074/mcp.M111.014423

37 Hebert, A. S., Merrill, A. E., Stefely, J. A., Bailey, D. J., Wenger, C. D., Westphall, M. S. *et al.* (2013) Amine-reactive Neutron-encoded Labels for Highly Plexed Proteomic Quantitation. *Molecular & Cellular Proteomics* **12**, 3360-3369

38 McAlister, G. C., Huttlin, E. L., Haas, W., Ting, L., Jedrychowski, M. P., Rogers, J. C. *et al.* (2012) Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Analytical Chemistry* **84**, 7469-7478, doi:10.1021/ac301572t

39 Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G. *et al.* (2003) Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry* **75**, 1895-1904, doi:10.1021/ac0262560

40 Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I. & Wright, P. C. (2009) iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly". *J. Proteome Res.* **8**, 5347-5355, doi:10.1021/pr900634c

41 Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. *et al.* (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat. Methods* **8**, 933-935, doi:10.1038/nmeth.1716

42 Still, A. J., Floyd, B. J., Hebert, A. S., Bingman, C. A., Carson, J. J., Gunderson, D. R. *et al.* (2013) Quantification of Mitochondrial Acetylation Dynamics Highlights Prominent Sites of

Metabolic Regulation. *Journal of Biological Chemistry* **288**, 26209-26219, doi:10.1074/jbc.M113.483396

43 Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G. *et al.* (2011) Less label, more free: Approaches in label-free quantitative mass spectrometry. *PROTEOMICS* **11**, 535-553, doi:10.1002/pmic.201000553

Chapter 2

Neutron-encoded mass signatures for multi-plexed proteome quantification

ASH designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Hebert, A. S.*, Merrill, A. E.*, Bailey, D. J., Still, A. J., Westphall, M. S., Strieter, E. R., Pagliarini, D. J. & Coon, J. J. (2013) Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Meth* **10**, 332-334, doi:http://www.nature.com/nmeth/journal/v10/n4/abs/nmeth.2378.html

*co-first author

Abstract

We describe a protein quantification method that exploits the subtle mass differences caused by neutron-binding energy variation in common stable isotopes. These mass differences are synthetically encoded into amino acids and incorporated into proteins with SILAC. Analysis with high mass resolving power (>100,000) reveals the isotopologue-embedded peptide signals permitting quantification. We conclude neutron encoding will enable high levels of multi-plexing (> 10), while maintaining the dynamic range and quantitative accuracy of traditional SILAC.

Introduction

Stable isotope incorporation with mass spectrometry (MS) analysis is a central technology for protein abundance measurements.[1,2] However the heavy isotopes are incorporated, metabolically[3,4] or chemically,[5-8] the aim is to differentially tag samples; the preferred spacing of ~ 4 Da limits the isotopic cluster overlap of the heavy and light peptides. Unfortunately, this multi-Da spacing restricts SILAC's quantitative capacity to triplex comparisons for two main reasons: (1) amino acid structures restrict the number of isotopes that can be added; and (2) mass spectral complexity increases as multiple isotopic clusters are introduced. Isobaric tagging provides up to 8-plexed analysis by concealing the quantitative information in the MS^1 scan and releasing it only upon MS/MS.[9-11] It does, however, suffer from severe dynamic range compression and loss of quantitative accuracy due to precursor interference.[12,13] And, quantitative data can only be obtained for peptides that are selected for MS/MS—a serious problem during replicate analysis, as there is high run-to-run variability in identifications (~60%).[14]

Through a fortuitous discovery, the multi-plexing capacity of TMT isobaric tags was recently expanded from six to eight: the concomitant swapping of a ^{12}C for a ^{13}C atom and a ^{15}N for a ^{14}N atom produces a new tag with a 6 mDa mass difference.[10,11] This mass change results from the discrepancy in energetics of neutron binding between isotopes of N and C, and can be distinguished with a mass resolution of 50,000 at *m/z* 130.[15] This creative concept still relies upon MS^2-based quantification, however, and does not resolve the accuracy and reproducibility issues of isobaric tagging. We reasoned that other elements, besides C and N, could encode neutron mass signatures. Indeed, mass defects can be induced with many common elements and their isotopes, for example: $^{12}C/^{13}C$ (+3.3 mDa), $^{1}H/^{2}H$ (+6.3 mDa), $^{16}O/^{18}O$ (+4.2 mDa), $^{14}N/^{15}N$ (-3.0 mDa), and $^{32}S/^{34}S$ (-4.2 mDa), among others.[15] We hypothesized that calculated incorporation of these isotopes into proteomes would generate a new MS^1-centric quantification technology that combines the accuracy of SILAC, with the multi-plexing capacity of isobaric tagging. We call this method neutron encoding (NeuCode).

Results

A straightforward method of embedding neutron signatures is to use amino acids during protein synthesis (NeuCode SILAC); consider a synthetic lysine molecule that is 8 Da heavier than unlabeled Lys (+8). **Figure 1a** describes the 39 isotopologues of this +8 Da amino acid, which span a mass range of 38.5 mDa, and are separated by approximately 1 mDa each. With infinite mass resolution, these 39 Lys isotopologues would permit 39-plex NeuCode SILAC; not having access to infinite mass resolution, we calculated the minimum resolvable mass difference using current technology. With a library of 71,499 identified tandem mass spectra, we modeled the percentage of the peptidome that would be quantifiable (*i.e.*, separated at full width at 1% max, only 1% overlap in peak areas) when labeled at intervals of 12, 18, or 36 mDa (**Fig. 1b**), at resolutions ranging from 10^3 to 10^6. Our calculation takes into account the diversity of precursor *m*, *z*, and *m/z* that is typically observed in a shotgun experiment. These data demonstrate that, at a resolving power of 480,000 > 85% of identified peptides can be quantified (i.e., resolved) when spaced 18 mDa apart. At 960,000 resolving power, achievable on both FT-ICR and Orbitrap MS systems, > 90% coverage with 12 mDa spacing could be achieved.[16-18]

NeuCode SILAC was tested by growing yeast on normal "*light*" lysine (+0 Da) and on two +8 Da heavy lysine isotopologues: one with six ^{13}C atoms and two ^{15}N atoms ("*heavy 1*", +8.0142 Da), and the other with eight ^{2}H ("*heavy 2*", +8.0502 Da). Peptides containing these Lys isotopologues differ in mass by 36 mDa (per lysine), and are easily distinguished at resolving powers in excess of 200,000. To our knowledge, of the 39 possible +8 Da Lys isotopologues, these are the only two that have been synthesized. Traditional SILAC samples were prepared by combining the "*light*" and "*heavy 1*" labeled peptides in ratios of 1:1 and 1:5. NeuCode SILAC samples were similarly prepared, except with "*heavy 1*" and "*heavy 2*" labeled peptides. Samples from each method were loaded onto a capillary nLC column and gradient-eluted into an ion trap-Orbitrap hybrid MS. For traditional SILAC, MS^1 analyses were performed at a resolving power of 30,000, with the top 10 most intense precursors selected for MS^2 analysis. For NeuCode SILAC, we implemented an additional MS^1 scan at a resolving power of 480,000.

The high resolution spectrum distinguished the NeuCode SILAC pairs, decoding the embedded quantitative data. Panel (c) of **Figure 1** displays a MS^1 scan (R = 30,000), and panel (d) presents the isotopic cluster of a selected precursor at *m/z* 827. Here we plot the signal generated with either the typical 30,000 resolving power (black), or the high resolution (480,000) quantification scan (red). The very close *m/z* spacing of the NeuCode SILAC partners is ideal for MS^2 scanning since both isotopologues are co-isolated, fragmented, and mass analyzed together to produce MS/MS spectra that are identical to non-multiplexed samples under normal resolution settings (**Fig. 1E**). Simply put, the encoded signatures are concealed, and spectral matching is unaffected. The high resolution scan does take ~1.6 seconds to complete, but, the system performs ion trap MS^2 analyses (top 10) during that time, so that very little effect on overhead is induced (16,974 *vs.* 18,074 MS^2 spectra, NeuCode SILAC *vs.* traditional SILAC).[19] The NeuCode SILAC experiment produced considerably more unique peptide spectral matches (PSMs) than traditional SILAC: 3,078 vs. 2,401, respectively. In traditional SILAC, each peptide precursor appears at two distinct *m/z* values, causing a redundancy in peptide identifications; the result is limited sampling depth. NeuCode SILAC eliminates this problem because all quantitative information is encoded within a single *m/z* peak for each precursor (**Fig. 1d**), meaning redundant MS^2 scans on partner peaks are not acquired.

Figure 2a captures quantitative metrics for both traditional and NeuCode SILAC: the dashed horizontal lines indicate the true ratio (red = 1:1, blue = 5:1), and boxplots demarcate the median (stripe), the 25th to 75th percentile (interquartile range, box), 1.5 times the interquartile range (whiskers), and outliers (open circles). Of the 3,078 PSMs posted by NeuCode SILAC, 87% were quantifiable (2,693). For traditional SILAC, 2,127 PSMs (89%) produced quantitative data. We conclude that NeuCode SILAC permits increased sampling depth, while maintaining highly comparable quantitative accuracy and precision. Identifications from the NeuCode SILAC data were generated using the MS^1 scans collected under low resolution settings (30,000, **Fig. 1d**, black trace). We plotted the distribution of mass error (ppm) as a function of identification e-value (~ significance) for both NeuCode SILAC and traditional

SILAC for all identifications (1% FDR). Here we note a very subtle decrease in mass accuracy for NeuCode SILAC – 3.5 *vs.* 2.5 ppm, but with comparable precision. This subtle increase in mass error stems from use of the low resolution (30K) MS^1 scan for NeuCode, where the isotopologues are not resolved; however, it is not problematic, as database searching typically allows precursor mass error tolerances of ± 10 to ± 25 ppm. Using the mass values from the high resolution MS^1 scan, where the isotopologues are resolved, could completely eliminate this difference.

Next, we benchmarked the NeuCode SILAC method against traditional SILAC by analyzing mouse C2C12 myoblasts and their differentiation to myotubes, a system that has been well characterized in the past by other MS methods.[3,20] NeuCode SILAC generated 5,747 quantifiable peptides, vs. 3,400 for traditional SILAC, translating to 45% more quantified proteins (1,458 vs. 1,031). **Figure 2b** displays the excellent correlation of measured protein abundance between NeuCode and traditional SILAC. Both methods detect expected proteome alterations induced by myogenic differentiation.

Here we report a fresh approach for protein quantification using stable isotopes. NeuCode exploits the subtle differences in neutron binding energy between isotopes.[15] The approach effectively compresses isotopic information into a very narrow *m/z* space (~ 5-40 mTh) so that it is easily concealed or revealed by varying mass resolution. FT-MS systems offer ultra-high resolution (> 1,000,000) and will permit the use of NeuCode labeled peptides separated by as little as ~10 mDa.[16-18] We envision synthesis of custom lysine isotopologues that offer four-plex quantification: +8 Da at 0, 12, 24, and 36 mDa spacings. Furthermore, such quad-plexed isotopologues could be generated with four, eight, or twelve additional neutrons. By combining these twelve isotopologues, NeuCode promises to facilitate 12-plex SILAC. Here, each peptide would be present in three isotopic clusters, just as in traditional three-plex SILAC; however, each cluster will reveal four distinct peaks upon high resolution analysis. We demonstrate this concept by combining duplex NeuCode SILAC with mTRAQ labeling to achieve six-plexed MS^1 quantification (**Fig. 2c**). Note that by using current commercial FT-MS mass resolution capability, NeuCode SILAC stands to easily deliver 9-plex quantification by easing Lys spacing to 18

mDa. NeuCode does not suffer from the pervasive problem of precursor interference that cripples quantitative accuracy in isobaric tagging, and, as mass resolution continues to improve, so will the multiplexing capacity of NeuCode. Finally, it has not escaped our notice that these neutron mass signatures could be encoded in chemical tags for samples not amenable to metabolic labeling.

Figure 1: NeuCode feasibility and scan sequence. (**a**) Mass calculations of the 39 isotopologues for a +8-Da lysine. Shown in solid black are the isotopologues used for the experiments presented here. The table below displays the elemental isotope compositions of each isotopologue. (**b**) Theoretical calculations depicting the percentage of peptides that are resolved (full width at 1% maximum peak height) when spaced 12, 18 or 36 mDa apart for resolving powers (*R*) of 15,000–1,000,000. (**c**) Top, MS1 scan collected with typical 30,000 resolving power from a nano–liquid chromatography-MS/MS analysis of yeast Lys-C peptides. Center, a selected precursor with *m*/*z* at 827 collected with 30,000 resolving power (black) and the signal recorded in a high-resolution MS1 scan (480,000 resolving power) (red). Bottom, MS/MS spectrum following collisionally activated dissociation (CAD) and ion trap *m*/*z* analysis of the NeuCode SILAC pair. The inset shows that the NeuCode SILAC pair is concealed at typical resolution. Quant, quantitative; b, fragment peaks from N terminus; y, fragment peaks from C terminus.

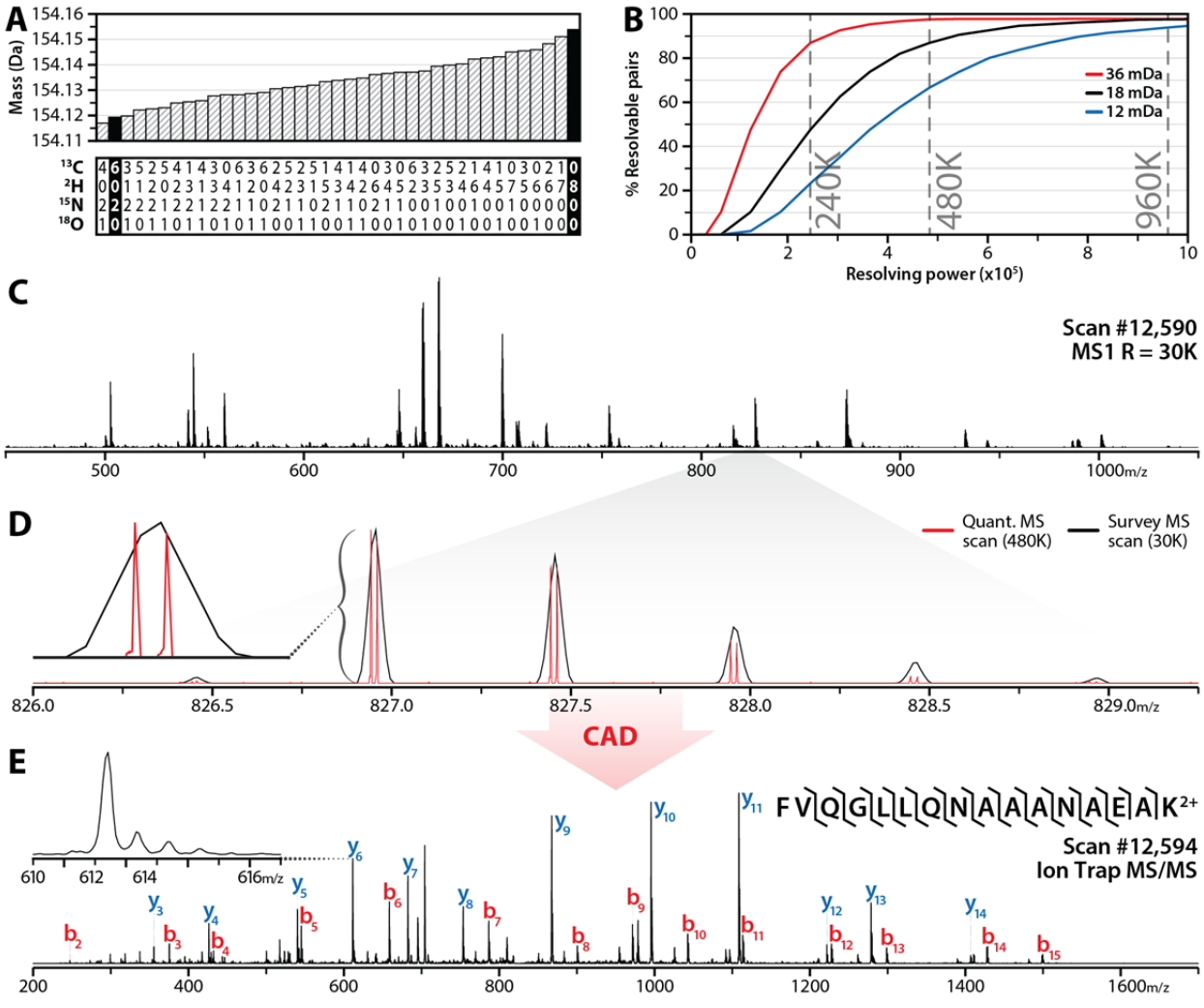
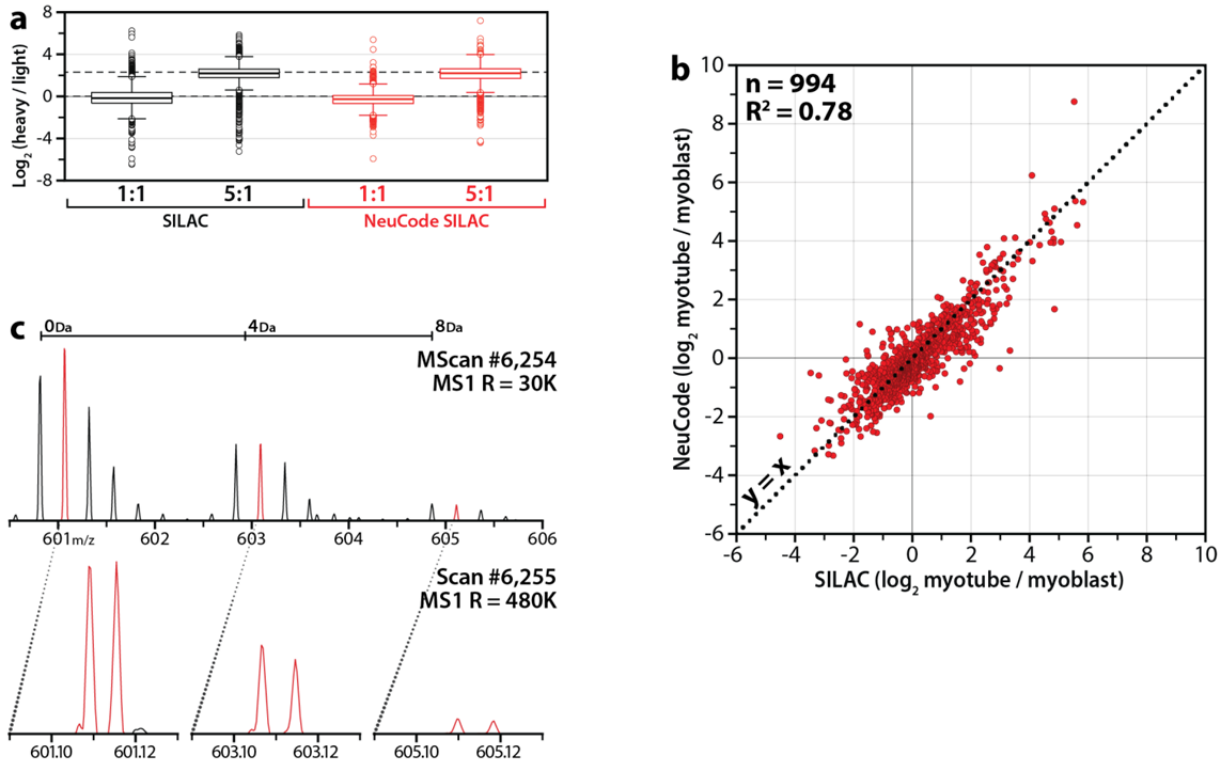


Figure 2: NeuCode SILAC quantitative results. (**a**) Box plots showing the measured (box and whiskers) and true (dashed lines) peptide ratios for both methods at mixing ratios of 1:1 and 5:1. Box plots demarcate the median (stripe), the 25th to 75th percentile (interquartile range, box), 1.5× the interquartile range (whiskers) and outliers (open circles) for both SILAC (black) and NeuCode SILAC (red). (**b**) NeuCode SILAC and traditional SILAC demonstrate a strong correlation for quantifying protein changes during the myogenic differentiation of mouse-derived C2C12 myoblasts (r^2 = 0.78). (**c**) The combination of NeuCode SILAC with mass differential tags for relative and absolute quantification (mTRAQ) labeling affords six channels of MS1-centric quantification. *R*, resolving power.



Experimental Procedures

Theoretical calculations

First**,** a library of 71,499 yeast endoproteinase LysC derived peptides identified by mass spectrometry was composed. The theoretical full width at 1% max peak height (*FWOM*) for each library peptide across resolving powers (*R*) from 15 thousand to 1 million is calculated by:

$$FWOM = 2.57756788 \times \frac{m/z}{R \times \sqrt{\frac{400}{m/z}}}$$

where resolving power is defined as the minimum *m/z* difference that can be resolved at 400 *m/z* and the coefficient (2.57756788) is derived from Gaussian peak shape modeling. The *m/z* difference (Δ *m/z*) for each theoretical isotope doublet assuming lysine isotopologue mass differences (Δ *I*) of 12, 18, and 36 mDa:

$$\Delta m/z = \frac{n \times \Delta I}{z}$$

where *n* is the number of lysines in the peptide sequence and *z* is the charge of the peptide. An isotopologue pair is only considered resolvable at the tested isotopologue mass difference and resolving power if Δ *m/z* > *FWOM*.

Sample preparation

Saccharomyces cerevisiae strain BY4741 Lys1Δ was grown in defined, synthetic-complete (SC, Sunrise Science) drop out media supplemented with either “light” unlabeled L-lysine (+0 Da), “heavy 1”

$^{13}C_6{}^{15}N_2$-L-lysine (+8.0142 Da, Cambridge Isotopes), or “heavy 2” $^{2}H_8$-L-lysine (+8.0502 Da, Cambridge Isotopes). Cells were allowed to propagate for a minimum of 10 doublings to ensure complete lysine incorporation. Upon reaching mid-log phase, the cells were harvested by centrifugation at 3,000 ×g for 3 minutes and washed three times with chilled ddH_2O. Cell pellets were re-suspended in 5mL lysis buffer (50mM Tris pH8, 8M urea, 75mM sodium chloride, 100mM sodium butyrate, 1mM sodium orthovanadate, protease and phosphatase inhibitor tablet), and total protein was extracted by glass bead milling (Retsch).

C2C12 cells were grown in DMEM lysine, arginine dropout culture media (Cambridge Isotopes) supplemented with 10% dialyzed FBS, antibiotics, 100 mg/L unlabeled L-arginine, and 100 mg/L of “light” unlabeled L-lysine, "heavy 1" $^{13}C_6{}^{15}N_2$-L-lysine, or "heavy 2" $^{2}H_8$-L-lysine for six passages. $1.3x10^6$ cells from these plates were seeded onto fresh plates with the same media type and allowed to grow for two days. Cells grown with “light” and "heavy 2" lysine were harvested, and the cells grown with “heavy 1” lysine were re-fed with DMEM supplemented with 2% dialyzed FBS, then allowed to differentiate for five additional days before harvesting. Cells were pelleted, and washed with ice-cold PBS. The cell pellets were resuspended in 8 M urea, 50 mM Tris pH 8.0, 5 mM $CaCl_2$, and protease inhibitors (Roche). Cells were lysed by sonication. For the label flipping experiment, cells were grown exactly as above, except “light”- undifferentiated and “heavy 2” - differentiated cells were harvested were for the SILAC experiment and “heavy 1” - undifferentiated and “heavy 2” - differentiated cells were harvested for the NeuCode experiment.

Lysate protein concentration was measured by BCA (Pierce). Protein was reduced by addition of 5 mM dithiothreitol and incubation for 30 minutes at ambient temperature. Free thiols were alkylated by addition of 15 mM iodoacetamide and incubated in the dark, at ambient temperature, for 30 minutes, followed by quenching with 5 mM dithiothreitol. Urea concentration was diluted to 4 M with 50 mM Tris pH 8.0. Proteolytic digestion was performed by addition of LysC (Wako), 1:50 enzyme to protein ratio,

and incubated at ambient temperature for 16 hours. The digest reaction was quenched by addition of TFA and desalted with a tC18 Sep-Pak (Waters).

SILAC yeast known ratios were prepared by mixing “light” = +0 Da and “heavy 1” = +8 Da labeled peptides in ratios 1:1 and 1:5 by mass. NeuCode SILAC ratios were prepared exactly the same, except “heavy 1” = +8.0142 Da and “heavy 2” = +8.0502 Da.

C2C12 mouse peptides were mixed 1:1 by mass. For the SILAC comparison, “light” peptides from myoblasts were combined with “heavy 1” peptides from myotubes. The NeuCode SILAC comparison combined “heavy 1” myotube peptides with “heavy 2” myoblast peptides.

6-plex samples were prepared by labeling each NeuCode SILAC yeast peptide with three mTRAQ tags (AB SCIEX), according to the manufacturer’s protocol, except that hydroxylamine was added to quench the labeling reaction after 2 hours. These peptides were mixed in the ratio 10:10:5:5:1:1 by mass.

LC-MS/MS

Each sample was injected onto a 75 μm capillary packed with 30 cm of 5 μm Magic C18 (Michrome) particles in mobile phase A (0.2% formic acid in water). Peptides were gradient-eluted with mobile phase B (0.2% formic acid in acetonitrile). The known ratios and 6-plex experiments employed a 60 minute gradient, while the mouse experiment used a 300 minute gradient. Eluted peptides were analyzed by an Orbitrap Elite mass spectrometer (Thermo Scientific). A survey scan was performed by the Orbitrap at 30,000 resolving power to identify precursors to sample for data dependent top-10 ion trap CAD tandem mass spectrometry. NeuCode SILAC analysis had an additional quantitative 480,000 resolving power scan immediately following the survey scan. The 30,000 resolving power scan assigned precursor charge state more often than the 480,000 resolving power scan and thus was used to trigger MS/MS events. The traditional SILAC sample mixed in a 1:1 ratio was additionally analyzed using 60,000 or 120,000 resolution for survey scans; increased resolving power did not lead to significant

improvements in unique peptide sequences identified, quantitative accuracy, nor precision. The 480,000 resolving power scan is available on Orbitrap Elite systems equipped with the developer's kit. Preview mode was enabled, and precursors with unknown charge, or charge = +1, were excluded from MS/MS sampling. MS1 and MS/MS target ion accumulation values were set to 1 x 10^6 and 4 x 10^4, respectively. Dynamic exclusion was set to 30 seconds for -0.55 *m/z* and +2.55 *m/z* of selected precursors.

Data analysis

MS raw files were converted to searchable text files and searched against a target-decoy database (*Saccharomyces* Genome Database (yeast), www.yeastgenome.org, February 3, 2011; Un-iProt (mouse), www.uniprot.org, October 1, 2011) using the Open Source Mass Spectrometry Search Algorithm (OMSSA). For all samples, methionine oxidation and cysteine carbamidomethylation were searched as variable and fixed modifications, respectively. SILAC samples were searched independently with an unmodified lysine and +8.014199 Da fixed modification, and later combined during false discovery rate filtering. NeuCode SILAC samples were searched with a single fixed modification representing the average mass increase of the $^{13}C_6{}^{15}N_2$ and $^{2}H_8$ isotopologues (+8.0322 Da) compared to unmodified lysine. Precursor mass tolerance was defined as 100 ppm and fragment ion mass tolerance was set to 0.5 Da. This relatively wide precursor mass tolerance was used to account for the mass difference observed between isotopologues. Search results were filtered to 1% FDR based on E-values. Peptides were grouped into proteins and filtered to 1% FDR according to rules previously described.

Quantification

Following database searching, the FDR-filtered list of peptide-spectral matches was first utilized to calculate the systematic precursor mass error associated with the data set. After adjusting precursor masses for this error, every high-resolution MS1 scan within 30 seconds of all PSMs identifying a unique peptide sequence was inspected; this retention time window ensured that quantitative data was extracted

throughout the entirety of a peptide's elution and accommodates any chromatographic shifts between isotopologues due to deuterium incorporation. In each MS1 scan (480,000 resolution for NeuCode; 30,000 resolution for traditional) quantitative pairs were isolated for the mono and first two isotopes of the isotopic cluster. If at least two peaks, with signal-to-noise greater than 3, were found within the specified tolerance (±5 ppm for NeuCode SILAC; (±10 ppm for SILAC), a SILAC pair was created. Any peaks below the noise level simply contributed a noise-based intensity to the appropriate missing channel. Peaks exhibiting possible peak coalescence, as determined by de-normalizing intensity by injection time, were excluded from quantification. The intensities for each channel were summed across the peptide's elution profile so that chromatographic shifts between isotopologues did not impair quantification. To eliminate the noise-capped peaks on the fringes of a peptide's elution profile compressing the quantitative ratio towards 1:1, peaks with intensities below $1/2e$ the maximum intensity were discarded (e refers to the mathematical constant); this peak filtering threshold is approximately the 1.75th standard deviation of the standard normal distribution (the assumed shape of a peptide's elution profile); the distribution area covered by ±1.75 standard deviations is > 90%. Peptides were required to have a minimum of 3 ratio-providing pairs (i.e., quantified across at least 3 MS1 scans) to be eligible for quantification. Protein quantification was accomplished by averaging the ratios of all corresponding peptides. The resulting protein ratios were normalized to a median fold-change around 1 to account for unequal mixing. This algorithm was utilized to quantify both traditional and NeuCode SILAC data sets.

To perform manual analysis of NeuCode SILAC peptides, zoom in on the precursor in the MS raw file, look for properly-spaced isotopologue peaks, and record their intensities. Ion chromatograms can be extracted by centering a 10 ppm window on the observed *m/z* (i.e., peak *m/z* ± 5 ppm) and integrating the peak areas.

To adapt existing quantification algorithms to analyze NeuCode data, set the lysine label masses to +8.0142 Da and +8.0502 Da and search for peaks within a ± 5 ppm tolerance of theoretical precursor masses that have been adjusted for systematic error.

Acknowledgements

We thank A.J. Bureta for help with figure illustrations, Rebecka Manis for critical proofreading, and Audrey Gasch for assistance in the culturing of yeast cells. This work was supported by the National Institutes of Health grant R01 GM080148 to J.J.C. A.E.M. gratefully acknowledges support from a National Institutes of Health-funded Genomic Sciences Training Program (5T32HG002760).

References

1 Domon, B. & Aebersold, R. (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology* **28**, 710-721, doi:Doi 10.1038/Nbt.1661

2 Altelaar, A. F. M., Frese, C. K., Preisinger, C., Hennrich, M. L., Schram, A. W., Timmers, H. T. M., Heck, A. J. R. & Mohammed, S. Benchmarking stable isotope labeling based quantitative proteomics. *Journal of Proteomics*, doi:10.1016/j.jprot.2012.10.009

3 Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386

4 Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **96**, 6591-6596

5 Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. & Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999, doi:10.1038/13690

6 Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. & Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904

7 Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169, doi:10.1074/mcp.M400129-MCP200

8 Hsu, J.-L., Huang, S.-Y., Chow, N.-H. & Chen, S.-H. (2003) Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Analytical Chemistry* **75**, 6843-6852, doi:10.1021/ac0348625

9 Choe, L., D'Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P. & Lee, K. H. (2007) 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* **7**, 3651-3660

10 McAlister, G. C., Huttlin, E. L., Haas, W., Ting, L., Jedrychowski, M. P., Rogers, J. C., Kuhn, K., Pike, I., Grothe, R. A., Blethrow, J. D. & Gygi, S. P. (2012) Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Anal Chem* **84**, 7469-7478, doi:10.1021/ac301572t

11 Werner, T., Becher, I., Sweetman, G., Doce, C., Savitski, M. M. & Bantscheff, M. (2012) High-Resolution Enabled TMT 8-plexing. *Anal Chem* **84**, 7188-7194, doi:10.1021/ac301553x

12 Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I. & Wright, P. C. (2009) iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* **8**, 5347-5355, doi:10.1021/pr900634c

13 Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. & Coon, J. J. (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Methods* **8**, 933-935, doi:10.1038/nmeth.1716

14 Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci U S A* **104**, 5860-5865, doi:0608638104 [pii] 10.1073/pnas.0608638104

15 Sleno, L. (2012) The use of mass defect in modern mass spectrometry. *J Mass Spectrom* **47**, 226-236, doi:10.1002/jms.2953

16 Denisov, E., Damoc, E., Lange, O. & Makarov, A. (2012) Orbitrap mass spectrometry with resolving powers above 1,000,000. *International Journal of Mass Spectrometry* **325–327**, 80-85, doi:10.1016/j.ijms.2012.06.009

17 Xian, F., Hendrickson, C. L., Blakney, G. T., Beu, S. C. & Marshall, A. G. (2010) Automated Broadband Phase Correction of Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Analytical Chemistry* **82**, 8807-8812, doi:10.1021/ac101091w

18 Schaub, T. M., Hendrickson, C. L., Horning, S., Quinn, J. P., Senko, M. W. & Marshall, A. G. (2008) High-performance mass spectrometry: Fourier transform ion cyclotron resonance at 14.5 Tesla. *Anal Chem* **80**, 3985-3990, doi:10.1021/ac800386h

19 Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Muller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J. J., Cox, J., Horning, S., Mann, M. & Makarov, A. (2012) Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* **11**, O111 013698, doi:O111.013698 [pii] 10.1074/mcp.O111.013698

20 Cui, Z., Chen, X., Lu, B., Park, S. K., Xu, T., Xie, Z., Xue, P., Hou, J., Hang, H., Yates, J. R., 3rd & Yang, F. (2009) Preliminary quantitative profile of differential protein expression between rat L6 myoblasts and myotubes by stable isotope labeling with amino acids in cell culture. *Proteomics* **9**, 1274-1292, doi:10.1002/pmic.200800354

Chapter 3

Amine-reactive neutron-encoded labels for highly-plexed proteomic quantitation

ASH designed research, synthesized labels, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Hebert, A. S., Merrill, A. E., Stefely, J. A., Bailey, D. J., Wenger, C. D., Westphall, M. S., Pagliarini, D. J. & Coon, J. J. (2013) Amine-reactive Neutron-encoded Labels for Highly Plexed Proteomic Quantitation. *Mol. Cell. Proteomics* **12**, 3360-3369

Abstract

We describe a novel amine-reactive chemical label that exploits differential neutron-binding energy between ^{13}C and ^{15}N isotopes. These neutron encoded (NeuCode) chemical labels enable up to 12-plex MS1-based protein quantification. Each structurally identical, but isotopically unique tag is encoded with a 12.6 mDa mass difference – relative to its nearest neighbor – so that peptides bearing these NeuCode signatures do not increase spectral complexity and are only detected upon analysis with very high mass resolving powers. We demonstrate that the method provides quantitative performance that is comparable to both metabolic labeling and isobaric tagging, while combining the benefits of both strategies. Finally, we employ the tags to characterize the proteome of *Saccharomyces cerevisiae* during the diauxic shift — a metabolic transition from fermentation to aerobic respiration.

Introduction

Proteome quantification is an increasingly essential component of modern biology and translational medicine[1,2]. Whether targeted or global, stable isotope incorporation with mass spectrometry (MS) analysis is a core technique for protein abundance measurements. There are numerous approaches to introduce stable isotopes into peptides, the most frequently used being stable isotope labeling with amino acids in cell culture (SILAC) and isobaric tagging (TMT, or iTRAQ)[3-7]. Both of these methods incorporate heavy isotopes to increase mass by at least 1 Da. Metabolic labeling with amino acids in cell culture (SILAC) is the quantification gold standard for global proteomic analysis. However, the SILAC approach is not easily adapted to tissue sample analysis; SILAC mouse labeling, for example, requires feeding mice a specialized diet for multiple generations[8]. Tissue samples can be analyzed by mixing with SILAC cell-culture-based labeled standards, but this strategy does not permit multi-plexing[9]. Isobaric labels, on the other hand, are conjugated to the primary amines of peptides following proteolytic digestion and, thus, have the advantage of being completely compatible with samples from virtually any source[10,11]. That said, isobaric tagging suffers from dynamic range suppression caused by co-isolation of precursor peptides[12]. Multiple studies have revealed that this problem greatly erodes quantitative accuracy – *e.g.*, ten-fold changes often are detected as much smaller ~four-fold changes[13,14]. Here we propose a new approach to protein quantification, one that achieves tissue compatible 4-plexed MS1-based quantification without increasing spectral complexity.

Recently, we described the use of mass defects to expand SILAC quantification from three-plex to 12-plex and beyond (*i.e.*, NeuCode SILAC)[15]. NeuCode SILAC exploits the subtle mass differences that exist in atoms due to the varying energies of nuclear binding in common stable isotopes (*e.g.*, carbon, nitrogen, hydrogen, and oxygen) by using the extremely high resolving power of modern Fourier transform mass spectrometer systems (FT-MS)[16,17]. For example, the multiplexing capability of TMT was increased from 6 to 8 by incorporating a difference in mass of 6.3 mDa in specific reporter ions by swapping ^{14}N for a ^{15}N atom while concomitantly switching a ^{13}C with a ^{12}C atom[18,19]. This method only

requires 30,000 resolving power to resolve the reporter ions, but still suffers from the interference problem described above. Repetition of this process, within the context of an analyte molecule, can generate several chemically identical isotoplogues that when analyzed under normal MS analysis conditions (resolving power, R,) are indistinguishable – *i.e.*, produce one *m/z* peak. Analysis of these NeuCode *m/z* peaks with high resolving power 480,000), however, often reveals distinct *m/z* peaks whose abundances can be extracted and used to determine analyte quantity across the sundry conditions. This strategy permits very high levels of MS1-based multi-plexing (> 10), which has several advantages. First, MS1 scans across the entire analyte elution profile can be averaged to increase quantitative accuracy and precision; Second, a tandem mass spectrum is not necessary for quantification. Mann and colleagues showed that once a peptide is identified by tandem MS analysis it can be confidently identified in other runs with only the exact mass and elution profile matching[20]. Second, MS1-based quantification does not suffer from the pervasive problem of precursor interference that cripples the quantitative accuracy of the isobaric tagging strategies[13].

The NeuCode SILAC approach, however, relies on the use of amino acid isotopologues and still requires metabolic incorporation. Chemical labeling strategies for proteome quantification can be convenient and, for certain systems, are requisite. We reasoned that our NeuCode strategy could be extended to create novel chemical reagents for proteome quantification. As such, we describe here the design, synthesis, and use of 12-plex NeuCode amine reactive labels for global proteome quantification.

Results

Theoretical considerations

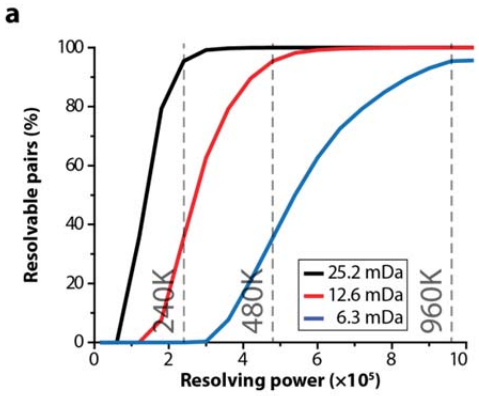
NeuCode SILAC requires that the mDa mass signatures be embedded into individual amino acids. Amino acids, of course, offer a limited palette – that is, there are only so many atoms to adjust[21]. For lysine, $C_6H_{14}N_2O_2$, up to 18 atoms can be adjusted – C6H9N2O1 – after excluding exchangeable atoms and losses during incorporation. Achieving the highest plexing with NeuCode SILAC will require stable

isotope labeled isotopologues that make use of all these atoms (*i.e.*, C, H, N, and O). Some elements, however, have greater mass defect than others – $^{12}C/^{13}C$ (+3.3 mDa), $^{1}H/^{2}H$ (+6.3 mDa), $^{16}O/^{18}O$ (+4.2 mDa), $^{14}N/^{15}N$ (-4.2 mDa). Deuterium offers the largest mass defect; however, it also can induce subtle chromatographic peak shifts[22]. We reasoned that deuterium should be avoided. Thus, we restricted our tag design to only use differential incorporation of C and N isotopes. In this scenario, each concomitant exchange of a ^{13}C atom for a ^{12}C and a ^{14}N for a ^{15}N imparts a 6.3 mDa mass defect.

To evaluate the feasibility and practicality of this approach we constructed a library containing 65,011 tryptic yeast peptides and generated (*in silico*) versions of each library entry that were spaced by multiples of the 6.3 mDa mass defect (6.3, 12.6, and 25.2 mDa). Next, we plotted the percentage of the proteome that was quantifiable (*i.e.*, resolved, full-width at ten percent max, FWTM) at resolving powers ranging from 15,000 to 10^6. The underlying calculation accounts for the decreased resolution at higher m/z, as well as the mass (*m*), charge (*z*), label mass, and the number of primary amines of each precursor (**Figure 1a**)[23]. In this calculation, *m/z* peaks overlapping by up to 3.2% peak area (FWTM) are considered resolvable and, thus, quantifiable. Partially resolved peaks overlapping by >10% to 100% full width of max, do not meet this criteria. The optimal mass defect produces the highest percentage of quantifiable identifications at the lowest resolving power and is also capable of the highest degree of multi-plexing (*i.e.*, smallest mass spacing). While Orbitrap mass analyzers can achieve resolution of over 10^6, current commercial implementations can routinely achieve up to 480,000 resolution (available when the Thermo Scientific Developer's software is installed). [24,25] From this analysis we conclude that 12.6 mDa is the current optimal mass defect for incorporation into NeuCode tags. While a mass defect of 6.3 mDa, the exchanging of a heavy N for a heavy C, would offer the highest plexing capacity, ~35% of the peptides carrying this signature would be quantifiable. Doubling this spacing to 12.6 mDa will render 95% of the proteome quantifiable at the standard commercial resolution of 480,000. We note the 6.3 mDa spaced tags will be more accessible in the coming years as resolution of MS instruments is expected

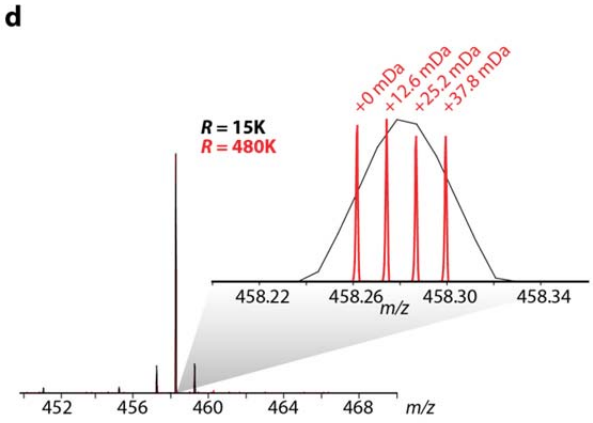
to continually improve. Fourier transform-ion cyclotron resonance-MS instruments (FT-ICR-MS) offer the highest resolution and can analyze samples labeled with the NeuCode tags described here[26,27].

Figure 1 | NeuCode amine-reactive label development and design. a, Theoretical calculation depicting the percentage of peptides that are resolved at full-width at 10% maximum peak height (FWTM) when spaced 6.3, 12.6, or 25.2 mDa for resolving powers 15 thousand to 1 million on an Orbitrap (n = 65,011). **b**, Acetylarginine-acetyllysine-glycine-NHS structure and isotopic composition. The (*) denotes each atom in which a ^{13}C or ^{15}N isotope can be incorporated. **c**, The specific isotopic compositions that encode the 12.6 mDa mass differences for each of the 4-plex labels are listed in the table. **d**, Spectra from synthesized acetylarginine-acetyllysine-glycine tripeptide encoding 12.6 mDa mass differences.



c

Label	AcArg		AcLys		Total		Isotope mass
	^{13}C	^{15}N	^{13}C	^{15}N	^{13}C	^{15}N	
1	0	4	0	2	0	6	5.9822
2	0	4	2	0	2	4	5.9948
3	2	0	2	2	4	2	6.0075
4	6	0	0	0	6	0	6.0201



Label characteristics.

With these theoretical calculations in hand, we next sought to develop a proof-of-concept NeuCode tag with the following objectives: (1) demonstrate compatibility of the NeuCode method with chemical labeling, (2) deliver high plexing capacity, and (3) confirm that quantitative accuracy and precision are comparable to other MS1-based quantification methods. Design of the tag structure must balance a number of considerations. The tag should have limited effect on nanoHPLC-MS/MS parameters, such as peptide chromatographic retention, ionization, and fragmentation. Thus, the tag should be reasonably hydrophilic (hydrophilic amino acids) and contain a basic moiety that cannot be a primary amine. Further, the number of nitrogen atoms should be maximized while limiting label size[28]. Incorporation of a large number of ^{15}N into a tag is more challenging than incorporation of ^{13}C, simply because most molecules have fewer N atoms than C atoms.

We reasoned that amino acids presented somewhat flexible building blocks from which to construct the NeuCode tagging reagents. Specifically, many isotope varieties of amino acids are commercially available and are readily combined by straightforward Fmoc solid phase synthesis[29,30]. Considering a number of properties, we decided on the sequence $_{acetyl}$Arg-$_{acetyl}$Lys-Gly with an NHS ester at the C-terminus (**Figure 1b**). Arg and Lys were selected for their relatively high N content. In fact, Arg has the highest ratio of N to mass (4 N:175 Da) of any amino acid. The presence of Arg also ensures that the tag contains a positive charge. Gly was present on the solid phase synthesis resin to simplify the initial coupling step. The acetyl groups block the primary amines and prevent the tag from polymerizing during NHS ester activation. The NHS ester moiety is reactive towards primary amines and is commonly used for conjugating labels to peptides prior to MS analysis[6,7]. Asp and Glu were avoided because of their potential to deamidate. **Figure 1c** presents the isotope incorporation plan to generate a four-plex NeuCode tag using the $_{acetyl}$Arg-$_{acetyl}$Lys-Gly tag scaffold. Our plan incorporates 6 heavy atoms in four combinations, $^{13}C_0{}^{15}N_6$, $^{13}C_2{}^{15}N_4$, $^{13}C_4{}^{15}N_2$, and $^{13}C_6{}^{15}N_0$, spanning 37.9 mDa. Using commercial components and standard Fmoc chemistry, we synthesized all four of these NeuCode tags and then

examined a 1:1:1:1 mixture of them using a linear ion trap-Orbitrap hybrid mass spectrometer (**Figure 1d)**. At 15,000 resolving power, only a single isotopic cluster is observed and the presence of the four isotopologues is cloaked; collection of the same MS1 scan at 480,000 resolution, however, reveals the four reagents, spaced by precisely 12.6 mDa. We note that, given its size, this label is not intended as a long term strategy for incorporation of mass defect signatures, but as a rather convenient scaffold to demonstrate the concept.

4-plex NeuCode labeling.

To evaluate the performance of these four-plex NeuCode tags we labeled four equivalent aliquots of a complex mixture of yeast tryptic peptides. We confirmed that each sample was >90% labeled, then mixed the labeled peptides in various ratios – 1:1:2:5 – and analyzed the mixture by nanoHPLC-MS/MS. The MS acquisition method included a fast MS1 survey scan (R = 30,000), a MS1 quantitation scan (480,000), and then MS/MS of the ten most intense precursors (collision-activated dissociation, CAD) using the ion trap (note in some instances we utilized HCD with Orbitrap mass analysis)[31]. Previously, we found that the added time required for the high resolution quantitation scan (~1.6 s) can be mitigated by use of the ion trap analyzer for MS/MS scanning[24]. We directly compared peptides from a yeast tryptic digest LC-MS/MS analysis, with a 60 min gradient elution, using a method with a 480,000 resolving power scan and one with only a 30,000 resolving power scan. The 480K method resulted in 824 unique protein groups, and 4,497 unique peptide identifications from 17,150 MS/MS while the 30K method resulted in 836 unique protein groups, and 4,725 unique peptides from 19,185 MS/MS scans. **Figure 2** presents representative spectra and results from this experiment. Here we consider a precursor at *m/z* 950 that was selected for MS/MS analysis. The high resolution quantification scan reveals the embedded NeuCode signature at the appropriate mixing ratios of 1:1:2:5, while the MS/MS scan, conducted in the low resolution ion trap analyzer, shows backbone fragmentation and no evidence of the isotopologue fingerprint. From this experiment 1,738 peptides were confidently identified and 1,448 were

quantified (83%). We note the overall MS/MS identification rate was somewhat depressed, as compared to unlabeled peptides, presumably due to the presence of the tag (see discussion below). For peptides that were confidently identified, however, the distribution of experimentally measured ratios were centered within ~20% of the expected values (median measured peptide ratios = 1.0:1, 1.7:1, 4.0:1, **Figure 2c**). Evaluation of median accuracy compares well with reported SILAC and isobaric labeling results. For example, when analyzed in a recent study the median value for a SILAC 2:1 ratio was 2.9:1, and our own laboratory discovered that 5:1 ratios in isobaric tagging are often detected as much lower, 2.5:1 on average[13,32]. Additionally, calculated standard deviations for each measurement (0.5 – 0.7, $\log_2$) are in line with typical performances for both SILAC (0.35 - 0.7) and isobaric labeling (0.7 – 0.8)[13,14,32,33]. Precision is improved (0.5 to 0.4, for the 1:1 ratio) when analyzing protein values as multiple peptide measurements are often included for calculating a single protein ratio. Results from the 1:1 ratio indicate that 99% of proteins fall within a 100% fold-change. We do note that higher ratios exhibit some dynamic range depression and also correlate with decreased precision, just as is often observed in isobaric labeling and traditional SILAC measurements[32,33]. Interestingly, a 4-plex tag with 12.6 mDa spacing, analyzed at 480,000 resolving power, typically encompasses a similar m/z space than a single peak analyzed at 30,000 resolving power. This suggests that MS1 interference of a 4-plex NeuCode tag will be similar to that of a label free or SILAC approach (**Figure 3**). We conclude that, while there is room for improvement, the overall accuracy, precision, and dynamic range afforded by these NeuCode tags is comparable to other current technology. And, with continued efforts, we anticipate that improvements in tagging chemistry, instrument control methods, and quantitative software are likely to boost each of these key Figures of Merit.

Figure 2 | NeuCode label example and quantitative Figures of Merit. a, MS1 scan collected from an nanoHPLC-MS/MS analysis of NeuCode four-plex labeled yeast tryptic peptides mixed in the ratio 1:1:2:5. The inset displays the 30,000 resolving power survey scan (black) and subsequent 480,000 resolving power quantitation scan (red) for a selected precursor having *m/z* of 950. **b**, Annotated MS2 spectrum and sequence, of the peptide from **a,** following CAD and ion trap *m/z* analysis. Inset displays that the SILAC pair is concealed at typical resolution. Fragment ions resulting from tag cleavage are denoted by (*). **c**, a) Boxplots showing the measured (box and whiskers) and true (dashed lines) values for mixing ratios of 1:1, 1:2, and 1:5 (n = 1,448). Boxplots demarcate the median (stripe), the 25th to 75th percentile (interquartile range, box), 1.5 times the interquartile range (whiskers), and outliers (open circles). The expected mixing ratio, median ratio and standard deviation of all measurements for each ratio are listed below the boxplots.

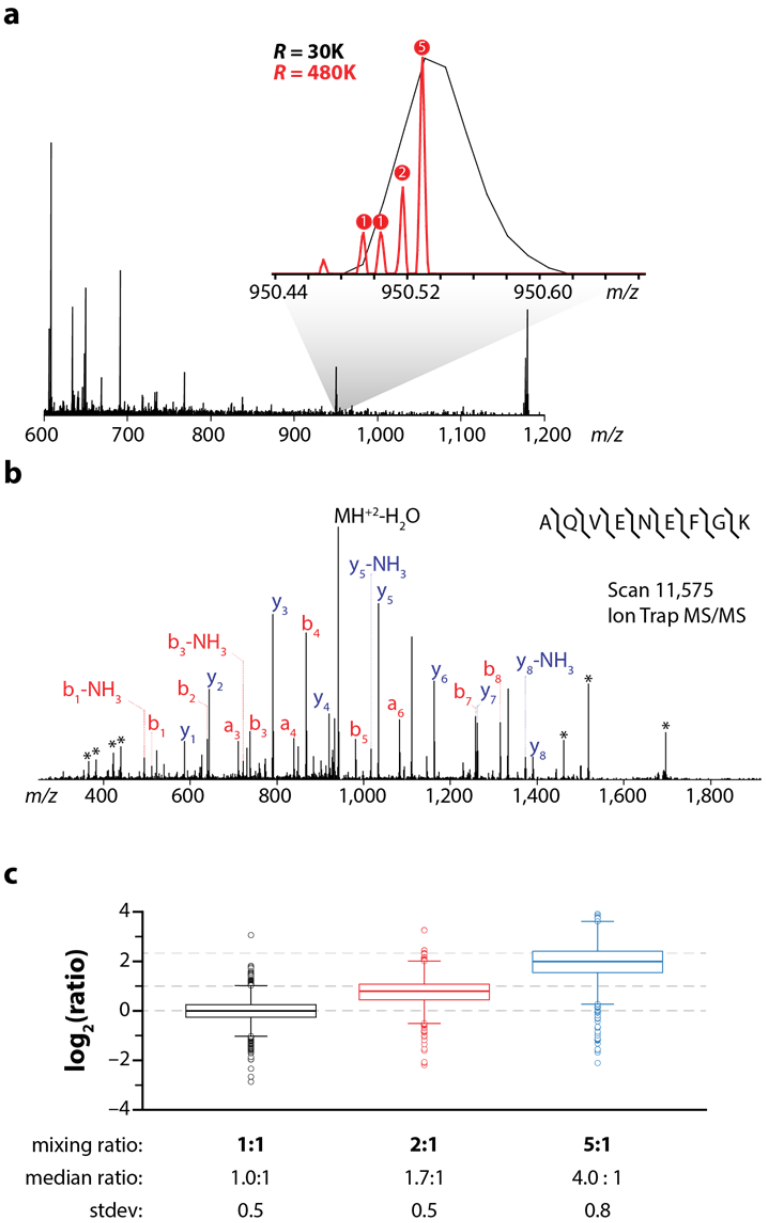
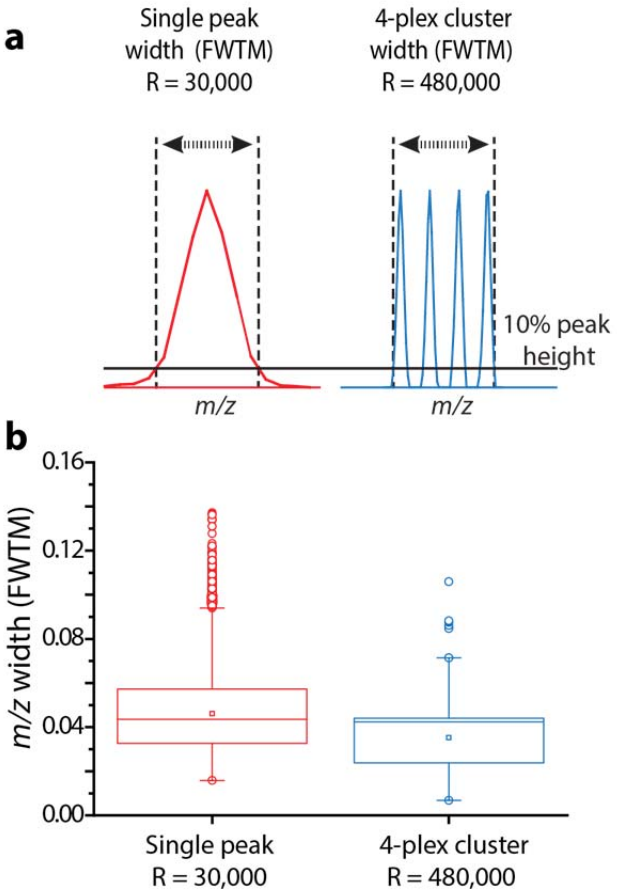


Figure 3 | Comparison between the widths of single peaks and 4-plex clusters. a, Single peak widths **(**red) were calculated as the FWTM when analyzed at R = 30,000. The width of a 4-plex cluster (blue) is calculated as the difference between the largest and smallest m/z that corresponds to 10% maximum peak height, when analyzed at R = 480,000. b, Box plot of calculated widths for single peaks and for 4-plex clusters from a library of 65,011 identified peptides.



Proteome remodeling during the diauxic shift

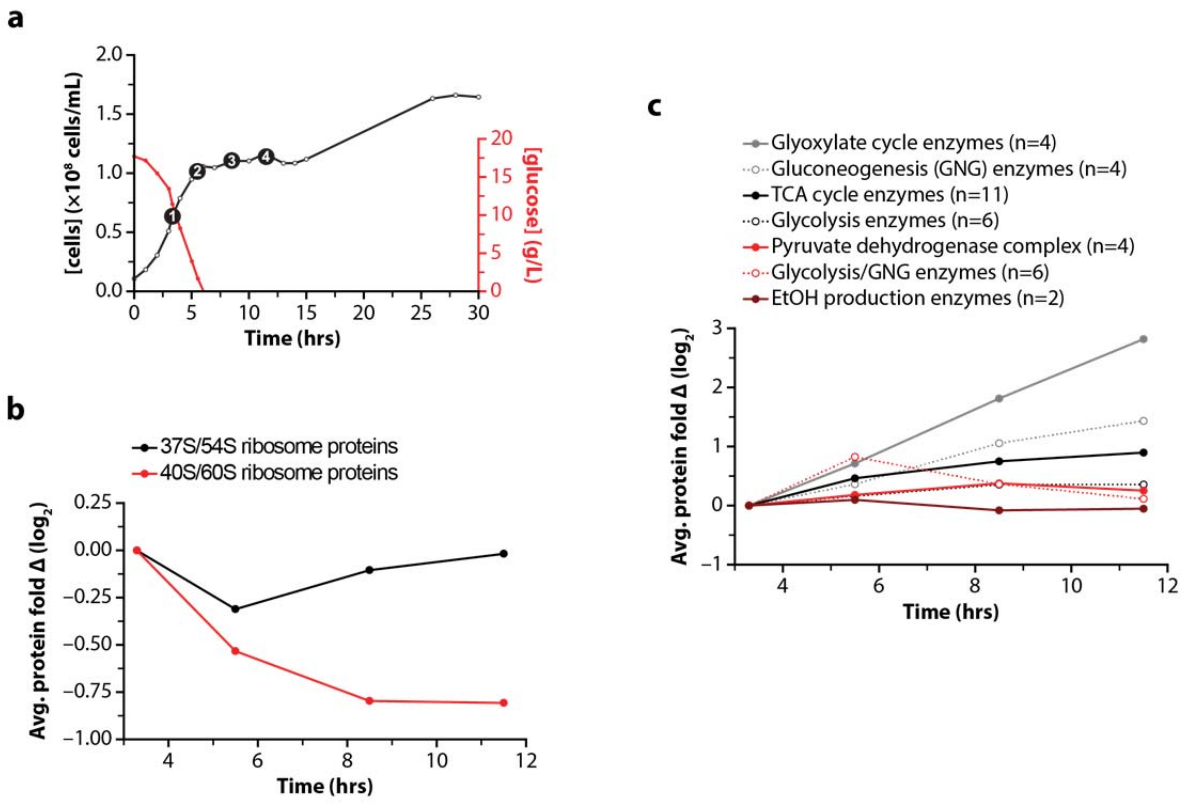
With these proof-of-concept data in hand we sought to test the performance of the four-plex NeuCode tags in quantifying proteome changes during the diauxic shift, a metabolic transition from fermentation to aerobic respiration, in *Saccharomyces cerevisiae*. This experiment provides an excellent test-bed as the phenomenon has been thoroughly investigated at the genetic and molecular level, but not yet at the global proteome plane[34-36]. Briefly, yeast can metabolize organic molecules for energy through two fundamentally different metabolic processes – fermentation and respiration. In the presence of abundant glucose, fermentation predominates. After consuming the available glucose, yeast restructure their metabolic machinery to enable respiration and to use fuel sources other than glucose. Yeast cultures undergo exponential growth in liquid media that contains abundant glucose, temporarily stop proliferating when the glucose is consumed, and then resume growth after a lag period, albeit at a slower rate. The changes in proliferation rate reflect complex changes in gene expression, protein abundance, and post-translational modifications during the diauxic shift[34-38].

Yeast were collected during log phase growth (3.3 hrs) while glucose was still available, and at three time points following glucose exhaustion (5.5, 8.5, and 11.5 hrs, **Figure 4a**). Yeast protein from these four samples was digested with trypsin and independently conjugated to the 4-plex NeuCode tags. Labeled peptides were combined and fractionated by high pH reversed-phase chromatography and analyzed by nanoHPLC-MS/MS. From these data, 1,543 proteins were confidently identified (1% FDR), enabling protein fold-change measurement over the time course. As anticipated, major proteomic transformations were observed: between the 3.3 hr and 11.5 hr time points, 281 proteins had abundances that changed by at least 2-fold. Approximately 1/3 of the proteins whose abundances decreased ($\geq$ 2-fold) during the diauxic shift were members of the 40S/60S ribosome, the major yeast cytosolic ribosome. The protein level data confirms transcriptomic results for these ribosomal genes during the diauxic shift, a response that reflects the decreased availability of nutrients needed to fuel protein synthesis and rapid cell division[34]. Alternatively, the abundances of the mitochondrial 37S/54S ribosomal proteins initially

decreased, but then increased following glucose depletion from the culture media (**Figure 4b**)[39]. The mitochondrial ribosome is essential for translation of proteins encoded by mitochondrial DNA, including proteins that are required for oxidative phosphorylation. The observed increase in mitochondrial ribosome abundance may enable an increase in respiratory capacity in response to decreased glucose availability. This result is consistent with previous analyses of the transcriptome and individual proteins of the yeast mitochondrial ribosome, such as *MRP20* and *MRP49*, which are more abundant in media without glucose[40].

A number of proteins whose abundances increased during the diauxic shift were highly-regulated enzymes in carbon metabolism, such as phospho-enolpyruvate carboxykinase, a key enzyme in gluconeogenesis. To further characterize these changes, we compared the average fold changes of enzymes from multiple pathways of central carbon metabolism (**Figure 4c**). As expected, enzymes involved in fermentation and ethanol production did not increase during the diauxic shift; however, enzymes of the tricarboxylic acid (TCA) cycle and the glyoxylate cycle, metabolic pathways that are tightly linked to respiration, increased during the diauxic shift. Further reflecting the decrease in glucose availability that drives the diauxic shift, an increase in enzymes of gluconeogenesis was observed. These data demonstrate that the NeuCode tagging technology can accurately determine changes in protein abundance from complex biological systems.

Figure 4 | Proteome remodeling during the diauxic shift. **a**, Growth curve for yeast cells (black) and plot of glucose concentration in the media (red) during the diauxic shift experiment. Each of the four selected time points are shown as black circles. **b**, Average expression change of quantified mitochondrial (red) and cytosolic (black) ribosomes over the time course. **c**, Average expression change of all quantified proteins involved in specific carbon metabolism pathways.



12-plex NeuCode labeling

A key benefit of multi-plexed analysis is the ability to simultaneously acquire data across multiple time points and biological replicates. By incorporation of biological replicate measurement one can establish a quantitative false discovery rate to distinguish statistically significant biological differences[41,42]. We reasoned that our four-plex NeuCode tag could be expanded to twelve-plex by superposition of nominal integer (~ 4 Da) mass differences onto the 12.6 mDa separated four-plex. To achieve this, we substituted 0, 4, or 8 additional ^{13}C atoms for ^{12}C atoms into the mass defect labels described above (**Figure 5a**). Peptides conjugated to these NeuCode labels will appear as three disparate isotopic distributions when analyzed at 30,000 resolutions, similar to triplex SILAC; however, each isotopic cluster peak is resolved into a set of four discrete *m/z* peaks separated by 12.6 mDa (per tag) when analyzed at higher resolution (~ 480,000, **Figure 5b**). To test the performance of these 12-plex NeuCode tags we analyzed the yeast diauxic shift time-course experiment in biological triplicate. Here, every isotopic cluster comprises quantitative data for each of the four time points of a complete biological replicate. From these data 1,056 proteins were detected, 760 of which were quantified across all twelve samples. The increased spectral complexity of the 12-plex experiments (three isotopic clusters), is likely responsible for the decreased proteomic coverage when compared to the 4-plex analysis. **Figure 6** demonstrates reproducibility (r^2 values ranging from 0.70 – 0.80) for the NeuCode tagging method between biological replicates. Indeed, recent work with isobaric tagging indicates typical correlation for that chemistry of ~ 0.70[43]. We leveraged the replicate measurements to perform rigorous statistical analysis to identify significant biological differences ($p < 0.05$). 85, 273, and 270 proteins significantly differed in abundance at 5.5, 8.5, and 11.5 hrs, as compared to 3.3 hrs, respectively (2-fold change and $p < 0.05$). Interestingly, only four proteins were found changing between time points three and four, suggesting the yeast proteome had adjusted to the lack of glucose by the third time point. Despite the tripled plexing capacity, the average standard is commensurate with the results achieved for the four-plex NeuCode tag and with both SILAC and isobaric tagging[13,14,32,33].

Figure 5 | NeuCode 12-plex labels. a. Specific isotopic compositions that encode the 12.6 mDa and 4 Da mass differences for each of the 12-plex labels. **b**, A peptide from *RPL17A* is observed as three isotopic distributions when analyzed with a 30,000 resolving power scan. A 480,000 resolving power scan, inset, reveals the mass differences imparted by the twelve labels.

a

Label	AcArg		AcLys		Total		Isotope mass
	^{13}C	^{15}N	^{13}C	^{15}N	^{13}C	^{15}N	
1	0	4	0	2	0	6	5.9822
2	0	4	2	0	2	4	5.9948
3	2	0	2	2	4	2	6.0075
4	6	0	0	0	6	0	6.0201
5	2	4	2	2	4	6	9.9956
6	6	4	0	0	6	4	10.0083
7	0	0	8	2	8	2	10.0209
8	8	0	2	0	10	0	10.0336
9	6	4	2	2	8	6	14.0090
10	8	4	2	0	10	4	14.0217
11	6	0	6	2	12	2	14.0343
12	6	0	8	0	14	0	14.0470

b

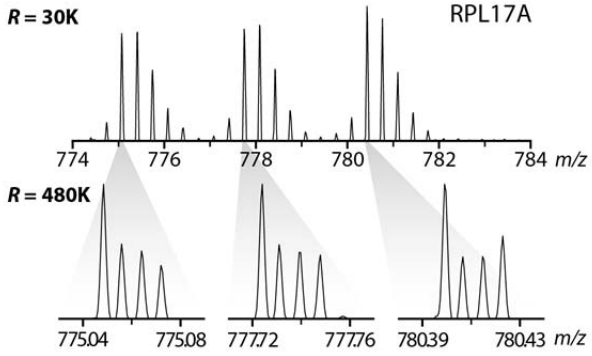
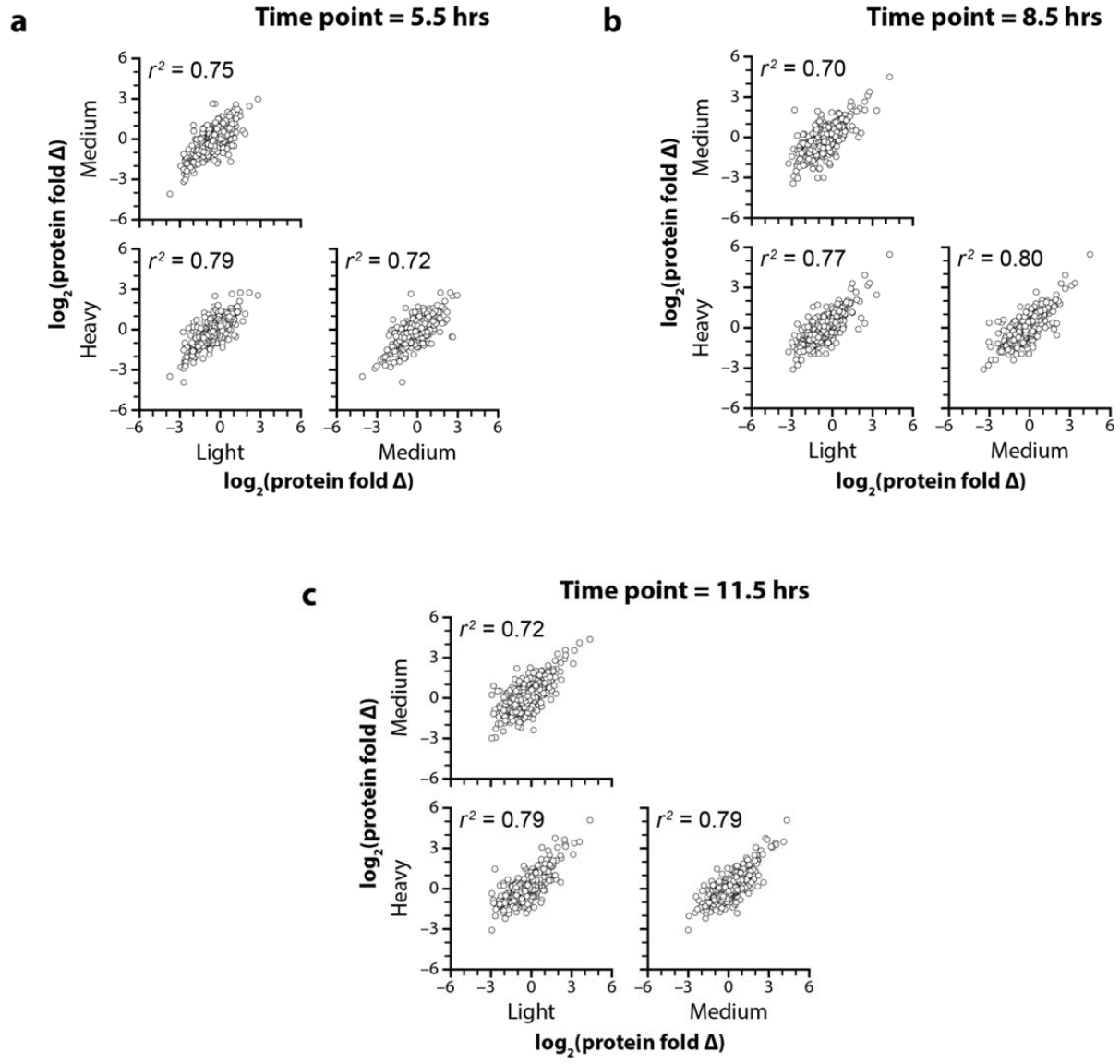


Figure 6 | NeuCode 12-plex multi-scatter plots. Multi-scatter plots for each time point compared to the 3.3 hr time point. The fold change for each protein with each nominal mass light (+0 Da), medium (+4 Da), and heavy (+8 Da) label is plotted against the fold change for the same protein, measured using a tag with the other two nominal masses. **a**, 5.5 hrs. **b**, 8.5 hrs. **c**, 11.5 hrs. demonstrates that the NeuCode tagging method provides excellent reproducibility between biological replicates, as indicated by r^2 values ranging from 0.70 – 0.80.



Discussion

Here we describe a new approach to protein quantification, one that combines the MS1-based quantification of SILAC with the plexing and tissue compatibility of isobaric labeling. The method stems from our recent description of neutron encoded signatures which compress quantitative information into a single isotopic cluster. That work, however, used amino acids and metabolic labeling (NeuCode SILAC) to insert the mDa spaced labels. Here we expand upon this concept by design, synthesis, and analysis of NeuCode chemical reagents, permitting NeuCode quantitative analysis of virtually any proteomic sample, including human tissues and fluids. Using amino acid building blocks and standard Fmoc coupling chemistry, we describe a strategy that delivers 12-plex MS1-based quantitation with spectral complexity that is commensurate with triplex SILAC. The specific label structure presented in this work is intended solely as a near-term demonstration for the overall concept of NeuCode labeling and is not intended as a long-term strategy. Using both quantitative models and a biological system of yeast undergoing the diauxic shift, we demonstrate strong performance across all quantitative Figure of Merit. Briefly, quantitative accuracy, as determined by the fixed ratio experiment, revealed NeuCode tags provided median measured peptide ratios within ~ 20% of known values. Measured standard deviations, *i.e.*, precision, ranged from 0.50 to 0.70 and r^2 correlation of biological replicate measurements (12-plex data) ranged from 0.7 to 0.8. Isobaric tagging, the only method offering comparable plexing capacity, requires some sort of gas-phase purification to achieve comparable quantitative accuracy, but even then, offers, at best, equivalent precision compared to these NeuCode tags. As expected, since both are MS1-based quantitative approaches, traditional SILAC and NeuCode tagging offer comparable precision (standard deviations of 0.4 vs. 0.5 for a 2:1 mixture, SILAC vs. NeuCode tags, respectively) and accuracy. Although beyond the scope of the current work, NeuCode tags should perform comparably for PTM quantitation.

Besides providing comparable quantitative accuracy and precision, as compared to SILAC and isobaric tagging, we conclude our guiding supposition that the NeuCode strategy is indeed compatible

with chemical labeling and that it promises to deliver highly plexed quantitation in a MS1 accessible format. The current tag structure, however, does reduce the overall number of peptide identifications, as compared to peptides having no tag. This result is primarily a function of the tag itself and not of the NeuCode data acquisition method. Supporting this conclusion, no significant loss in peptide identifications was incurred using the same data acquisition strategy for our NeuCode SILAC work[15]. Further, approximately 5% of tandem mass spectra bearing the NeuCode tags were mapped to sequence compared to about ~30% for unlabeled peptides. Based on the structure of the NeuCode tag, this reduction in sequencing performance is not entirely unexpected. First, the tag adds five additional amide bonds to each primary amine. Fragmentation of any of these generates sequence uninformative product ions. Second, the label is relatively large; ~ 435 Da is added to each primary amine – *i.e.*, one to two per peptide. Third, Arg is the most basic amino acid and may inhibit random backbone fragmentation by sequestering protons[44,45]. The reduction in proteomic coverage observed in the 12-plex experiments, as compared to the 4-plex data, is likely caused by increased spectral complexity; that is, three isotopic clusters are observed for every unique peptide sequence so that lower abundance species are less often selected for MS/MS[43,46]. This exact phenomenon is observed during duplex, and especially, triplex SILAC.

Altogether, this study demonstrates that NeuCode chemical reagents will provide a direct route to very highly plexed MS1-based quantification, in a format (MS1) that offers improved accuracy over the current state-of-the-art isobaric tagging chemistry. We posit that future tag architecture should move away from amino acid building blocks and toward small molecule organic synthesis with tunable ^{15}N and ^{13}C incorporation. Such a strategy will allow for maximal ^{15}N incorporation, so as to increase the plexing capacity, while reducing or even eliminating the number of amide linkages of the current tag. And, as routine MS resolving power improves to 10^6, the required NeuCode spacing of 12.6 mDa will be reduced to 6.3 mDa, doubling the plexing capacity for a given tag system.

Experimental Procedures

Theoretical calculations

First**,** a library of 65,011 yeast trypsin derived peptides identified by mass spectrometry was composed. The theoretical full width at 10% max peak height (*FWTM*) for each library peptide across resolving powers (*R*) from 15 thousand to 1 million is calculated by:

$$FWTM = 1.82262 \times \frac{m/z}{R \times \sqrt{\frac{400}{m/z}}}$$

where resolving power is defined as the minimum *m/z* difference that can be resolved at 400 m/z and the coefficient (1.82262) is derived from Gaussian peak shape modeling. The *m/z* difference (Δ*m/z*) for each theoretical isotope doublet assuming isotopologue mass differences (Δ*I*) of 6.3, 12.6, and 25.2 mDa:

$$\Delta m/z = \frac{n \times \Delta I}{z}$$

where *n* is the number of primary amines (lysines and N-terminii) in the peptide sequence and *z* is the charge of the peptide. An isotopologue pair is only considered resolvable at the tested isotopologue mass difference and resolving power if $\Delta m/z > FWTM$.

Label synthesis

NeuCode labels were synthesized on pre-loaded Gly-Wang resin (EMD Millipore) by typical Fmoc solid phase synthesis. Mass defects were incorporated by generating tri-peptides with isotopic distributions described in Figure 1c using protected amino acids (EMD Millipore) and protected amino acid isotopes (Cambridge Isotopes). The N-terminii and lysine side chains were reacted with either $^{13}C_4$ acetic anhydride (Cambridge Isotopes) or light acetic anhydride (Sigma Aldrich). NHS esters were formed by reaction of the tri-peptide in DMSO with 10 equivalents of diisopropylcarbodiimide, 1.5 equivalents of N-hydroxysuccinimide, and 10 equivalents of pyridine. NHS esters were precipitated by drop-wise addition to ice-cold ether. Precipitate was dissolved in DMSO.

Yeast cultures and Lysis

The wild-type W303 *Saccharomyces cerevisae* strain was obtained from the laboratory of Jarred Rutter. In biological triplicate, a single colony of yeast was used to inoculate a starter culture of YPD media and incubated (30 °C, 230 rpm) for 12 h. YPD media (1 L) was inoculated with yeast cells from the starter culture, incubated (30 °C, 230 rpm). Yeast cell pellets were isolated at four points during the culture (1–4, Figure 3a) by centrifugation and immediately frozen in $N_{2(l)}$. The concentration of glucose in the media was determined with a Glucose (HK) Assay Kit (Sigma). The entirety of the following yeast lysis procedure was performed at 4 °C. SEH lysis buffer (0.6 M sorbitol, 20 mM HEPES, pH 7.4, 2 mM $MgCl_2$, 1 mM EGTA, protease inhibitors) and glass beads (BioSpec) were added to the yeast cell pellet. The cells were lysed by bead beating. The beads were allowed to settle to the bottom of the tube, and the supernatant (lysate) was frozen in $N_{2(l)}$ and stored at -80 °C.

nanoHPLC-MS/MS analysis

The samples were reduced and alkylated, digested overnight with trypsin (Promega), desalted and dried. Peptides were re-suspended in DMSO and TEA. Labels were added to individual samples, incubated at ambient temperature for 2 hours and combined. The 4- and 12-plex samples were fractionated by high pH reversed phase (pH 10). Each sample was analyzed by nano-reversed phase liquid chromatography coupled to and Orbitrap Elite mass spectrometer (Thermo Scientific). The MS acquisition method is described in the Results section.

Data analysis

Raw files were converted to searchable text files using COMPASS 1.2.2.3. Data was searched using semi-tryptic enzyme specificity against a target-decoy database containing 13,306 entries (*Saccharomyces cerevisiae,* downloaded from UniProt January 20, 2013) using the Open Source Mass

Spectrometry Search Algorithm (OMSSA 2.1.8), allowing for up to three missed cleavages. For all samples, methionine oxidation and cysteine carbamidomethylation were searched as variable and fixed modifications, respectively. 4-plex and known ratio samples were searched with the average 4-plex NeuCode label mass (+431.2399) as a fixed modification on peptide N-terminii and lysines and as a variable modification on tyrosines. 12-plex was searched exactly like the 4-plex except the data was searched independently with for NeuCode light (+431.2399), medium (+435.2533), and heavy (+439.2667). Survey scan mass tolerances were set to 3.5 Da. Fragment mass tolerances were 0.1 Da for Orbitrap HCD spectra and 0.35 for ion-trap CAD spectra. Little difference in the number of identifications was observed between spectra cleared of tag ions and those not, this step was not included in the analyses. Search results were filtered to 1% FDR based on E-values and mass error[47]. The maximum mass error for all identified peptides was 15 ppm. Peptides were grouped into proteins and filtered to 1% FDR according to rules previously described[48]. Tagging efficiency was calculated from peptide identification counts. The number of labels on primary amines was divided by the total number of amines.

Quantification

The general algorithm employed for quantification of NeuCode peptide isotopologues is described elsewhere[15]. This approach was modified in the following ways to accommodate sets of four isotopologues: sets of peaks were extracted from an *m/z* range extending from 10 ppm below the adjusted *m/z* for the lightest isotopologue to 10 ppm above the adjusted m/z for the heaviest isotopologue; to be eligible for quantification, a set must contain at least two peaks that are properly spaced (*i.e.*, within the spacing tolerance of the theoretical spacing ± 3.5 mDa); a noise-based value was added to missing isotopologue channels for peak sets meeting the abovementioned requirements. This algorithm was applied independently to isotopologue sets in different isotope clusters (*e.g.*, for the 12-plex experiments). For protein quantification, the intensities for each isotopologue channel were summed for all a protein's

observed peptides and the resulting protein ratios were normalized such that the distribution median was centered at a ratio of 1. Resolution was not used as criteria for acceptance of quantitative data. Known contaminants were not quantified.

Acknowledgements

This work was supported by the National Institutes of Health grant R01 GM080148 to J.J.C. A.E.M. gratefully acknowledges support from a National Institutes of Health-funded Genomic Sciences Training Program (5T32HG002760). D.J.P was supported by a Searle Scholars Award and Shaw Scientist award. J.A.S. was supported by a NIH Ruth L. Kirschstein National Research Service Award (1F30AG043282).

References

1 Grimsrud, P. A., Swaney, D. L., Wenger, C. D., Beauchene, N. A. & Coon, J. J. (2010) Phosphoproteomics for the Masses. *ACS Chemical Biology* **5**, 105-119, doi:10.1021/cb900277e

2 Altelaar, A. F. M., Munoz, J. & Heck, A. J. R. (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* **14**, 35-48

3 Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences* **96**, 6591-6596, doi:10.1073/pnas.96.12.6591

4 Jiang, H. & English, A. M. (2002) Quantitative Analysis of the Yeast Proteome by Incorporation of Isotopically Labeled Leucine. *Journal of Proteome Research* **1**, 345-350, doi:10.1021/pr025523f

5 Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics* **1**, 376-386, doi:10.1074/mcp.M200025-MCP200

6 Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T. & Hamon, C. (2003) Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry* **75**, 1895-1904, doi:10.1021/ac0262560

7 Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. (2004) Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics* **3**, 1154-1169, doi:10.1074/mcp.M400129-MCP200

8 Krüger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C. A., Forner, F., Schmidt, S., Zanivan, S., Fässler, R. & Mann, M. (2008) SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell* **134**, 353-364

9 Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Meth* **7**, 383-385, doi:http://www.nature.com/nmeth/journal/v7/n5/suppinfo/nmeth.1446_S1.html

10 Choe, L., D'Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P. & Lee, K. H. (2007) 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *PROTEOMICS* **7**, 3651-3660, doi:10.1002/pmic.200700316

11 Gouw, J. W., Krijgsveld, J. & Heck, A. J. R. (2010) Quantitative Proteomics by Metabolic Labeling of Model Organisms. *Molecular & Cellular Proteomics* **9**, 11-24, doi:10.1074/mcp.R900001-MCP200

12 Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I. & Wright, P. C. (2009) iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly". *Journal of Proteome Research* **8**, 5347-5355, doi:10.1021/pr900634c

13 Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. & Coon, J. J. (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Meth* **8**, 933-935, doi:http://www.nature.com/nmeth/journal/v8/n11/abs/nmeth.1716.html#supplementary-information

14 Ting, L., Rad, R., Gygi, S. P. & Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Meth* **8**, 937-940, doi:http://www.nature.com/nmeth/journal/v8/n11/abs/nmeth.1714.html#supplementary-information

15 Hebert, A. S., Merrill, A. E., Bailey, D. J., Still, A. J., Westphall, M. S., Strieter, E. R., Pagliarini, D. J. & Coon, J. J. (2013) Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Meth* **advance online publication**, doi:http://www.nature.com/nmeth/journal/vaop/ncurrent/abs/nmeth.2378.html#supplementary-information

16 Atwood, J. A., Cheng, L., Alvarez-Manilla, G., Warren, N. L., York, W. S. & Orlando, R. (2007) Quantitation by Isobaric Labeling: Applications to Glycomics. *Journal of Proteome Research* **7**, 367-374, doi:10.1021/pr070476i

17 Sleno, L. (2012) The use of mass defect in modern mass spectrometry. *Journal of Mass Spectrometry* **47**, 226-236, doi:10.1002/jms.2953

18 Werner, T., Becher, I., Sweetman, G., Doce, C., Savitski, M. M. & Bantscheff, M. (2012) High-Resolution Enabled TMT 8-plexing. *Analytical Chemistry* **84**, 7188-7194, doi:10.1021/ac301553x

19 McAlister, G. C., Huttlin, E. L., Haas, W., Ting, L., Jedrychowski, M. P., Rogers, J. C., Kuhn, K., Pike, I., Grothe, R. A., Blethrow, J. D. & Gygi, S. P. (2012) Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Analytical Chemistry* **84**, 7469-7478, doi:10.1021/ac301572t

20 Nagaraj, N., Alexander Kulak, N., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O. & Mann, M. (2012) System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Molecular & Cellular Proteomics* **11**, doi:10.1074/mcp.M111.013722

21 Molina, H., Yang, Y., Ruch, T., Kim, J.-W., Mortensen, P., Otto, T., Nalli, A., Tang, Q.-Q., Lane, M. D., Chaerkady, R. & Pandey, A. (2008) Temporal Profiling of the Adipocyte Proteome during Differentiation Using a Five-Plex SILAC Based Strategy. *Journal of Proteome Research* **8**, 48-58, doi:10.1021/pr800650r

22 Ong, S.-E., Foster, L. J. & Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods* **29**, 124-130, doi:http://dx.doi.org/10.1016/S1046-2023(02)00303-1

23 Makarov, A., Denisov, E. & Lange, O. (2009) Performance Evaluation of a High-field Orbitrap Mass Analyzer. *Journal of the American Society for Mass Spectrometry* **20**, 1391-1396, doi:http://dx.doi.org/10.1016/j.jasms.2009.01.005

24 Denisov, E., Damoc, E., Lange, O. & Makarov, A. (2012) Orbitrap mass spectrometry with resolving powers above 1,000,000. *International Journal of Mass Spectrometry* **325–327**, 80-85, doi:http://dx.doi.org/10.1016/j.ijms.2012.06.009

25 Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Müller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J.-J., Cox, J., Horning, S., Mann, M. & Makarov, A. (2012) Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Molecular & Cellular Proteomics* **11**, doi:10.1074/mcp.O111.013698

26 Schaub, T. M., Hendrickson, C. L., Horning, S., Quinn, J. P., Senko, M. W. & Marshall, A. G. (2008) High-Performance Mass Spectrometry: Fourier Transform Ion Cyclotron Resonance at 14.5 Tesla. *Analytical Chemistry* **80**, 3985-3990, doi:10.1021/ac800386h

27 Marshall, A. G. & Hendrickson, C. L. (2008) High-Resolution Mass Spectrometers. *Annual Review of Analytical Chemistry* **1**, 579-599, doi:doi:10.1146/annurev.anchem.1.031207.112945

28 Pichler, P., Köcher, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G. & Mechtler, K. (2010) Peptide Labeling with Isobaric Tags Yields Higher Identification Rates Using iTRAQ 4-

Plex Compared to TMT 6-Plex and iTRAQ 8-Plex on LTQ Orbitrap. *Analytical Chemistry* **82**, 6549-6558, doi:10.1021/ac100890k

29 Merrifield, R. B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85**, 2149-2154, doi:10.1021/ja00897a025

30 Sieber, P. (1987) A new acid-labile anchor group for the solid-phase synthesis of C-terminal peptide amides by the Fmoc method. *Tetrahedron Letters* **28**, 2107-2110, doi:http://dx.doi.org/10.1016/S0040-4039(00)96055-6

31 McAlister, G. C., Phanstiel, D., Wenger, C. D., Lee, M. V. & Coon, J. J. (2009) Analysis of Tandem Mass Spectra by FTMS for Improved Large-Scale Proteomics with Superior Protein Quantification. *Analytical Chemistry* **82**, 316-322, doi:10.1021/ac902005s

32 Altelaar, A. F. M., Frese, C. K., Preisinger, C., Hennrich, M. L., Schram, A. W., Timmers, H. T. M., Heck, A. J. R. & Mohammed, S. Benchmarking stable isotope labeling based quantitative proteomics. *Journal of Proteomics*, doi:http://dx.doi.org/10.1016/j.jprot.2012.10.009

33 Bakalarski, C. E., Elias, J. E., Villén, J., Haas, W., Gerber, S. A., Everley, P. A. & Gygi, S. P. (2008) The Impact of Peptide Abundance and Dynamic Range on Stable-Isotope-Based Quantitative Proteomic Analyses. *Journal of Proteome Research* **7**, 4756-4765, doi:10.1021/pr800333e

34 DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997) Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* **278**, 680-686, doi:10.1126/science.278.5338.680

35 Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D. & Brown, P. O. (2000) Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Molecular Biology of the Cell* **11**, 4241-4257

36 Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. (2009) Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. *Cell* **138**, 795-806, doi:http://dx.doi.org/10.1016/j.cell.2009.05.051

37 Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. & Garrels, J. I. (1999) A Sampling of the Yeast Proteome. *Molecular and Cellular Biology* **19**, 7357-7368

38 Haurie, V., Sagliocco, F. & Boucherie, H. (2004) Dissecting regulatory networks by means of two-dimensional gel electrophoresis: Application to the study of the diauxic shift in the yeast Saccharomyces cerevisiae. *PROTEOMICS* **4**, 364-373, doi:10.1002/pmic.200300564

39 Graack, H. R. & Wittmann-Liebold, B. (1998) Mitochondrial ribosomal proteins (MRPs) of yeast. *Biochem. J.* **329**, 433-448

40 Fearon, K. & Mason, T. L. (1992) Structure and function of MRP20 and MRP49, the nuclear genes for two proteins of the 54 S subunit of the yeast mitochondrial ribosome. *Journal of Biological Chemistry* **267**, 5162-5170

41 Storey, J. D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498, doi:10.1111/1467-9868.00346

42 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. (2011) Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Meth* **8**, 821-827, doi:http://www.nature.com/nmeth/journal/v8/n10/abs/nmeth.1699.html#supplementary-information

43 Mertins, P., Udeshi, N. D., Clauser, K. R., Mani, D. R., Patel, J., Ong, S.-e., Jaffe, J. D. & Carr, S. A. (2011) iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. *Molecular & Cellular Proteomics*, doi:10.1074/mcp.M111.014423

44 Dikler, S., Kelly, J. W. & Russell, D. H. (1997) Improving mass spectrometric sequencing of arginine-containing peptides by derivatization with acetylacetone. *Journal of Mass Spectrometry* **32**, 1337-1349, doi:10.1002/(sici)1096-9888(199712)32:12<1337::aid-jms599>3.0.co;2-x

45 Fuchs, S. W., Sachs, C. C., Kegler, C., Nollmann, F. I., Karas, M. & Bode, H. B. (2012) Neutral Loss Fragmentation Pattern Based Screening for Arginine-Rich Natural Products in Xenorhabdus and Photorhabdus. *Analytical Chemistry* **84**, 6948-6955, doi:10.1021/ac300372p

46 Boersema, P. J., Aye, T. T., van Veen, T. A. B., Heck, A. J. R. & Mohammed, S. (2008) Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *PROTEOMICS* **8**, 4624-4632, doi:10.1002/pmic.200800297

47 Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. (2011) COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *PROTEOMICS* **11**, 1064-1074, doi:10.1002/pmic.201000616

48 Nesvizhskii, A. I. & Aebersold, R. (2005) Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Molecular & Cellular Proteomics* **4**, 1419-1440, doi:10.1074/mcp.R500012-MCP200

Chapter 4

Gas-phase purification enables accurate, large-scale, multiplexed proteome quantification

ASH designed experiments, performed experiments, and analyzed data.

This chapter has been published:

Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. & Coon, J. J. (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Meth* **8**, 933-935, doi:http://www.nature.com/nmeth/journal/v8/n11/abs/nmeth.1716.html

Abstract

We describe a mass spectrometry method, QuantMode, which improves the accuracy of isobaric tag–based quantification by alleviating the pervasive problem of precursor interference—co-isolation of impurities—through gas-phase purification. Analysis of a model interference sample with a 10:1 ratio improved from 4.4:1 to 8.5:1, with a minimal loss of spectral identifications (21%). This technique substantially improves the accuracy of multiplexed quantitative proteomics with isobaric tagging.

Introduction

Protein identification technologies have rapidly matured such that constructing catalogs of the thousands of proteins comprised by a cell using mass spectrometry is now relatively straightforward[1]. Knowing how the abundance of these molecules change under various circumstances is not[2]. Stable isotope labeling by amino acids in cell culture (SILAC) provides a means to make binary or ternary comparisons[3,4]. By interlacing these two- or three-way measurements, higher-order comparisons can be obtained[5]. Such experiments are invaluable; however, constructing this type of multi-faceted proteomics study is a time-intensive undertaking.

Isobaric tagging[6,7] is an elegant solution to this problem, allowing relative quantification of up to eight proteomes simultaneously[8,9]. Further, it is compatible with mammalian tissues and biofluids, unlike metabolic approaches. Despite its potential, isobaric tagging has not been widely embraced for large-scale studies[10]—chiefly because of precursor interference. This problem does not exist for SILAC because abundance measurements are obtained from high-resolution survey mass spectra (MS^1). Even for very complex samples having hundreds of co-eluting peptides, high–resolving power mass analyzers can easily distinguish the target from neighboring peaks less than 0.01 Th away.

In the isobaric approach, the target peptide is isolated at much lower resolution (typically 1–3 Th), then dissociated to produce reporter tags. Therefore, the quantitative signal in the reporter region is compiled from every species in the isolation window[11]. For highly complex mixtures, co-isolation of multiple species is the rule, not the exception (*vide infra*). This erodes quantitative accuracy, as measured ratios tend to be compressed toward the median ratio of 1:1, and thus has restricted isobaric tagging to applications with lower sample complexity.

Results

To document the extent of interference, we created a *precursor purity model* by labeling tryptic yeast peptides with the TMT^6-126 isobaric tag. We combined these peptides with tryptic human peptides which had been labeled with the TMT^6-131 tag. By incorporating human peptides as the interference, we effectively model the precursor contamination characteristic of a human proteomic analysis. Following nanoflow liquid chromatography–tandem mass spectrometry (nLC–MS/MS) we examined the reporter *m/z* regions of MS/MS spectra that were uniquely mapped to yeast. On average, only 68% of reporter ion signal originated from the target peptide (**Figure 1a**). Only 3% of MS/MS spectra were from ultrapure (≥99%) precursors.

We next wondered: what effect does this prevalent interference have on isobaric tag–based quantification? To address this question, we created a *quantitative accuracy model* by labeling tryptic yeast peptides with TMT^6 tags and mixing them in ratios of 10:1:3:2:5:1.5. This sample was combined with an equal mass of tryptic human peptides, also labeled with TMT^6 but mixed in ratios of 1:1:1:1:1:1. By examining reporter ion ratios in yeast peptide spectra, we assess the degradation of quantitative accuracy. Because we analyzed this mixture directly with nLC–MS/MS, without prior fractionation, we test our technology on the worst-case interference scenario. Shockingly, the yeast 10:1 ratio was measured at a median of 4.4:1—a 2.3-fold underestimation or a 66% relative error (**Figure 1b**).

One approach to reduce interference is to narrow the MS/MS isolation width so that fewer contaminant ions are present during precursor activation. Application of this concept produced only minor improvements (4.3:1 to 5.4:1 for 3 to 1 Th). Note that isolation is less effective as window widths narrow, resulting in fewer identifications (3,348 to 1,723 yeast PSMs for 3 to 1 Th). Chromatographic pre-fractionation did not reduce ratio truncation relative to an unfractionated analysis (4.3:1 to 4.3:1).

Another strategy is to reject quantitative information from precursors having interference above a specified threshold. For the above dataset, we used post-acquisition filtering (PAF) to remove MS/MS spectra if the precursor's purity was below 75% within the 3 Th MS/MS isolation window of the preceding MS^1 spectrum. This technique (**Figure 1b**, middle boxplot) improves quantification (to 6.2:1),

but comes at the expense of 66% of the data: 3,098 versus 1,068 quantified yeast PSMs. We devised a method, real-time target filtering (RTF), to execute this logic in real time so that only precursors without major interference were sampled. RTF boosts quantifiable PSMs relative to post-acquisition filtering and yields comparable accuracy. However, as the correct 10:1 ratio was not achieved, these data evince that background contaminants are not always detectable.

A fresh approach is to combat interference directly via gas-phase purification: that is, to deconvolve the co-isolated contaminants from the precursor in *m/z* space by manipulating either mass or charge before performing a second isolation. Expedient proton-transfer ion/ion reactions (PTR)[12] efficiently reduce ion charge state and integrate easily into instrument methods. And because ion/ion reactions can be conducted on a variety of instruments, including ion traps, ion trap hybrids (i.e., orbitrap and FT-ICR), and quadrupole–time-of-flight (Q-TOF)[13] mass spectrometers, this approach is broadly accessible. Note that standalone ion trap systems, which are capable of performing beam-type CAD[14], are wholly compatible with the approach presented here—such a combination offers a benchtop solution to multiplexed quantification.

Consider a doubly charged precursor at 500 Th co-isolated with a triply charged contaminant also at 500 Th. Following PTR, the precursor is now positioned at *m/z* 999 (+1), whilst the interfering species is moved to *m/z* 749.5 (+2). Subsequent isolation of *m/z* 999 yields a purified precursor population from which to generate accurate quantification. In essence, this facilitates isolation on the basis of both *m/z* and *z* rather than simply *m/z*. Contaminants having charge identical to the precursor are still spread in *m/z* space; for example, a +2 interference at *m/z* 501 will be transformed to *m/z* 1001 and effectively eliminated from the second isolation window (997.5–1000.5 Th).

We devised a scan function to implement this concept, which we call QuantMode, comprising the following steps: (1) first cation injection (quadrupole linear ion trap, QLT), (2) first cation precursor isolation (QLT), (3) anion injection (QLT), (4) PTR (QLT), (5) charge-reduced precursor isolation (QLT), (6) HCD of the charge-reduced precursor (HCD cell), (7) transfer of HCD products (c-trap), (8)

second cation injection (QLT), (9) second cation precursor isolation (QLT), (10) resonant-excitation collision-activated dissociation (CAD, QLT), (11) transfer CAD products (c-trap), (12) transfer HCD/CAD products (orbitrap), and (13) mass analysis of reporter and sequence ions together (orbitrap). The typical QuantMode scan is ~64% longer than an HCD scan, although markedly shorter than an MS/MS/MS experiment (~115% longer).

The purifying effects of QuantMode are evident in **Figure 2**. Examination of a 3 Th precursor isolation window (*m/z* 595.72, +3) from the preceding MS^1 scan reveals extensive contamination—merely 49% purity. This impurity does not hinder sequence identification upon HCD (peptide sequence RINELTLLVQK, OMSSA expectation value 2×10^{-10}). It does, however, cripple quantitative accuracy: the 10:1 true value is recorded as 2.5:1. A 30 ms PTR step on this impure population efficiently deconvolves the target (~45% for +3→+2). The doubly protonated precursor (893.08 Th) is then isolated from the contaminants (85% purity); note this mass analysis following PTR, while illustrative, is not necessary, as the precursor's charge-reduced *m/z* is easily calculated. This untainted charge-reduced precursor population is then dissociated under HCD conditions favorable to reporter ion generation, the products of which are stashed in the c-trap. Next, sequence-informative products are formed through re-injection of the original triply charged precursor (*m/z* 595.72), isolation, and fragmentation in the QLT. After combination in the c-trap, the mixed ion population is mass analyzed in the orbitrap. The QuantMode scan, like its HCD-only counterpart, yields a high-confidence match to the same peptide (OMSSA expectation value 2×10^{-8}). Yet in stark contrast to the truncated 2.5:1 ratio, QuantMode obtains a 9.9:1 ratio, virtually equivalent to the expected 10:1.

The QuantMode scan affords manifold benefits. First, independent control of dissociation parameters can simultaneously improve quantitative accuracy and spectral identifications. Second, decoupling permits the use of isobaric tagging with dissociation methods otherwise incompatible (e.g., resonant-

excitation CAD, ECD/ETD[15], etc.). Third, beyond decoupling benefits, QuantMode elegantly combines multiple disparate steps into a single scan to improve duty cycle.

To validate the efficacy of this method we re-analyzed the mixed organism models with QuantMode. For the *precursor purity model*, purity improved from 68% to 88% among the same 1,297 precursors. Even more striking is the surge in ultrapure (≥99%) precursors—from 3% to 23%. But does this enhanced purity translate to quantitative accuracy? Yes: applying QuantMode to analyze the *quantitative accuracy model* shifted the measured ratio from 4.4:1 with HCD to 8.5:1, much nearer to the true value of 10:1 (**Figure 1c**). QuantMode incurs a 21% loss (3,098 to 2,459) in identified yeast PSMs; however, all are suitable for quantification.

We have demonstrated that isobaric tagging suffers from systemic loss of quantitative accuracy on account of pervasive and inherent precursor interference. Here we describe a new data acquisition method, QuantMode, which mitigates this problem through gas-phase purification. QuantMode substantially increases quantitative accuracy without severely penalizing identifications. In a non-laborious fashion, this method facilitates the critically important measurement of protein and post-translation modification dynamics, as well as biological replicates for proper statistical treatment. This initial implementation will doubtless evolve to include a repertoire of dissociation methods, enhancing both sequence and reporter ion generation, and to improve duty cycle. We conclude QuantMode will substantially augment multiplexed proteome quantification with isobaric tagging.

Figure 1: Analysis of the precursor purity model and quantitative accuracy model samples with either HCD MS/MS or QuantMode. (**a**) Distribution of precursor purity as determined by examining abundance of reporter tag signal at *m/z* of 126 (yeast) and 131 (human) for yeast-identified sequences using either HCD MS/MS or QuantMode. (**b**) Analysis of quantitative accuracy via HCD MS/MS, HCD MS/MS with postacquisition filtering (PAF), and QuantMode. Shown is the true ratio of 10:1 (dashed red line), and boxplots indicate the median (stripe), 25th–75th percentile (interquartile range, box), 1.5 times the interquartile range (whiskers) and outliers (open circles). The number of quantified yeast PSMs (top) and median ratio (bottom) are given for each method.

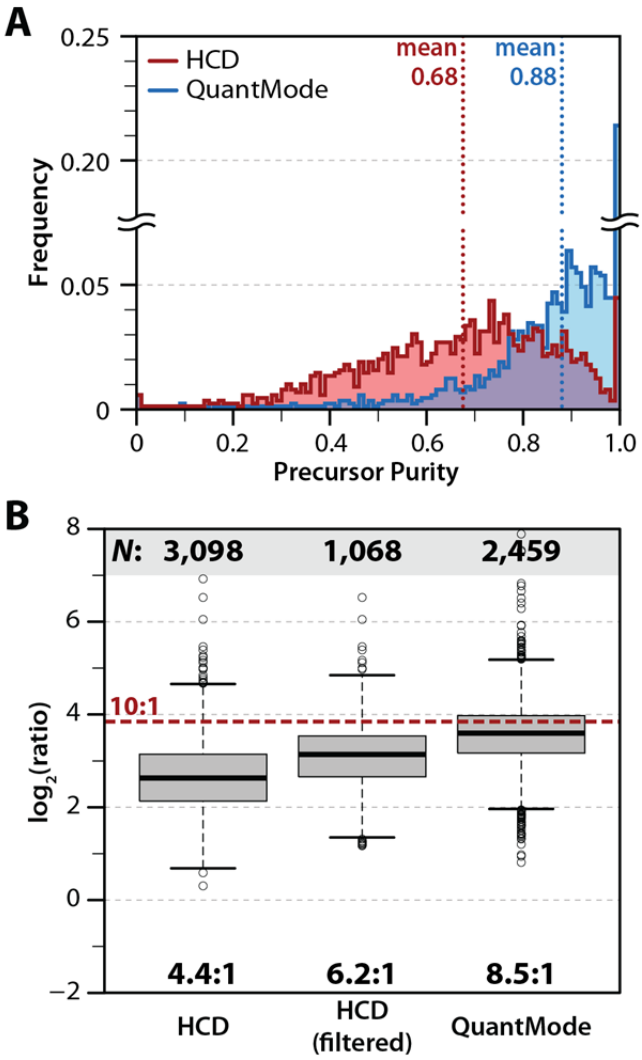
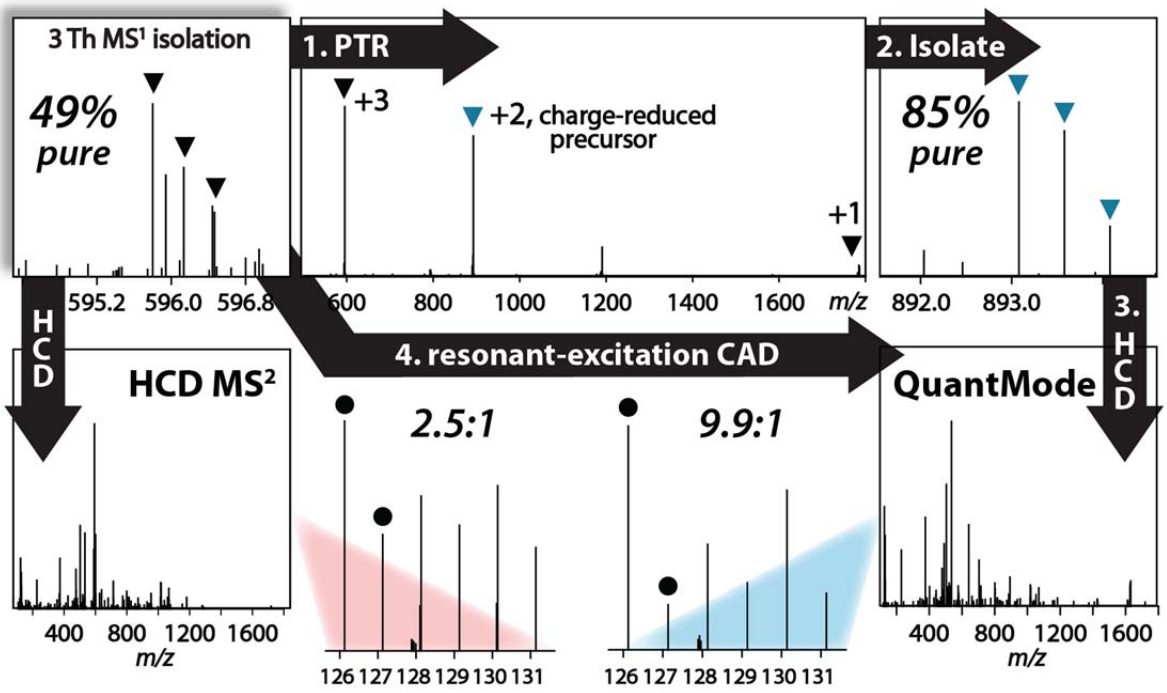


Figure 2: Overview of QuantMode. (**a**) A triply charged precursor at *m/z* 595.72 was isolated with a 3 Th window. The precursor isotopic cluster occupies only 49% of the total ion current in this region. (**b**) QuantMode begins with PTR. (**c**) Isolation of the charge-reduced precursor (+2 charge state) purifies this target to 85%; HCD converts these purified precursors to reporter ions; resonant-excitation CAD follows reinjection and reisolation of the triply charged precursor. The HCD and CAD products are combined in the c-trap before orbitrap mass analysis. (**d**,**e**) The conventional HCD MS/MS scan for this impure precursor (**d**) is juxtaposed against this QuantMode scan (**e**). Insets are close-ups of the gray shaded regions, showing the reporter-ion regions (identical intensity scale) and the quantitative accuracy achieved by both approaches for the 10:1 ratio.



Experimental Procedures

Sample preparation.

Wild-type yeast (*Saccharomyces cerevisiae*) was grown in rich media to an OD_{600} of 0.6. Cells were harvested and centrifuged for 10 min at 4 °C. The resulting cell pellet was washed twice with sterile water and centrifuged at 1,100 × g for 5 min. Lysis buffer of approximately three times the cell pellet volume was added. The lysis buffer contained 8 M urea, 75 mM NaCl, 50 mM tris (pH 8), 10 mM sodium pyrophosphate, 100 mM sodium butyrate, complete mini ETDA-free protease inhibitor (Roche Diagnostics), and phosSTOP phosphatase inhibitor (Roche Diagnostics). The cells were lysed using a French press where the sample was pressed three times and centrifuged for 15 min at 24,000 × g and 4 °C.

Human H1 embryonic stem cells were maintained in a feeder-independent system, as previously described[16]. Cells were harvested by individualizing for 10 min with an adequate volume of pre-warmed (37 ^{o}C) 0.05% Tryp-LE to cover the culture surface. Following cell detachment, an equivalent volume of either ice-cold DPBS (Invitrogen) was added before collecting the cells. Cell pellets were subsequently washed twice in ice-cold PBS and stored at −80 ^{o}C. Cells were lysed via sonication in lysis buffer containing 30 mM NaCl, 50 mM tris (pH 8), 2 mM $MgCl_2$, 50 mM NaF, 1 mM sodium orthovanadate, 6 mM sodium pyrophosphate, mini EDTA-free protease inhibitor (Roche Diagnostics), and phosSTOP phosphatase inhibitor (Roche Diagnostics).

For both yeast and human proteins, cysteine residues were reduced with DTT, alkylated using iodoacetamide, and digested in a two-step process (separately). Proteinase Lys-C (Wako Chemicals) was added at an enzyme:protein ratio of 1:100 and incubated for approximately 4 hr at 37 °C in lysis buffer. Samples were then diluted with 50 mM tris (pH 8.0) to a final urea concentration of 1.5 M and digested with sequencing-grade trypsin (Promega) at an enzyme:protein ratio of 1:50 at 37 °C overnight. Reactions were quenched using trifluoroacetic acid. Samples were desalted using C18 solid-phase extraction

columns (SepPak; Waters) and dried to completion. Yeast and human peptides were then split into six equal mass aliquots. TMT labeling was then performed with each of these aliquots, independently, as described previously[17].

The *precursor purity model* sample comprised equal masses of a yeast aliquot labeled with the TMT^6-126 tag and a human aliquot labeled with the TMT^6-131 tag. The *quantitative accuracy model* sample comprised six yeast and six human aliquots labeled with TMT^6-126−131 tags. The yeast aliquots were mixed in the mass ratios 10:1:3:2:5:1.5, respectively; the human aliquots were mixed in the mass ratios 1:1:1:1:1:1, respectively. These samples were subsequently combined in equal masses. Chromatography fractions of the *quantitative accuracy model* sample were prepared by separation across a Gemini C18 column (Phenomenex) in 20 mM ammonium formate (pH 10) with an acetonitrile gradient. Five 1.5 min fractions were dried down prior to analysis.

Liquid chromatography–mass spectrometry

Online chromatography was performed with a NanoAcquity UPLC system (Waters) coupled to an ETD-enabled LTQ Orbitrap Velos (Thermo Scientific). Samples were loaded onto a precolumn (75 μm i.d., packed with 10 cm, 5 μm Magic C18 particles; Michrom Bioresources) for 10 min at 98:2 0.2% formic acid:acetonitrile with 0.2% formic acid at a flow rate of 1 μL/min. Samples were then eluted over an analytical column at a flow rate of 250 nL/min (50 μm i.d., packed with 25 cm, 5 μm Magic C18 particles; Michrom Bioresources) using a gradient with an initial steep rise to 8% B (acetonitrile with 0.2% formic acid) within 5 min, followed by a linear gradient to 30% B over 120 min and a ramped up to 70% B over 10 min and held for 5 min. The column was equilibrated with 2% B for an additional 25 min. The column-making procedure was previously described[18].

Unless otherwise noted, the instrument method consisted of a data-dependent top-10 experiment with MS^1 resolution of 30,000 (orbitrap) followed by either ten HCD MS/MS or QuantMode scans, mass

analyzed in the orbitrap at 7,500 resolution. All isolation widths were 3 Th. HCD used a normalized collision energy (NCE) setting of 45 for 30 ms. QuantMode employed PTR, the anions for which were generated by the commercial ETD source. For best PTR performance, we employed the nitrogen adduct of fluoranthene at *m/z* 216. We optimized source conditions and all associated ion optics, using the instrument's automated tuning procedures, for this PTR anion.

Following accumulation of the precursor and reagent ions, the two populations were mixed via charge-sign independent trapping, as in an ETD scan. Reaction conditions were tailored to produce the maximal amount of the single charge reduction—e.g., the ion/ion reaction time for a triply charged precursor was set such that the maximum amount of doubly charge product ion was produced. For these experiments the following reaction times were employed: doubly charged precursors were reacted for 80 ms, triply charged precursors were reacted for 30 ms, quadruply charged precursors were reacted for 20 ms, and all precursors with five or more charges were reacted for 10 ms.

Following PTR, the resulting charge-reduced population was isolated and subjected to HCD (NCE 45, 30 ms). Note the charge-reduced species was used for determining HCD energetics, not the initial precursor. These products are then stored in the c-trap. Following HCD, but prior to mass analysis in the orbitrap, a second population of precursor ions (not charge reduced, i.e., impure) was injected into the high-pressure QLT, isolated, and subjected to resonant-excitation CAD (*q*-value 0.25, NCE 35, 10 ms). These products, which are devoid of the reporter region on account of the QLT low-mass cutoff, were mixed with the HCD products in the c-trap and then simultaneously mass analyzed in the orbitrap.

Automatic gain control (AGC) target values were 1,000,000 (MS^1), 50,000 (HCD), 200,000 (QuantMode: PTR), and 10,000 (QuantMode: resonant-excitation CAD). Note more recent data indicates a reduction in QuantMode PTR AGC target to 50,000 results in neither substantial loss of quantitative accuracy nor identification rate. Precursors of unknown or +1 charge state were rejected. Dynamic exclusion was enabled for 60 s after one precursor activation.

For the QuantMode overview experiment (**Figure 2**), we instead used a data-dependent top-3 instrument method consisting of a 30,000-resolution MS^1 scan in the orbitrap followed by (1) normal HCD MS/MS, (2) PTR MS/MS, and (3) HCD MS/MS/MS of the charge-reduced precursor followed by CAD MS/MS of the original precursor, in three separate scans, all mass analyzed at 7,500 resolution in the orbitrap. For the isolation width experiments, standard HCD was used with MS/MS isolation widths from 3 Th to 1 Th in increments of 0.5 Th. For HCD NCE experiments, standard HCD was used with NCEs of 30 to 70 in increments of 5. In this case only the yeast component of the *quantitative accuracy model* sample was used with no human interference.

Data analysis

The resulting data was processed with the COMPASS software suite[19]. OMSSA[20] searches were performed against a target–decoy[21] database containing both yeast (*Saccharomyces* Genome Database[22], http://www.yeastgenome.org, version February 3, 2011, "all" version including all systematically named open reading frames (ORFs), including verified, uncharacterized, and dubious ORFs and pseudogenes) and human (International Protein Index[23], http://www.ebi.ac.uk/IPI, version 3.80) proteins using an average precursor mass tolerance of ±5.0 Da and a monoisotopic product mass tolerance of ±0.01 Da. Carbamidomethylation of cysteines (+57 Da), TMT^6 on peptide N-termini (+229 Da) and TMT^6 on lysines (+229 Da) were specified as fixed modifications. Oxidation of methionine (+16 Da) and TMT^6 on tyrosines (+229 Da) were specified as variable modifications.

All analyses were independently filtered to 1% FDR at the unique peptide level. Peptides that could be derived from human proteins, regardless of enzymatic specificity and treating leucine and isoleucine as equivalent, were removed from consideration so only peptides of yeast origin were evaluated for quantitative analysis.

Acknowledgements

We thank A.J. Bureta for figure illustrations, A. Williams for critical proofreading, A. Ledvina and D. Bailey for assistance with instrument firmware code modifications, S. Hubler for theoretical calculations regarding yeast and human peptides, and J. Brumbaugh and J. Thomson for culturing the human cells. We thank J. Syka, J. Schwartz, V. Zabrouskov, J. Griep-Raming, and D. Nolting (all of Thermo Fisher Scientific) for helpful discussions. This work was supported by the National Institutes of Health grant R01 GM080148 to J.J.C. D.H.P. acknowledges support from an NIH predoctoral traineeship—the Genomic Sciences Training Program, NIH 5T32HG002760.

References

1 de Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F., Walther, T. C. & Mann, M. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251-1255, doi:Doi 10.1038/Nature07341

2 Ong, S. E. & Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252-262, doi:10.1038/nchembio736

3 Jiang, H. & English, A. M. (2002) Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *J. Proteome Res.* **1**, 345-350, doi:Doi 10.1021/Pr025523f

4 Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376-386, doi:DOI 10.1074/mcp.M200025-MCP200

5 Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S. & Mann, M. (2010) Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. *Sci. Signal.* **3**, ra3, doi:ARTN ra3
DOI 10.1126/scisignal.2000475

6 Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T. & Hamon, C. (2003) Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895-1904, doi:Doi 10.1021/Ac0262560

7 Ross, P. L., Huang, Y. L. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces

cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154-1169, doi:DOI 10.1074/mcp.M400129-MCP200

8 Choe, L., D'Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P. & Lee, K. H. (2007) 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* **7**, 3651-3660

9 Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R. & Sanchez, J. C. (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem* **80**, 2921-2931, doi:10.1021/ac702422x

10 Lu, R., Markowetz, F., Unwin, R. D., Leek, J. T., Airoldi, E. M., MacArthur, B. D., Lachmann, A., Rozov, R., Ma'ayan, A., Boyer, L. A., Troyanskaya, O. G., Whetton, A. D. & Lemischka, I. R. (2009) Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* **462**, 358-U126, doi:Doi 10.1038/Nature08575

11 Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I. & Wright, P. C. (2009) iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly". *J. Proteome Res.* **8**, 5347-5355, doi:Doi 10.1021/Pr900634c

12 Reid, G. E., Shang, H., Hogan, J. M., Lee, G. U. & McLuckey, S. A. (2002) Gas-phase concentration, purification, and identification of whole proteins from complex mixtures. *J Am Chem Soc* **124**, 7353-7362, doi:ja025966k [pii]

13 Liang, X. & McLuckey, S. A. (2007) Transmission mode ion/ion proton transfer reactions in a linear ion trap. *J Am Soc Mass Spectrom* **18**, 882-890, doi:S1044-0305(07)00111-0 [pii]
10.1016/j.jasms.2007.02.001

14 McAlister, G. C., Phanstiel, D. H., Brumbaugh, J., Westphall, M. S. & Coon, J. J. (2011) Higher-energy Collision-activated Dissociation Without a Dedicated Collision Cell. *Mol Cell Proteomics* **10**, O111 009456, doi:O111.009456 [pii]
10.1074/mcp.O111.009456

15 Phanstiel, D., Zhang, Y., Marto, J. A. & Coon, J. J. (2008) Peptide and protein quantification using iTRAQ with electron transfer dissociation. *J Am Soc Mass Spectrom* **19**, 1255-1262, doi:S1044-0305(08)00485-6 [pii]
10.1016/j.jasms.2008.05.023

16 Ludwig, T. E., Bergendahl, V., Levenstein, M. E., Yu, J. Y., Probasco, M. D. & Thomson, J. A. (2006) Feeder-independent culture of human embryonic stem cells. *Nat. Methods* **3**, 637-646, doi:Doi 10.1038/Nmeth902

17 Lee, M. V., Topper, S. E., Hubler, S. L., Hose, J., Wenger, C. D., Coon, J. J. & Gasch, A. P. (2011) A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol Syst Biol* **7**, 514, doi:10.1038/msb.2011.48
msb201148 [pii]

18 Martin, S. E., Shabanowitz, J., Hunt, D. F. & Marto, J. A. (2000) Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **72**, 4266-4274, doi:Doi 10.1021/Ac000497v

19 Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. (2011) COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* **11**, 1064-1074, doi:DOI 10.1002/pmic.201000616

20 Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X. Y., Shi, W. Y. & Bryant, S. H. (2004) Open mass spectrometry search algorithm. *J. Proteome Res.* **3**, 958-964, doi:Doi 10.1021/Pr0499491

21 Elias, J. E. & Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207-214, doi:Doi 10.1038/Nmeth1019

22 Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A., Dwight, S. S., Hester, E. T., Jia, Y., Juvik, G., Roe, T., Schroeder, M., Weng, S. & Botstein, D. (1998) SGD: Saccharomyces Genome Database. *Nucleic Acids Res* **26**, 73-79, doi:gkb024 [pii]

23 Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E. & Apweiler, R. (2004) The International Protein Index: An integrated database for proteomics experiments. *Proteomics* **4**, 1985-1988, doi:DOI 10.1002/pmic.200300721

Chapter 5

Calorie restriction and SIRT3 trigger global reprogramming of the mitochondrial protein acetylome

ASH designed experiments, performed experiments, and analyzed data.

This chapter has been published:

Hebert, Alexander S.*, Dittenhafer-Reed, Kristin E.*, Yu, W., Bailey, Derek J., Selen, Ebru S., Boersma, Melissa D., Carson, Joshua J., Tonelli, M., Balloon, Allison J., Higbee, Alan J., Westphall, Michael S., Pagliarini, David J., Prolla, Tomas A., Assadi-Porter, F., Roy, S., Denu, John M. & Coon, Joshua J. (2013) Calorie Restriction and SIRT3 Trigger Global Reprogramming of the Mitochondrial Protein Acetylome. *Molecular Cell* **49**, 186-199, doi:http://dx.doi.org/10.1016/j.molcel.2012.10.024

*co-first author

Abstract

Calorie restriction (CR) extends lifespan in diverse species. Mitochondria play a key role in CR adaptation, however, the molecular details remain elusive. We developed and applied a quantitative mass spectrometry method to probe the liver mitochondrial acetyl-proteome during CR vs. control diet in mice that were wild-type or lacked the protein deacetylase SIRT3. Quantification of 3,285 acetylation sites – 2,193 from mitochondrial proteins rendered a comprehensive atlas of the acetyl-proteome and enabled global site-specific, relative acetyl occupancy measurements between all four experimental conditions. Bioinformatic and biochemical analyses provided additional support for the effects of specific acetylation on mitochondrial protein function. Our results (1) reveal widespread reprogramming of mitochondrial protein acetylation in response to CR and SIRT3, (2) identify three biochemically distinct classes of acetylation sites, and (3) provide evidence that SIRT3 is a prominent regulator in CR adaptation by coordinately deacetylating proteins involved in diverse pathways of metabolism and mitochondrial maintenance.

Introduction

Calorie restriction (CR) is the only regimen known to extend the lifespan and healthspan in a spectrum of organisms that include yeast, mice, and non-human primates [1-3]. Reducing food consumption of mice 25%-60% without undernutrition extends the lifespan of rodents up to 50% [3] and delays the onset of age-related maladies like cardiovascular disease, cancer, and diabetes in different animal models [1,4]. The positive effects of CR are linked to major metabolic reprogramming towards efficient fuel utilization and a reduction in oxidative damage to macromolecules [5,6]. Mitochondrial dysfunction plays an important role in cellular aging [7] and reduces fuel utilization plasticity. As metabolic centers for fuel utilization and primary producers of cellular reactive oxygen species, mitochondria are poised to act as mediators of reprogramming under CR. However, the molecular basis for global metabolic adaptation induced by CR remains unknown.

Emerging evidence suggest that reversible protein acetylation might be a dynamic protein modification that regulates metabolism through multiple mechanisms. The sirtuin family of NAD^+-dependent protein deacetylases are implicated as possible regulators of the cellular adaption to CR, and directly deacetylate specific metabolic enzymes. In yeast, sirtuins are reported as critical components of the CR dependent extension of lifespan [8], and in multiple organisms sirtuins are linked with improved healthspan and metabolic regulation . A critical link between an anti-aging phenotype associated with CR and sirtuins was recently demonstrated for the mitochondrial Sirtuin3 (SIRT3), which was essential for the prevention of age-related hearing loss in mice under CR [9,10]. In response to CR or prolonged fasting, the expression of SIRT3 is induced considerably [9,11]. SIRT3 regulates the function of several mitochondrial proteins involved in oxidative phosphorylation, fatty acid oxidation, the urea cycle, and the antioxidant response system [10-18]. SIRT3 is postulated to be the main, if only, protein mitochondrial deacetylase, but the extent of its capacity to modulate mitochondrial processes remains unclear.

Protein acetylation has emerged as a major post-translational modification and accumulating evidence indicates that acetylation rivals phosphorylation as a regulatory modification [19-25]. With respect to CR and

the role of SIRT3, a broad and deep understanding of acetylation dynamics remains elusive. In a handful of cases, mass spectrometry (MS) has been applied to study mitochondrial acetylation in response to diet changes or SIRT3 expression on a subset of potential protein substrates [9,11,26-29]; however, to understand the diverse mechanisms of reversible acetylation that modulate mitochondrial metabolism and maintenance, a robust, quantitative approach that maps physiologically relevant changes across the mitochondrial acetylome is required.

Here we describe a quantitative acetyl-proteomic method that combines isobaric tagging for multiplexed quantification, immunoenrichment, two-dimensional chromatography, and high resolution, high accuracy MS. We applied this approach to examine the mitochondrial-wide changes to the acetyl-proteome during CR in mice and concomitantly examined the function of resident protein deacetylase SIRT3 in mediating the CR response. We identified 3,285 acetylation sites, of which 2,193 were localized to mitochondrial proteins, and achieved near complete coverage of the mitochondrial proteome. This quantitative acetylation atlas provides a key resource for understanding mitochondrial regulation and reversible protein acetylation. With these data, we establish that SIRT3 is a major regulator of the mitochondrial acetylome in response to CR and discover new processes controlled by SIRT3.

Results

Metrics

We developed and applied a quantitative acetyl-proteomic method to examine the mitochondrial-wide reprogramming of the acetylome and proteome in response to calorie restriction (CR) and the role of SIRT3 dependent regulation of the mitochondrial acetylome under CR. Our approach (**Figure 1**) utilized four distinct components: (1) isobaric tags for multiplexed quantification, (2) an antibody-based acetylpeptide enrichment technique, (3) two-dimensional chromatography, and (4) high resolution, high mass accuracy MS. We compared liver mitochondrial proteomes and acetyl-proteomes of 12 mice (C57BL/6J) following a CR dietary study. Specifically, 3 WT and 3 *Sirt3*$^{-/-}$ mice, at 2 months of age, had

their caloric intake reduced by 25% (i.e., CR group) whilst 3 WT and 3 *Sirt3*$^{-/-}$ animals were maintained on a control diet (i.e., CD group, **Figure 1**). All mice were euthanized at 12 months of age. Altogether, we analyzed, in biological triplicate, the mouse liver mitochondrial proteomes and acetyl-proteomes across four conditions: WT-CD, WT-CR, *Sirt3*$^{-/-}$-CD, and *Sirt3*$^{-/-}$-CR. From these experiments we detected and quantified 3,285 acetylation sites – the vast majority of which (2,193) were localized to mitochondrial proteins (434). We likewise achieved very deep quantification of the mitochondrial proteome (568 of the 701 confirmed members) allowing for the most comprehensive mitochondrial acetyl-proteome study to date.

To profile acetylome dynamics, as regulated by CR and/or SIRT3, intact mitochondria were purified from mouse livers (vide supra) via differential centrifugation. The purified mitochondrial proteins were subjected to trypsin digestion followed by isotopic labeling (isobaric tag, TMT). Labeled peptides were then combined (**Figure 1**), fractionated by strong cation exchange (SCX) chromatography, and immunopurified to isolate acetylated peptides. A small portion (~5%) was excluded from acetyl-enrichment to obtain protein abundance. All fractions were analyzed using a nanoflow liquid chromatography-mass spectrometry method (nHPLC-MS) on a high resolution and high mass accuracy orbitrap mass spectrometer. Eluting peptide cations were isolated and subjected to tandem mass spectrometry (MS/MS) to obtain both sequence identity and quantitative information in a single scan (**Figure 2**). Employing a recently developed gas-phase ion/ion reaction scheme, QuantMode [30], we collected ~1.5 million MS/MS scans, each of which was processed using a custom in-house developed software suite (COMPASS) [31,32]. Only peptide and protein identifications below 1% false discovery rate (FDR) are reported here; acetylation site assignments were subsequently localized to specific lysines using a modified A-score algorithm[33].

Alterations of protein abundance, either *in vivo* or during preparation, can cause apparent acetylation changes, even when the acetylation site occupancy of a protein remains unaffected. Measuring absolute occupancy of global acetylation sites would require the synthesis of thousands of isotopically labeled

internal standards and is, thus, currently impractical. Instead we implemented an experimental design to generate relative acetylation site occupancy information by quantifying both acetylation and protein abundance. The spectrum shown in **Figure 2A** depicts an MS/MS scan mapped to the peptide EMLVK(ac)LAK, a product of digested ribosome releasing factor (MRRF). Note the inset displays the reporter tag mass-to-charge (*m/z*) region, which indicates strong up-regulation of this site in WT-CR (2-fold), *Sirt3*$^{-/-}$-CD (7.5-fold), and *Sirt3*$^{-/-}$-CR (11-fold) animals. To confirm this difference is a result of increased acetylation, rather than an artifact of MRRF abundance, we examined all detected un-modified MRRF peptides (**Figure 2B**). In total, five MS/MS scans were mapped to 3 unique un-modified peptides from MRRF. Summation of these 5 reporter ion regions provides a protein abundance measurement; combination of these data with acetylation fold-change confirms that the acetylation site occupancy is indeed dynamic (**Figure 2C**). This process was conducted for each of the 3,285 acetylation sites.

From these data we conclude our method is both specific and reproducible. In each of the 3 experimental replicates, approximately 70% of the acetylation sites were from mitochondrial proteins. The average enrichment efficiency observed in this study (46%), i.e., number of acetyl peptides identified divided by the total number of identified peptides, was considerably higher than values typically reported (**Figure 3A**) [23,34]. Improved enrichment efficiency may have resulted from reducing sample complexity through fractionation prior to immunoenrichment or the increased abundance of the acetylation modification in the *Sirt3*$^{-/-}$ mitochondria. **Figure 3B** presents the overlap of acetylation sites detected across the replicates – over half of all acetyl sites were found in every experiment. The inter-experiment overlap of our acetylation identifications is higher than what is commonly observed for phosphorylation, where only 1/3 of all phosphorylation sites would likely be observed in all 3 experiments [35]. The high overlap provides both a basis for strong statistical analysis, as most sites are detected in all 3 biological replicates, and strong confidence in our identifications. From this comprehensive atlas we report that 65% of mitochondrial proteins have at least one identifiable acetylation site (**Figure 3C**); the majority of these proteins have multiple sites of acetylation. This observation was unexpected and goes well beyond the

recent estimates that ~20% of the mitochondrial proteome is acetylated [19]. Site occupancy measurements indicate that of the 1,275 mitochondrial acetylation sites identified in all of the experiments 8%, 32%, and 43% exhibit ≥ 2-fold statistically significant changes ($p < 0.05$, Welch's t-test with Storey correction) in the CR, *Sirt3*$^{-/-}$, and *Sirt3*$^{-/-}$-CR conditions, respectively (**Figure 3D)**. Altogether, our data indicate that the mitochondrial proteome is (1) extensively acetylated and (2) pervasively altered by CR and genetic loss of SIRT3.

Figure 1. Experimental design and mass spectrometry workflow for quantitation of acetylated mitochondrial proteins between four biological conditions. (A) WT and *Sirt3*$^{-/-}$ mouse liver mitochondria were compared in biological triplicate from mice fed either the control or calorie restricted (CR) diet. Numbers correspond to the isobaric mass tag used for that sample. (B) Extracted protein from purified liver mitochondria was trypsin digested followed by labeling with TMT reagents. Labeled peptides were combined and fractionated by strong cation exchange chromatography. Each fraction was enriched for acetyl peptides with a pan anti-acetyl lysine antibody and analyzed via nano-RPLC-MS/MS.

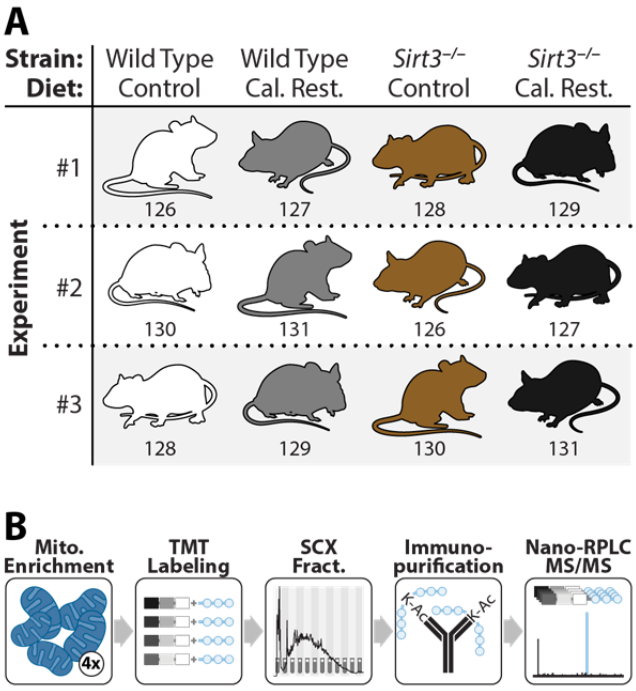


Figure 2. Example acetyl occupancy quantitation (relative), for site K195 on the MRRF protein. The peptides used for calculations are highlighted within the MRRF protein sequence. Blue: Un-modified peptide sequences, spectra with reporter ion magnification and protein fold change for each condition compared to WT-CD condition. Note, the three unique peptides from five total identifications are shown here. Yellow: The MRRF K195 acetylated peptide sequence, spectra, and acetyl fold change for each condition compared to WT-CD. Bottom center: Results for acetyl fold change normalized to protein fold change for each condition compared to WT-CD.

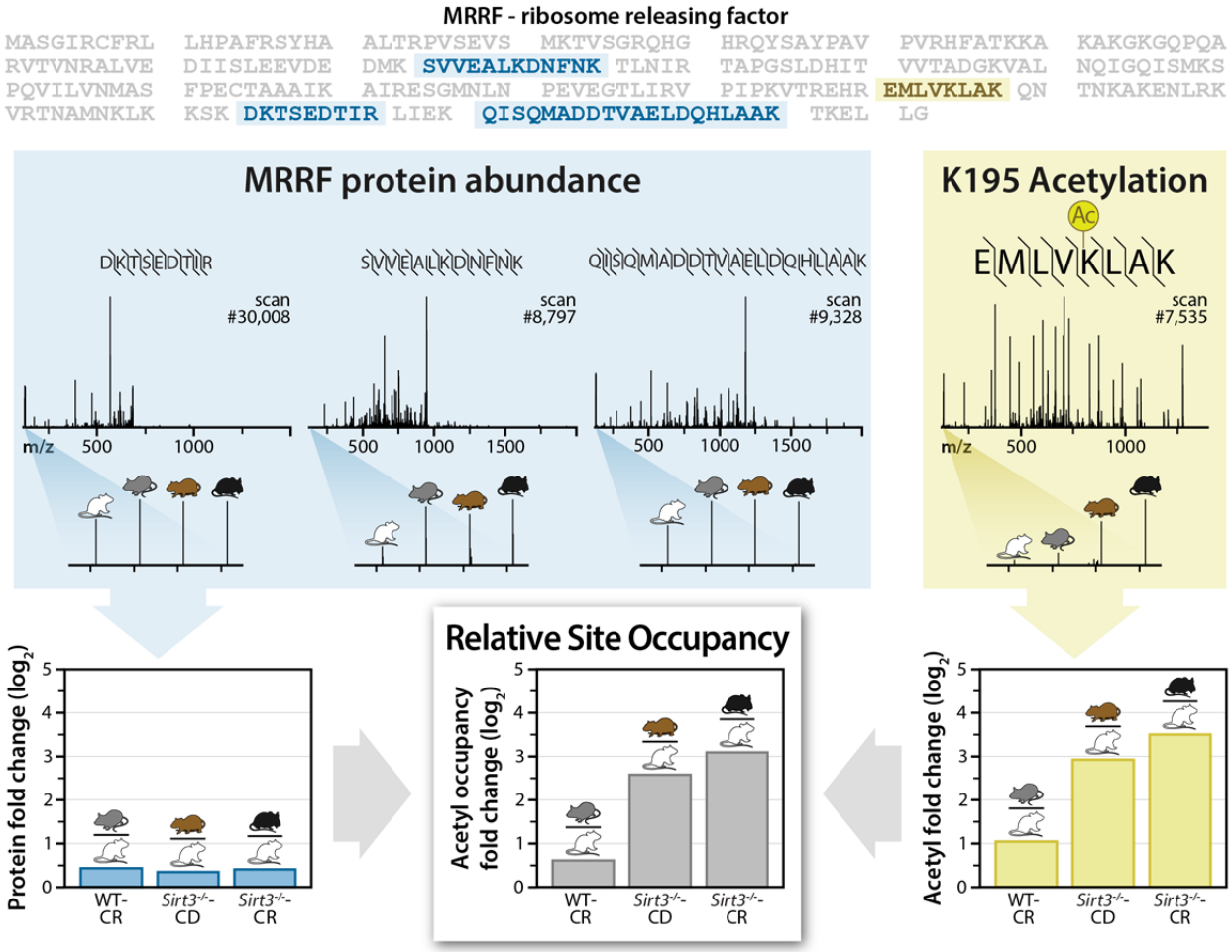
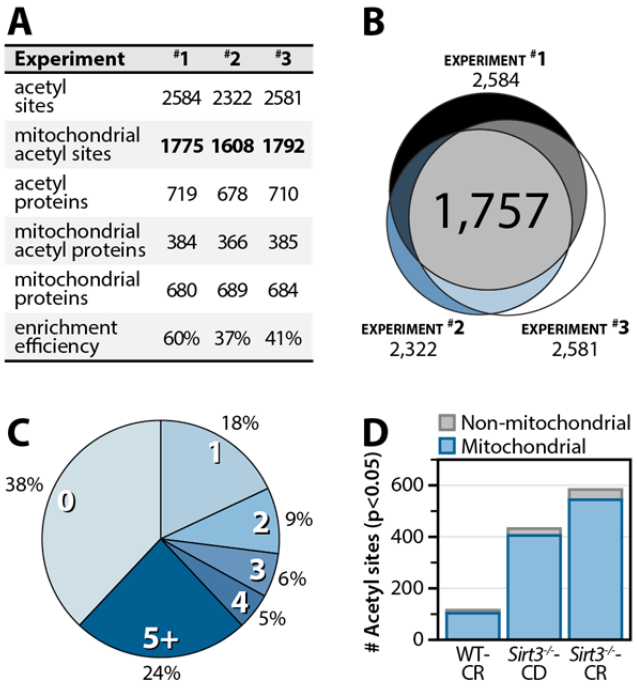


Figure 3. Acetylomic experimental results and metrics. (A) Experimental metrics for each experiment. Enrichment efficiency is calculated as the ratio of acetyl peptide spectral matches to the total number of peptide spectral matches for enriched samples.(B) Venn diagram indicating overlap of the acetyl sites quantified in each experiment. Experimental overlap of acetyl sites identified in three biological replicates yields 1,757 quantifiable acetylation sites. (C) Pie chart of the number of acetylation sites detected per mitochondrial protein. (D) Number of acetyl sites significantly changing ≥ 2-fold for each condition compared to WT control diet ($p < 0.05$, Welch's t-test with Storey Correction).



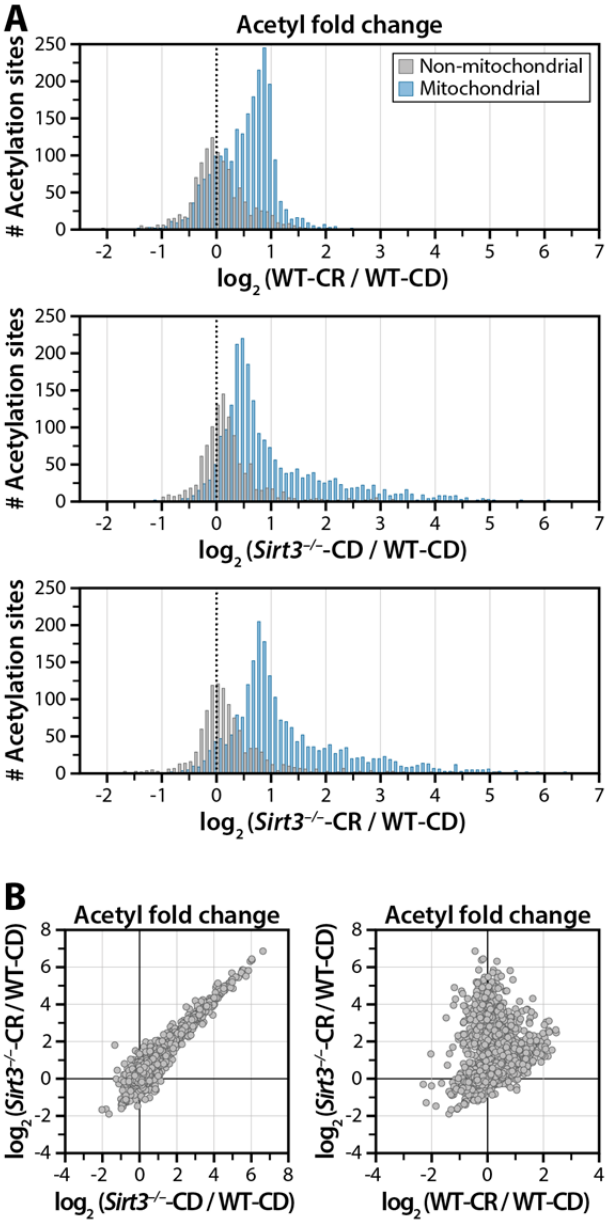
Experiment	#1	#2	#3
acetyl sites	2584	2322	2581
mitochondrial acetyl sites	**1775**	**1608**	**1792**
acetyl proteins	719	678	710
mitochondrial acetyl proteins	384	366	385
mitochondrial proteins	680	689	684
enrichment efficiency	60%	37%	41%

Calorie restriction dramatically alters the mitochondrial protein acetylation landscape

CR dynamically alters mitochondrial protein acetylation, with two notable trends. CR increases acetylation site occupancy of 135 acetyl sites ($p < 0.05$) by >2-fold when compared to CD (**Figure 3D**), while a subset of mitochondrial acetylation sites (>100) display decreased acetylation under CR. The two modes of the distribution in **Figure 4A** reflect both the acetylation sites that decrease under CR and a segregation based upon mitochondrial localization, with acetyl sites localized to the mitochondria exhibiting increased acetylation and the nearly 500 non-mitochondrial acetyl sites showing minimal change. These results indicate CR is a major regulator of protein acetylation, but these changes are predominantly mitochondrial. A major benefit of our approach is the quantitation of total protein in addition to acetyl sites. Compared with acetylation, there is relatively little change in the overall distribution of protein-fold changes across conditions, as only 33 proteins change more than 1.5 fold, supporting a hypothesis that changes in acetylation, but not overall protein levels, drive metabolic reprogramming under CR.

CR induced significant deacetylation of a 108 acetyl sites ($p < 0.05$) by 20% or greater. SIRT3 expression is stimulated under chronic CR [9,15,29] and both deacetylates and regulates the function of several mitochondrial proteins in response to fasting and/or CR, including OTC, LCAD, HMGCS2, and IDH2 [9,11,15,16]. As the only validated deacetylase in mitochondria and because SIRT3 expression is stimulated under CR, we attribute acetyl sites displaying decreased acetylation under CR as SIRT3 substrates. Further, the majority of acetyl sites with decreased acetylation in CR show dramatically increased acetylation (8- to 100-fold, $p < 0.05$) in *Sirt3*$^{-/-}$ (**Figure 4B**). We conclude that SIRT3 is responsible for the striking loss of acetylation in a number of proteins in diverse mitochondrial pathways.

Figure 4. Quantitative distribution of acetyl site data and scatterplot analysis reveals distinct populations. (A) Bimodal distribution of acetyl sites quantified (n=3) segregates between mitochondrial (blue) and non-mitochondrial (gray) proteins. Mitochondrial acetyl sites exhibit increased acetylation, while non-mitochondrial acetyl sites show minimal change. (Top plot) Analysis of acetyl sites for WT-CR vs. WT-CD indicates CR alters mitochondrial protein acetylation. (Middle and lower plots) Analysis of acetyl sites for *Sirt3*$^{-/-}$-CD/CR vs WT-CD indicates that loss of SIRT3 dominates the acetylation content of mitochondrial proteins with increased acetylation apparent in *Sirt3*$^{-/-}$ under both CD and CR. (B) Scatter plot (Top) of acetyl fold change between *Sirt3*$^{-/-}$-CR and *Sirt3*$^{-/-}$-CD, displays near-linear correlation and indicates SIRT3 expression dominates global acetylation status of the mitochondrial proteome. (Lower plot) Scatter plot comparison of the acetyl fold change of *Sirt3*$^{-/-}$-CR to WT-CR exposes unique populations of acetyl sites.



SIRT3 expression is a dominant regulator of the mitochondrial acetylome

We detected a vast number (~400) of mitochondrial acetyl sites increasing (i.e., >2-fold) in the absence of SIRT3 – 120 of these exhibited >8-fold increased acetylation (**Figure 3D**). CR treatment of *Sirt3*$^{-/-}$ mice induces an additional 34% increase in the number of acetyl sites significantly increasing ≥ 2-fold (**Figure 3D**). Regardless of dietary treatment, the overall mitochondrial acetylation landscape is dominated by loss of SIRT3 (*Sirt3*$^{-/-}$) as evinced by the near linear correlation between the acetyl fold change of *Sirt3*$^{-/-}$-CR and *Sirt3*$^{-/-}$-CD (**Figure 4B**).

Importantly, a number of acetyl sites previously identified as SIRT3 substrates were identified, including those in HMGCS2, IDH2, and SDHA. The trends we observed for HMGCS2 [16] are consistent with published results as regulated lysine residues (K327, K333, K473) display low acetylation in CR and considerably increased acetylation in *Sirt3*$^{-/-}$. K413 of IDH2 is a SIRT3 target [10], and here we observed 50% increased acetylation in *Sirt3*$^{-/-}$. Comparing *Sirt3*$^{-/-}$-CR to WT-CD results, a similar analysis revealed numerous proteins exhibiting 10- to 100-fold increases in specific acetylation, while many lysine residues displayed minimal perturbation.

Cluster analysis reveals distinct classes of acetyl isoforms

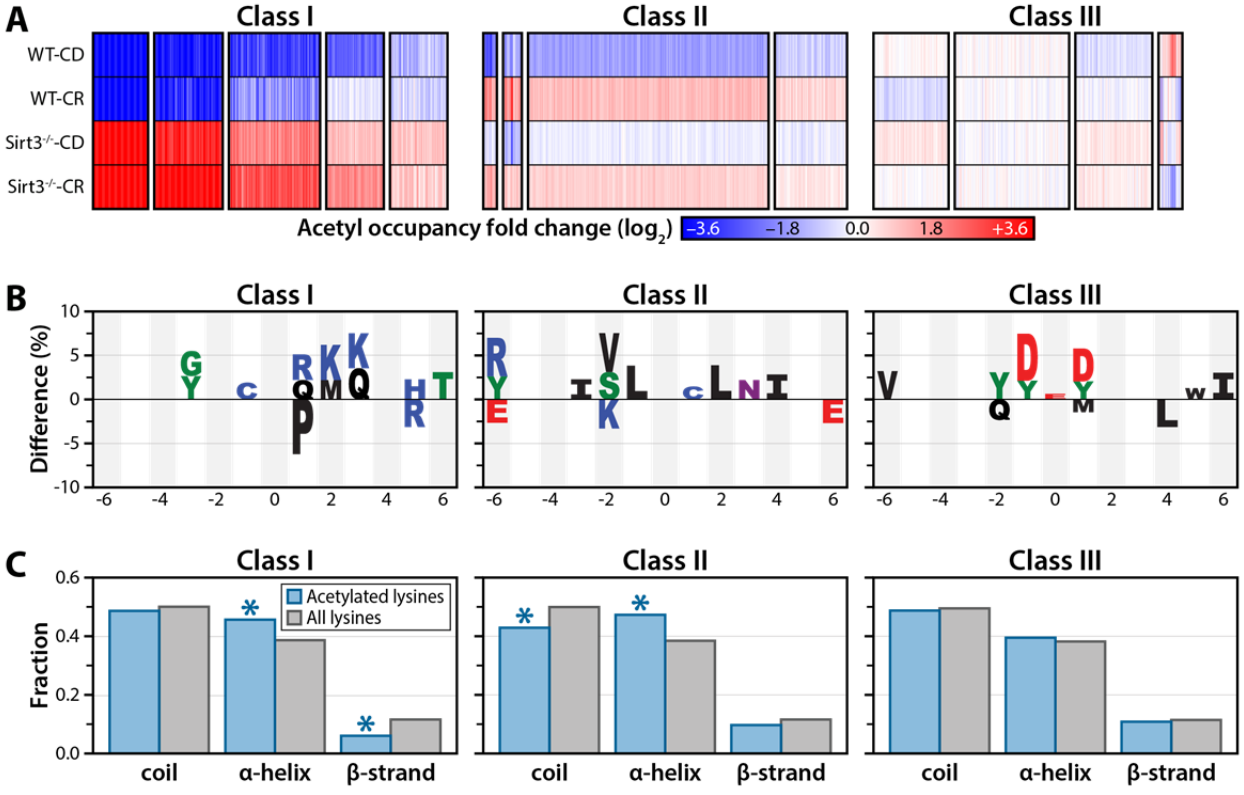
To determine whether the quantifiable mitochondrial acetyl sites could be categorized into subpopulations based on their specific patterns across all four conditions (WT-CD, WT-CR, *Sirt3*$^{-/-}$-CD, *Sirt3*$^{-/-}$-CR), we used a probabilistic clustering algorithm based on a Gaussian mixture model [36]. The clustering algorithm generated 15 clusters (**Figure 5A**) that segregate into three distinct classes: Class 1) acetyl sites that are controlled by SIRT3 protein expression, exhibiting low acetylation under CR when SIRT3 is maximally expressed and hyperacetylation when SIRT3 expression is lost: Class 2) acetyl sites that increase in response to CR: Class 3) sites that display minimal change in response to either dietary regimen or SIRT3 expression.

We postulated that each major acetylation class might comprise unique physical properties. To investigate this we interrogated the sequence and secondary structure of acetyl sites among each class sub-population. Initially, the frequencies of neighboring amino acids were determined for acetyl sites from each class and compared against all acetyl peptides quantified, avoiding false motifs that may arise from antibody bias by immunoenrichment. We observed distinct preferences for amino acids localized near the acetyl lysine for each class (**Figure 5B**). Positively charged residues (i.e., Lys and Arg) near the acetyl lysine are more prevalent in clusters (Class 1) with increased acetylation in the absence of SIRT3. A strong preference for basic peptides is consistent with previous SPOT-library and machine learning screens of SIRT3 substrate specificity [37] and with the structure of SIRT3. The peptide-binding pocket of SIRT3 contains negatively charged residues, particularly C-terminal to the acetyl lysine binding site [38]. In Class 2 (acetyl sites displaying increased acetylation in response to CR), surrounding sequences are largely hydrophobic and uncharged residues. The nature of Class 2 sites and the inability of SIRT3 to induce deacetylation suggests that these acetyl lysines are within hydrophobic regions and inaccessible to SIRT3. In Class 3 (acetyl sites displaying little change in response to dietary treatment or SIRT3 expression), amino acids surrounding acetyl lysine are enriched for negatively charged residues at positions -1 and +1. Given the known in vitro peptide specificity of SIRT3 [37], Class 3 acetyl lysines are expected to be poor SIRT3 substrates.

For each acetyl isoform, the conformational tendencies for α helix, β strand, or coil were determined using PSIPRED [39], a protein structure server used to predict protein secondary structure, against a background set containing all lysines of mitochondrial proteins identified in this study. The local secondary structures allow for interrogation of differentially enriched regions for acetyl lysine in each cluster (**Figure 5C**). Previous global acetyl-proteomics studies indicated that acetyl lysine sites are significantly enriched in α-helical and β-strand regions [19]. Our data reveal distinct secondary structure preferences among different subpopulations of acetyl sites. Class 1 acetyl sites are significantly enriched in □-helical regions ($p < 3.0 \times 10^{-3}$, hyper-geometric test) and depleted in β-strand regions ($p < 2.0 \times 10^{-4}$)

(**Figure 5C**). Acetyl sites increasing in response to CR (Class 2) are enriched in α-helical regions ($p < 2.5 \times 10^{-5}$) and depleted in coils ($p < 8.0 \times 10^{-4}$). Interestingly, Class 3 acetyl sites, which are unchanging across conditions, display no significant enrichment in any secondary structure element.

Figure 5. Cluster analysis resolves acetyl sites into three distinct classes. (A) Cluster analysis identified 15 clusters, with 13 grouped into 3 classes based on their overall trends between the four biological conditions: **Class I** - Acetylation sites that respond dynamically to SIRT3 expression. **Class II** - Acetylation sites that are largely independent of SIRT3 expression but respond to CR. **Class III** - Acetylation sites that display minimal perturbation in response to either loss of SIRT3 expression or CR. Each column represents a quantified acetyl site. Blue: decreased acetyl occupancy, White: no change, Red: increased acetyl occupancy. (B) Amino acid motifs identified for clusters displaying the largest magnitude fold change for each class (from A). The motif shows significant ($p < 0.05$) under or over representation of a particular amino acid at each site flanking the acetylated lysine (position 0). (C) Predicted protein secondary structures of acetylation sites identified in each class. Probabilities for coil, α-helix and β-strands were compared with the probabilities of these secondary structure elements in all lysines identified in the study. (Class I α -helix: $p < 3.0 \times 10^{-3}$; Class I β -strand: $p < 2.0 \times 10^{-4}$; Class II coil: $p < 8.0 \times 10^{-4}$; Class II α -helix: $p < 2.5 \times 10^{-5}$, hyper-geometric analysis).



SIRT3 coordinately deacetylates multiple targets in mitochondrial metabolic pathways

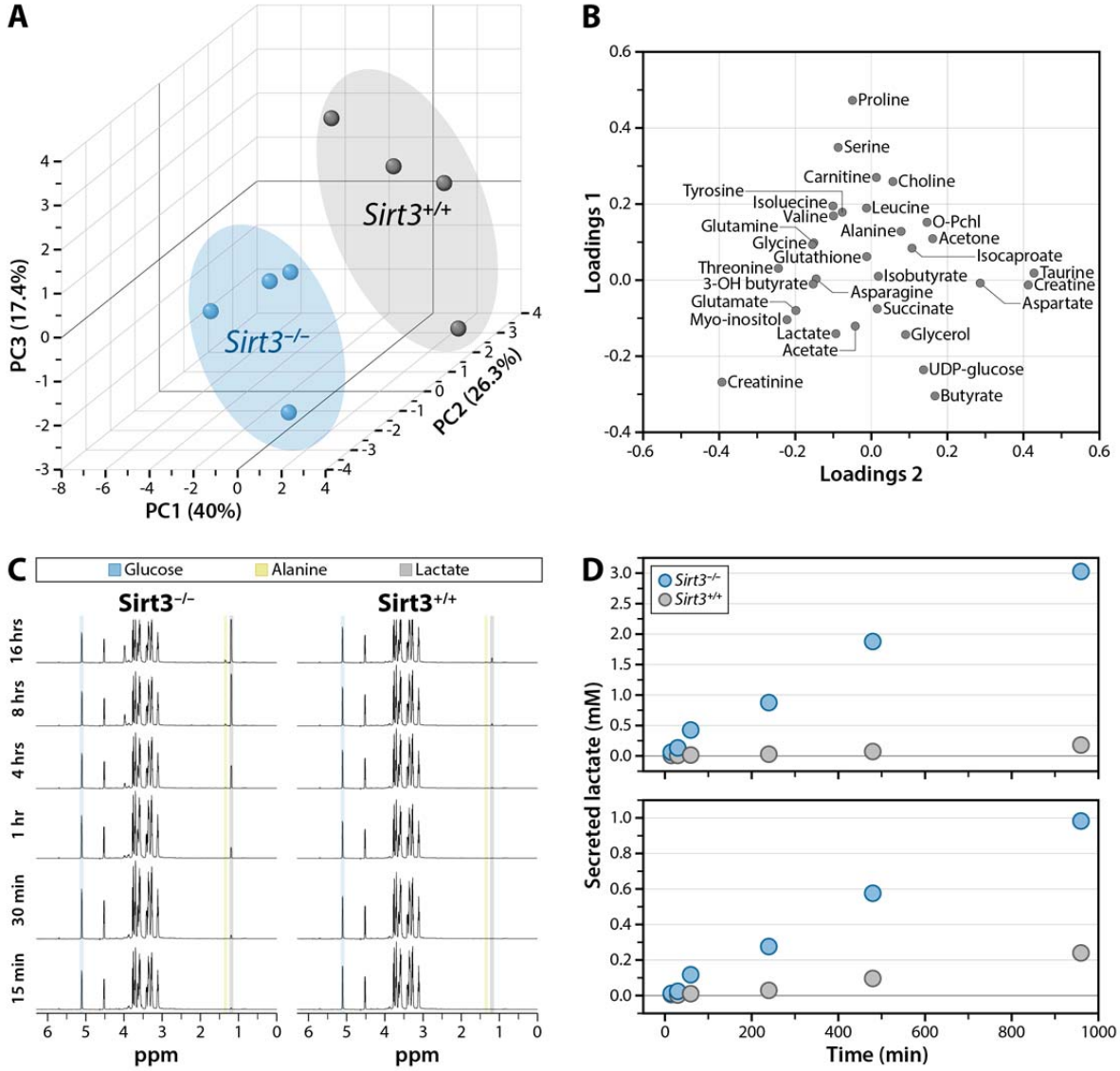
Dysregulation of mitochondrial energy metabolism is a hallmark of aging and disease; however, CR apposes these trends and leads to increased metabolic gene expression and changes in metabolism that protect cells and tissues from aging. The reprogramming of metabolism towards more efficient fuel utilization and energy production requires coordinate regulation of a number of mitochondrial metabolic pathways. Cluster (**Figure 5**) and scatterplot (**Figure 4B**) analysis identified a large pool of regulated sites dominated by SIRT3. To identify affected pathways and understand the biological functions of these acetyl sites responding to CR and *Sirt3*$^{-/-}$, we performed a network analysis using functional association data [40]. Proteins were grouped according to their involvement in biological processes using Gene Ontology (GO) annotations and networks were visualized in Cytoscape [41]. Significant enrichment of major mitochondrial processes was apparent, including fatty acid metabolism, electron transport / ATP production, acetyl-CoA metabolism, amino acid catabolism, and mitochondrial integrity. Strikingly, these results reveal two major features: (1) SIRT3 regulates the acetylation status of multiple proteins within a given pathway, and (2) SIRT3 regulates pathways not previously linked to SIRT3 function. These include branched-chain amino-acid (BCAA) catabolism, one-carbon metabolism, mitochondrial genome maintenance and transcription, and iron homeostasis.

Metabolite analysis of Sirt3$^{-/-}$ MEFs verifies pathways regulated by acetylation

To provide complementary evidence of pathways affected by SIRT3, we profiled metabolite changes resulting from genetic loss of SIRT3. This was accomplished by comparing, over a time course, the metabolic differences in *Sirt3*$^{-/-}$ and *Sirt3*$^{+/+}$ mouse embryonic fibroblasts (MEFs). Although a number metabolites in liver tissue and plasma were determined in previous analyses of *Sirt3*$^{-/-}$ mice [11,15], here the utilization of homogenous MEFs provided a number of benefits: (1) eliminated complexity of entire organism analysis and multi-tissue compensation; (2) allowed for controlled growth conditions; and (3) provided the opportunity to utilize isotopic metabolites and quantitatively monitor flux; ultimately

resulting in clear identification of pathways affected by loss of SIRT3 and subsequent increased protein acetylation. Briefly, MEFs were cultured to 80% confluence and media was replaced with modified minimal medium supplemented with 2 mM [U-^{13}C]-glucose and 10 mM glutamine. Cells and their corresponding conditioned media were collected at 0.25, 0.5, 1, 4, 8 and 16 hours. Lysed cells were subjected to NMR-based metabolomics analysis. A total of 44 cellular metabolites were identified in ^{1}H spectra. To determine the distribution of experimental groups, an unsupervised principal component analysis (PCA) [42,43] was performed with 31 metabolites that were quantifiable in all samples (**Figure 6**). Three principal components (PC) were used in the analysis, capturing 84% of the total variance in the data. PCA analysis resulted in a clear segregation between *Sirt3*$^{-/-}$ and *Sirt3*$^{+/+}$ MEFs (**Figure 6A**), demonstrating these populations are metabolically distinct. The loading plot (**Figure 6B**) identified metabolites that significantly contribute to the biochemical segregation of *Sirt3*$^{-/-}$ and *Sirt3*$^{+/+}$ MEFs. This metabolomic analysis identified perturbations in many metabolic pathways, which likewise were revealed from our SIRT3-dependent acetylome investigation, providing corroborating evidence that SIRT3 regulates the acetylation and activity of diverse proteins in many mitochondrial processes. Below, we highlight these pathways and briefly discuss supporting biochemical, structural and biological evidence.

Figure 6. Functional validation of SIRT3 in mitochondrial metabolism. (A-D) NMR-metabolomics analysis of *Sirt3*$^{+/+}$ and *Sirt3*$^{-/-}$ MEFs. (A) PCA analysis 3-D score plot for the first 3 principal components (PC) with corresponding contribution percentage for each PC. PCA segregates *Sirt3*$^{+/+}$ and *Sirt3*$^{-/-}$ MEFs. (B) Metabolites contributing to the segregation of *Sirt3*$^{+/+}$ and *Sirt3*$^{-/-}$ MEFs are identified in the loading plot for PC1 and PC2. Metabolites localized further from the origin (0,0) have a greater contribution to the segregation of *Sirt3*$^{+/+}$ and *Sirt3*$^{-/-}$. (C) ^{13}C-edited ^{1}H-NMR spectra of metabolites secreted into the media. ^{13}C-Glucose, ^{13}C-lactate and ^{13}C-alanine peaks are identified. (D) Time-dependent increases in ^{13}C-lactate (top plot) and ^{13}C-alanine (bottom plot) concentrations relative to glucose concentration. (E) Acetyl CoA concentration (left) from mouse liver tissue displayed as nmol/g wet weight of tissue and acetyl CoA / CoA ratio (right). Error bars represent percent error calculated from isotopic acetyl CoA standard. (F) MDH2 activity is reduced by K239 acetyl mimic. MDH2-MYC was immunoprecipitated with anti-MYC beads from HEK293 cell lysates and was analyzed by western blotting. Activity was determined by measuring a decrease in absorbance at 340 nm from the oxidation of NADH or NADPH. Rates are calculated relative to WT, error bars represent standard deviation (n = 3, n=2 for Vec), Vec is pcDNA3.1 vector control.



Amino acid catabolism

CR requires metabolic adaption to lowered energy input, commonly resulting in enhanced gluconeogenesis fueled by precursors generated from protein breakdown. In CR mice, the transcription and enzymatic activity of transamination and nitrogen disposal enzymes are increased, supporting the importance of amino acid catabolism in the adaption to CR [44]. Our analysis identified 8 highly regulated acetyl sites on proteins involved in transamination and nitrogen detoxification, including carbamoyl phosphate synthase 1 (CPS1, K1356), glutaminase (GLS2, K329), glutamate dehydrogenase (GLUD1, K480), and ornithine carbamoyltransferase (OTC, K275), demonstrating the extent to which SIRT3 regulates amino acid catabolism and the urea cycle in response to CR. Metabolite analysis revealed an accumulation of many amino acids in *Sirt3*$^{-/-}$ MEFs, including glutamate, glutamine, serine, threonine and tyrosine. These data are supportive of significant defects in amino acid catabolism resulting from the loss of SIRT3. The accumulation of glutamine over time in *Sirt3*$^{-/-}$ MEFs provides evidence that acetylation down-regulates GLS2 activity, a primary enzyme responsible for glutamine catabolism. Here, we identified K329 within GLS2 with a 4.7-fold increase in acetyl occupancy in *Sirt3*$^{-/-}$-CR. Based on the homologous site in the cytoplasmic isoform of GLS2, K329 lies in the homodimeric interface and would be predicted to affect activity [45].

Amino acid catabolism is an important anapleurotic pathway for other mitochondrial processes. The breakdown of BCAA (isoleucine, leucine, valine) results in the production of acetyl-CoA and succinyl CoA, metabolites used in the citric acid cycle (TCA) or ketone body synthesis, pathways important in CR responses. Acetylome data reveal BCAA catabolism as a highly enriched KEGG pathway among mitochondrial proteins regulated by SIRT3 in response to CR. Nearly half of the proteins involved in this pathway contain significantly increasing acetyl sites in *Sirt3*$^{-/-}$-CR**.** These include two subunits of branched-chain alpha-keto acid dehydrogenase (BCKDHA and AUH), the enzyme catalyzing the first, shared step in breakdown of all BCAA in the liver. Supporting the conclusion that acetylation decreases BCKDHA activity, isoleucine, leucine, and valine are elevated in *Sirt3*$^{-/-}$ MEFs and in *Sirt3*$^{-/-}$ liver tissue.

Deficiency in acetyl-CoA acetyltransferase (ACAT1) in humans has been associated with a buildup of branched chain amino-acid metabolites [46]. ACAT1 plays a critical role in BCAA catabolism as well as acetyl-CoA metabolism and ketone body synthesis [47]. Our acetylome analysis provided strong evidence that ACAT1 is a major target of SIRT3. Of the 22 acetyl lysines quantified in ACAT1, three (K260, K263, K265) display a 16- to 32-fold acetylation increase in the absence of SIRT3. These residues are located within the CoA binding pocket of ACAT1, within 3-5 Å from the ribosyl-phosphate group of CoA [48], thus we predict acetylation of these residues will alter CoA binding and regulate ACAT1 activity.

Additional evidence for SIRT3-dependent regulation of BCAA catabolism comes from the elevated levels of 3-hydroxyisovalerate in *Sirt3*$^{-/-}$ MEFs at early time points. Increased excretion of 3-hydroxyisovalerate is an indicator of BCAA catabolic deficiency, organic acidemia, and is associated with a rare human recessive autosomal disorder caused by isovaleryl-CoA dehydrogenease (IVD) deficiency [49]. Our acetylome analysis revealed that the acetylation state of IVD at K316 is controlled by SIRT3 in a CR-dependent manner. K316 undergoes deacetylation during CR, but a 3.27-fold increase in acetylation in the absence of SIRT3. K316 is found near the subunit interface of IVD and may affect tetrameric assembly [50].

Citric acid cycle

An increase in respiration efficiency is a hallmark of the mitochondrial response to CR [51]. This relies on precisely controlled production and consumption of redox coenzymes NADH and $FADH_2$, produced in the TCA cycle. We identified nine acetyl sites on four enzymes of the TCA cycle that display increased acetylation of >4-fold, including citrate synthase (CS, K49|K52), aconitase (ACO2, K144), malate dehydrogenase (MDH2, K239) and succinate dehydrogenase (SDHA, K179). Metabolite analysis identified an increase in succinate in *Sirt3*$^{-/-}$ MEFs and provides further evidence for acetylation decreasing succinate dehydrogenase activity. Consistent with a role for dynamic acetylation in the TCA cycle, alterations in glutamate, lactate, and aspartate were also apparent in *Sirt3*$^{-/-}$ MEFs.

One of the most striking examples of hyperacetylation upon loss of SIRT3 is malate dehydrogenase (MDH2). Among seven quantified acetyl sites on MDH2, K239 displayed a 66-fold increase in acetylation in *Sirt3*$^{-/-}$ mice. This same site exhibited decreased acetylation under CR with WT animals. Within the MDH2 structure, K239 lies within the dimeric interface and near to the substrate-binding channel (PDB 2DFD). Thus K239 is a likely target of SIRT3 in response to CR. Within the MDH2 structure, K239 lies within the dimeric interface and near to the substrate-binding channel (PDB 2DFD). To provide evidence that acetylation of K239 would affect MDH2 activity, we generated the acetyl mimic (K239Q) and compared activity against WT protein. Indeed, the K239Q mutant exhibited a dramatic loss in activity (>5-fold) when compared to WT, suggesting acetylation of this site dynamically regulates MDH2 activity (**Figure 6F**). This example provides additional validation of our acetylome atlas, and demonstrates the utility of mining the data for functional importance and defining specific roles for SIRT3.

One-carbon metabolism

One-carbon (1C) transfer reactions are essential for amino acid catabolism, transfer of methyl groups to DNA, RNA, and proteins, and in formylation of the initiator tRNA required for mitochondrial protein synthesis [52]. Serine, glycine, dimethylglycine, sarcosine, and choline serve as mitochondrial 1C donors. Four key proteins involved in 1C transfer reactions contain acetyl sites responsive to CR and *Sirt3*$^{-/-}$. Serine hydroxymethyltransferase (SHMT2, K464) is involved in breakdown of serine to glycine, with concomitant conversion of tetrahydrofolate to 5,10-methylenetetrahydrofolate. Lysine 464 displays 4-fold increased acetylation in response to *Sirt3*$^{-/-}$. Serine contributes significantly to the segregation between *Sirt3*$^{-/-}$ and *Sirt3*$^{+/+}$ MEFs by PCA and is elevated in *Sirt3*$^{-/-}$ MEFs**,** supporting the role of SIRT3 in regulating SHMT2 activity and a defect in serine utilization. Mitochondrial dimethylglycine dehydrogenase (DMGDH) converts dimethylglycine to sarcosine and contains an acetyl site (K852) whose relative acetyl occupancy increases a dramatic 36-fold in *Sirt3*$^{-/-}$-CR.

Choline is a major source of dietary methyl groups in mammals, and oxidation of choline to betaine, which ultimately supplies 1C units to the mitochondria as dimethylgylcine, takes place primarily in the liver. Choline dehydrogenase (CHDH, K582) contains an acetyl site displaying increased acetylation of 6.8 fold. Betaine, choline, and homocysteine, a metabolite necessary for the conversion of betaine to dimethylglycine, are all altered in liver tissue in response to CR and *Sirt3*$^{-/-}$. Together these results provide compelling evidence that SIRT3 controls donation of 1C units to carbon carriers in the mitochondria.

Mitochondrial integrity

CR is reported to enhance mitochondrial proliferation [51,53]. Consistent with new protein synthesis, over 40% of gene transcripts overexpressed in response to CR [54] were identified in the current study with increased (≥40%) relative protein level in CR. However, overall proteome quantitation argues against widespread mitochondrial biogenesis during CR, as discussed previously [55]. We observed a slight overall enrichment of mitochondrial proteins relative to non-mitochondrial proteins during CR. Our data support reversible acetylation as a major mechanism that drives CR-dependent metabolic reprogramming in mitochondria.

It is known that CR-adapted mitochondria use less oxygen and produce less ROS, while maintaining ATP production. One molecular mechanism to enhance mitochondrial performance while minimizing new protein synthesis might be the stimulation of effective quality control pathways for mitochondrial proteins. Our study supports this hypothesis. We identified significant SIRT3-dependent changes in acetylation (>6-fold, $p<0.05$) on major enzymatic regulators of mitochondrial proteome homeostasis, including the ATP-dependent Lon-like protease (LONP1, K378), and the chaperone proteins, 100 kDa chaperone (CLPX, K127), heat shock protein 70 (HSPA9, K653), 10 kDa heat shock protein (HSPE1, K36).

Our analysis reveals a potential role for SIRT3 in coordinating the regulatory network required for maintaining mitochondrial integrity, including protein quality control, mtDNA transcription and

translation, and iron homeostasis. Mitochondria play a central role in iron metabolism and serve as a key location for iron sulfur cluster (Fe-S) and heme biogenesis [56]. Several proteins involve in homeostasis are well-represented as major targets of SIRT3 deacetylation. These include frataxin (FXN), ACO2 and GLRX5.

Mitochondrial genome maintenance, transcription and translation

A number of components from mitochondrial transcription and translational machinery are highlighted in the network analysis of SIRT3-regulated acetylation sites. Mitochondrial transcription factor A (TFAM, K70) plays an essential role in maintenance, expression, and organization of mitochondrial DNA [57]. TFAM is required for efficient promoter recognition by mitochondrial RNA polymerase and binds and induces significant conformational changes in mtDNA [58]. We identified a previously unreported acetyl lysine in TFAM, K70 that exhibits a 8-fold increase in acetylation in *Sirt3*$^{-/-}$, and a slight decrease in acetylation under CR when SIRT3 expression is induced. Reversible acetylation of TFAM could control mtDNA binding and subsequent transcription, akin to the role of histone acetylation regulating nuclear gene expression. Mitochondrial ribosomal proteins (MRPS9, K174 and MRPL12, K145) play an important role in coordinating mitochondrial transcription and translation and contain acetyl sites that increase 5- and 18-fold, respectively in the *Sirt3*$^{-/-}$-CR condition. MRPL12 interacts with mitochondrial RNA polymerase and is thought to enhance mitochondrial transcription [59]. The essential mitochondrial elongation factor Ts (TSFM, K83) is also a target of SIRT3-dependent acetylation. Mutations in TSFM result in severe infantile liver failure [60]. TSFM functions as an activator of elongation factor Tu, which delivers the aminoacyl-tRNA to the ribosome during elongation of protein synthesis. K83 exists at a kink in alpha helix 3 of TSFM, connecting two major contact regions with elongation factor Tu, and would be predicted to alter protein:protein interaction [61].

^{13}C-glucose flux analysis confirms major defects in oxidative glucose metabolism in Sirt3$^{-/-}$ MEFs

Our acetylome and metabolite analyses suggest a profound up-regulation of oxidative metabolism by the dynamic regulation of protein acetylation induced by CR and mediated by SIRT3. To provide support that wholesale alteration to the acetylome results in a major shift in glucose metabolism, we measured the metabolic flux of [U-^{13}C]-glucose in *Sirt3*$^{-/-}$ and *Sirt3*$^{+/+}$ MEFs. Conditioned-media from cells treated with [U-^{13}C]-glucose were analyzed by NMR spectroscopy and the consumption of glucose and production of ^{13}C-labeled metabolites was monitored over 16 hours (**Figure 6D**). Strikingly, the rate of ^{13}C-lactate and ^{13}C-alanine production and excretion to media was increased dramatically in *Sirt3*$^{-/-}$ MEFs 17- and 4-fold, respectively, compared to *Sirt3*$^{+/+}$ (**Figure 6E**). These flux studies indicate a major defect in oxidative glucose metabolism in the absence of SIRT3. Glucose uptake is slightly increased in *Sirt3*$^{-/-}$ MEFs and a buildup of glucose derived metabolites such as myo-inositol and UDP-glucose, was apparent in ^{1}H-spectral analysis (**Figure 6B**). These results reinforce the hypothesis that SIRT3 plays a critical role in mitochondrial metabolic reprogramming towards oxidative metabolism and are consistent with previous reports [62].

Sirt3*$^{-/-}$ *mice display inability to lower liver acetyl-CoA levels during CR

The metabolic flux results suggest decreased ability of *Sirt3*$^{-/-}$ MEFs cells to perform full oxidation of glucose. The consequences on acetyl-CoA and CoA levels were assessed in liver from *Sirt3*$^{-/-}$ and WT mice, under either CD or CR diet. In WT animals, we observed a considerable (5x) decrease in liver acetyl-CoA under CR compared with CD (**Figure 6E**). In contrast, *Sirt3*$^{-/-}$ mice were unable to lower their acetyl-CoA levels in response to CR (**Figure 6E**). The ratios of acetyl-CoA/CoA displayed similar trends (**Figure 6E**), indicating that the difference in these ratios reflect differences primarily in acetyl-CoA. Lower levels of acetyl-CoA during CR are consistent with increased acetyl-CoA utilization in energy requiring processes, and validate the role of SIRT3 in mitochondrial reprogramming towards efficient oxidative metabolism. *Sirt3*$^{-/-}$ mice display lower acetyl CoA under either dietary condition and were unable to lower their acetyl CoA levels in response to CR. These findings are consistent with a

deficiency in mitochondrial oxidative metabolism due to misregulation of protein acetylation in the absence of SIRT3.

Experimental procedures

Mitochondrial preparation

Mitochondria were isolated by differential centrifugation by methods previously reported[63]. All steps were performed at 4˚C. Tissue was suspended in isolation buffer [(220 mM mannitol, 70 mM sucrose, 5 mM HEPES-KOH, pH 7.4, 1 mM EGTA) supplemented with 1.0 mg/mL bovine serum albumin (BSA), protease inhibitor cocktail (Roche cOmplete tablets 1 tablet per 50 mL buffer), phosphatase inhibitor cocktail (Roche PhosSTOP) and deacetylase inhibitors (10 mM nicotinamide, 10 μM TSA, 5 μM MS257, 10 mM sodium butyrate, 2 μM SAHA)]. Liver tissue was suspended to 0.1 g/mL and homogenized with four strokes of a Potter-Elvehjem Teflon homogenizer connected to a motorized stirrer at 1000 rpm. The homogenate was decanted and spun at 1000 rpm for 10 minutes. Any lipids were removed from the top of the supernatant by aspiration. The supernatant containing mitochondria was removed and transferred to a Beckman Ultra-clear centrifuge tube and spun at 8000 rpm for 10 minutes. The supernatant was discarded and the pellet was re-suspended in isolation buffer. The crude mitochondria were transferred to a 1.5 mL microfuge tube and spun at 8000 rpm for 10 minutes in a bench-top centrifuge. The supernatant was removed and the pellet was re-suspended in resuspension buffer [(220 mM mannitol, 70 mM sucrose, 5 mM HEPES-KOH, pH 7.4, 1 mM EGTA)]. The mitochondria were pelleted by centrifugation at 8000 rpm for 10 minutes in a bench top centrifuge. The supernatant was aspirated and the pellet was frozen immediately in liquid nitrogen.

Sample Preparation

Purified mitochondria were suspended in 8 M urea, 50 mM Tris pH 8.0, 5 mM $CaCl_2$, 100 mM NaCl, protease inhibitors, and deacetylase inhibitors. Protein was extracted by sonication with a probe sonicator on ice and quantified by BCA (Pierce). One mg of protein from each sample was reduced and alkylated with 5 mM DTT and 15 mM iodoacetamide respectively, each for 30 min at ambient temperature. The samples were then digested with LysC (Wako) for 2 hrs at ambient temperature. Following 8 fold dilution with 50 mM Tris pH 8.0, 100 mM NaCl, 5 mM $CaCl_2$ the samples were further digested with trypsin (Promega) overnight at ambient temperature. Digestions were quenched by TFA acidification, and desalted with a tC18 sep-Pak (Waters). Desalted material was labeled with TMT isobaric labels (Thermo-Pierce), only four of the six labels were used per experiment. Tagged peptides were combined and desalted. Labeling efficiency was evaluated by analyzing a test mixture by LC/MS/MS for each experiment. All experiments had ≥ 97% labeling efficiency, calculated by the number of N-terminal labeled peptides divided by the total number of peptide identifications.

Fractionation and 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; D: Water. The gradient was generated by a Surveyor LC quaternary pump (Thermo) at 3 ml/min flow rate. Peptides were eluted over the following gradient: 0-2 min, 100% A; 2-5 min, 0-15% B; 5-35 min 15-100% B; 35-45 min, 100% B; followed by washes with C and D prior to re-equilibration with mobile phase A. Twenty fractions were collected and desalted. A small portion, 5%, of each was retained for protein analysis. The remaining material was pooled into 7 fractions for acetyl lysine enrichment.

Acetyl Lysine Enrichment

Peptide fractions were dissolved in 50 mM HEPES pH 7.5, 100 mM KCl and to each fraction approximately 50 µL pan-acetyl lysine antibody-agarose conjugate (Immunechem) was added. The samples were rotated overnight at 4 C and then 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.

LC-MS/MS

Acetyl enriched and protein fractions were analyzed by reverse phase liquid chromatography on a nanoAcuity LC (Waters) coupled to an Orbitrap Elite (Thermo). Samples were loaded onto a 75 µm inner diameter column packed with 5 µm magic C18 AQ particles (Michrome) The elution portion of the gradient was 8% to 30% B (A: water/0.2% formic acid; B: acetonitrile/0.2% formic acid) over 135 min for acetyl enriched fractions and 90 min for protein fractions. Acetyl enriched fractions were each analyzed in duplicate.

Mass spectrometry instrument methods all started with one MS1 survey scan (resolution = 60,000; 300 Th – 1,500 Th) followed by data dependent MS2 fragmentation and analysis (resolution = 15,000) of the ten most intense precursors. To reduce reporter ion interference caused by co-isolation of multiple precursors the QuantMode instrument control method was used [64]. This method generates sequence ions by collisional-activated dissociation (CAD; normalized collision energy = 35%; target value = $5e^4$) and reporter ions by beam-type CAD (HCD; normalized collision energy = 80%; target value = $2e^5$) from charge reduced precursors by gas phase reaction with a proton transfer reagent (PTR). The PTR used in all experiments was SF_6. The exclusion duration was 30 seconds for -1.55 Th to +2.55 Th of the sampled precursor. Ions with unassigned charge state or a charge of +1 were excluded.

Database search and FDR filtering

Spectra were converted to searchable text files using DTA generator. Generated text files were searched for fully tryptic peptides with up to three missed cleavages against a UniProt target-decoy

database populated with mouse canonical plus isoforms supplemented with common contaminants (downloaded 28Nov2011) using the Open Mass Spectrometry Search Algorithm (OMSSA) [65]. Mass tolerance was set to 20 ppm for precursors and 0.01 Th for fragment ions. Carbamidomethylation (+57.021464) of cysteine, TMT 6plex on lysine (+229.162932), and TMT 6plex on peptide N-terminus (+229.162932) were searched as fixed modifications for all samples. Methionine oxidation (+15.994915) and TMT 6plex on tyrosine (+229.162932) were searched as variable modifications for all samples. Enriched fractions were additionally searched for variable acetylation modifications. In this case the acetylation mass shift was set to -187.1523 Da, the difference between an acetyl group and a TMT 6plex tag. Using this custom mass shift allows TMT 6plex on lysine to remain as a fixed modification even for acetylated peptides. With the COMPASS software suite search results were filtered to 1% FDR at the unique peptide level [31]. TMT quantitation of identified peptides was performed within COMPASS as previously described [32]. Peptides were grouped into proteins according to previously reported rules and protein identifications were further filtered to 1% FDR [66]. 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 quantitation.

Acetylation Localization

Acetylation events were localized to specific residues using probabilistic methods described elsewhere [67]. Briefly, for each peptide spectral match (PSM) that contains an acetyl modification, every possible peptide isoform was generated and fragmented in silico to produce theoretical fragmentation spectra. Each theoretical spectrum was compared to the experimental spectrum at 10 PPM *m/z* tolerances; the number of matching peaks was recorded and a p-value was calculated using a cumulative binomial distribution. An AScore (i.e. the difference of p-values) was calculated between every pair of isoforms. A peptide was declared “localized” if all AScores for a particular isoform were larger than the minimum value (AScore = 13, p-value < 0.05) for every comparison. Localized acetylated peptides were grouped

together if they share identical modification sites and the reporter ion intensities were summed; peptides with C-terminal acetylation are excluded from quantitation. For simplicity acetyl isoforms are referred to as acetyl sites.

Protein Normalization

All quantitative values were log2 transformed and mean normalized. 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 statistically 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 using Welch's t-test with Storey correction for multiple hypotheses testing [68]. Acetyl sites identified in only a single replicate experiment were excluded from p-value calculations. When an acetyl isoform or its corresponding protein had no intensity in a reporter ion channel, it was excluded from the protein normalization calculation and was not considered quantifiable.

Mitochondrial assignment and site novelty

Proteins were identified as mitochondrial or non-mitochondrial based on inclusion or exclusion from the mitoCarta compendium of mitochondrial mouse proteins. MitoCarta EntrezID identifiers were converted to UniProt identifiers with the UniProt ID mapping function. This list was limited to contain only proteins that were in the mouse canonical database used for searching. Mitochondrial acetylation sites were considered novel if they were absent from both the UniProt and PhosphoSitePlus databases.

Gene Ontology (GO) Analysis

GO enrichment analysis was performed using the functional annotation tool of the DAVID bioinformatics resource [69,70]. Redundant proteins were removed from the data set and searched against a background that included all mitochondrial proteins in the MitoCarta database [22]. This background set was chosen to identify enriched mitochondrial processes and processes were sorted by p-value. Protein interaction network analysis was performed using GeneMANIA functional association data [40] and visualized in Cytoscape [41].

SIRT3 expression and global acetylation analysis

Antibodies for western blotting included anti-mouse SIRT3 antibody and pan-anti acetylated lysine [9] prepared by GeneTel Laboratories LLC, Madison, WI. The rabbit anti-SIRT3 antibody was generated with a synthetic peptide corresponding to the mouse SIRT3 sequence DLMQRERGKLDGQDRKLH. Anti-VDAC1 (Abcam ab14734) was used as a mitochondrial loading control. Mitochondrial samples were boiled with SDS loading buffer and subjected to western blotting.

Acknowledgments

This work was funded by NIA grant AG038679 to J.M.D. and T.A.P, NIH grant GM065386 to J.M.D., NIH GM080148 to J.J.C., and Searle Scholars Award and NIH grant RC1DK086410 to D.J.P. K.E.D. was funded by a NSF Graduate Research Fellowship and NIH traineeship (5T32GM08349). We thank AJ Bureta for figure design.

References

1 Colman, R. J., Anderson, R. M., Johnson, S. C., Kastman, E. K., Kosmatka, K. J., Beasley, T. M. *et al.* (2009) Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science* **325**, 201-204, doi:10.1126/science.1173635

2 Mattison, J. A., Roth, G. S., Beasley, T. M., Tilmont, E. M., Handy, A. M., Herbert, R. L. *et al.* (2012) Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study. *Nature*, doi:10.1038/nature11432

3 Weindruch, R., Walford, R. L., Fligiel, S. & Guthrie, D. (1986) The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. *J Nutr* **116**, 641-654

4 Koubova, J. & Guarente, L. (2003) How does calorie restriction work? *Genes Dev* **17**, 313-321, doi:10.1101/gad.1052903

5 Sohal, R. S. & Weindruch, R. (1996) Oxidative stress, caloric restriction, and aging. *Science* **273**, 59-63

6 Anderson, R. M. & Weindruch, R. (2012) The caloric restriction paradigm: implications for healthy human aging. *Am J Hum Biol* **24**, 101-106, doi:10.1002/ajhb.22243

7 Wallace, D. C. (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* **39**, 359-407, doi:10.1146/annurev.genet.39.110304.095751

8 Lin, S. J., Defossez, P. A. & Guarente, L. (2000) Requirement of NAD and SIR2 for life-span extension by calorie restriction in Saccharomyces cerevisiae. *Science* **289**, 2126-2128

9 Someya, S., Yu, W., Hallows, W. C., Xu, J., Vann, J. M., Leeuwenburgh, C. *et al.* (2010) Sirt3 mediates reduction of oxidative damage and prevention of age-related hearing loss under caloric restriction. *Cell* **143**, 802-812, doi:10.1016/j.cell.2010.10.002

10 Yu, W., Dittenhafer-Reed, K. E. & Denu, J. M. (2012) SIRT3 protein deacetylates isocitrate dehydrogenase 2 (IDH2) and regulates mitochondrial redox status. *J Biol Chem* **287**, 14078-14086, doi:10.1074/jbc.M112.355206

11 Hirschey, M. D., Shimazu, T., Goetzman, E., Jing, E., Schwer, B., Lombard, D. B. *et al.* (2010) SIRT3 regulates mitochondrial fatty-acid oxidation by reversible enzyme deacetylation. *Nature* **464**, 121-125, doi:10.1038/nature08778

12 Lombard, D. B., Alt, F. W., Cheng, H. L., Bunkenborg, J., Streeper, R. S., Mostoslavsky, R. *et al.* (2007) Mammalian Sir2 homolog SIRT3 regulates global mitochondrial lysine acetylation. *Mol Cell Biol* **27**, 8807-8814, doi:10.1128/MCB.01636-07

13 Hallows, W. C., Lee, S. & Denu, J. M. (2006) Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. *Proc Natl Acad Sci U S A* **103**, 10230-10235, doi:10.1073/pnas.0604392103

14 Schwer, B., Bunkenborg, J., Verdin, R. O., Andersen, J. S. & Verdin, E. (2006) Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-CoA synthetase 2. *Proc Natl Acad Sci U S A* **103**, 10224-10229, doi:10.1073/pnas.0603968103

15 Hallows, W. C., Yu, W., Smith, B. C., Devries, M. K., Ellinger, J. J., Someya, S. *et al.* (2011) Sirt3 promotes the urea cycle and fatty acid oxidation during dietary restriction. *Mol Cell* **41**, 139-149, doi:10.1016/j.molcel.2011.01.002

16 Shimazu, T., Hirschey, M. D., Hua, L., Dittenhafer-Reed, K. E., Schwer, B., Lombard, D. B. *et al.* (2010) SIRT3 deacetylates mitochondrial 3-hydroxy-3-methylglutaryl CoA synthase 2 and regulates ketone body production. *Cell Metab* **12**, 654-661, doi:10.1016/j.cmet.2010.11.003

17 Qiu, X., Brown, K., Hirschey, M. D., Verdin, E. & Chen, D. (2010) Calorie restriction reduces oxidative stress by SIRT3-mediated SOD2 activation. *Cell Metab* **12**, 662-667, doi:10.1016/j.cmet.2010.11.015

18 Schlicker, C., Gertz, M., Papatheodorou, P., Kachholz, B., Becker, C. F. & Steegborn, C. (2008) Substrates and regulation mechanisms for the human mitochondrial sirtuins Sirt3 and Sirt5. *J Mol Biol* **382**, 790-801, doi:10.1016/j.jmb.2008.07.048

19 Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J. *et al.* (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* **23**, 607-618, doi:10.1016/j.molcel.2006.06.026

20 Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T. *et al.* (2010) Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000-1004, doi:10.1126/science.1179689

21 Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C. *et al.* (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840, doi:10.1126/science.1175371

22 Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S. E. *et al.* (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-123, doi:10.1016/j.cell.2008.06.016

23 Lundby, A., Lage, K., Weinert, B. T., Bekker-Jensen, D. B., Secher, A., Skovgaard, T. *et al.* (2012) Proteomic Analysis of Lysine Acetylation Sites in Rat Tissues Reveals Organ Specificity and Subcellular Patterns. *Cell Rep*, doi:10.1016/j.celrep.2012.07.006

24 Henriksen, P., Wagner, S. A., Weinert, B. T., Sharma, S., Bacinskaja, G., Rehman, M. *et al.* (2012) Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in Saccharomyces cerevisiae. *Mol Cell Proteomics*, doi:10.1074/mcp.M112.017251

25 Chen, Y., Zhao, W., Yang, J. S., Cheng, Z., Luo, H., Lu, Z. *et al.* (2012) Quantitative acetylome analysis reveals the roles of SIRT1 in regulating diverse substrates and cellular pathways. *Mol Cell Proteomics*, doi:10.1074/mcp.M112.019547

26 Fritz, K. S., Galligan, J. J., Hirschey, M. D., Verdin, E. & Petersen, D. R. (2012) Mitochondrial acetylome analysis in a mouse model of alcohol-induced liver injury utilizing SIRT3 knockout mice. *J Proteome Res* **11**, 1633-1643, doi:10.1021/pr2008384

27 Kendrick, A. A., Choudhury, M., Rahman, S. M., McCurdy, C. E., Friederich, M., Van Hove, J. L. *et al.* (2011) Fatty liver is associated with reduced SIRT3 activity and mitochondrial protein hyperacetylation. *Biochem J* **433**, 505-514, doi:10.1042/BJ20100791

28 Law, I. K., Liu, L., Xu, A., Lam, K. S., Vanhoutte, P. M., Che, C. M. *et al.* (2009) Identification and characterization of proteins interacting with SIRT1 and SIRT3: implications in the anti-aging and metabolic effects of sirtuins. *Proteomics* **9**, 2444-2456, doi:10.1002/pmic.200800738

29 Schwer, B., Eckersdorff, M., Li, Y., Silva, J. C., Fermin, D., Kurtev, M. V. *et al.* (2009) Calorie restriction alters mitochondrial protein acetylation. *Aging Cell* **8**, 604-606, doi:10.1111/j.1474-9726.2009.00503.x

30 Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. *et al.* (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Meth* **8**, 933-935, doi:http://www.nature.com/nmeth/journal/v8/n11/abs/nmeth.1716.html#supplementary-information

31 Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. (2011) COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* **11**, 1064-1074, doi:10.1002/pmic.201000616

32 Phanstiel, D. H., Brumbaugh, J., Wenger, C. D., Tian, S., Probasco, M. D., Bailey, D. J. *et al.* (2011) Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Meth* **8**, 821-827, doi:http://www.nature.com/nmeth/journal/v8/n10/abs/nmeth.1699.html#supplementary-information

33 Wu, R., Dephoure, N., Haas, W., Huttlin, E. L., Zhai, B., Sowa, M. E. *et al.* (2011) Correct Interpretation of Comprehensive Phosphorylation Dynamics Requires Normalization by Protein Expression Changes. *Molecular & Cellular Proteomics* **10**, doi:10.1074/mcp.M111.009654

34 Guan, K. L., Yu, W., Lin, Y., Xiong, Y. & Zhao, S. (2010) Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides. *Nat Protoc* **5**, 1583-1595, doi:10.1038/nprot.2010.117

35 Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences* **104**, 5860-5865, doi:10.1073/pnas.0608638104

36 Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistcal Learning*. (Springer, 2009).

37 Smith, B. C., Settles, B., Hallows, W. C., Craven, M. W. & Denu, J. M. (2011) SIRT3 substrate specificity determined by peptide arrays and machine learning. *ACS Chem Biol* **6**, 146-157, doi:10.1021/cb100218d

38 Jin, L., Wei, W., Jiang, Y., Peng, H., Cai, J., Mao, C. *et al.* (2009) Crystal structures of human SIRT3 displaying substrate-induced conformational changes. *J Biol Chem* **284**, 24394-24405, doi:10.1074/jbc.M109.014928

39 Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202, doi:10.1006/jmbi.1999.3091

40 Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-220, doi:10.1093/nar/gkq537

41 Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431-432, doi:10.1093/bioinformatics/btq675

42 Bathe, O. F., Shaykhutdinov, R., Kopciuk, K., Weljie, A. M., McKay, A., Sutherland, F. R. *et al.* (2011) Feasibility of identifying pancreatic cancer based on serum metabolomics. *Cancer Epidemiol Biomarkers Prev* **20**, 140-147, doi:10.1158/1055-9965.EPI-10-0712

43 Janes, K. A. & Yaffe, M. B. (2006) Data-driven modelling of signal-transduction networks. *Nat Rev Mol Cell Biol* **7**, 820-828, doi:10.1038/nrm2041

44 Hagopian, K., Ramsey, J. J. & Weindruch, R. (2003) Caloric restriction increases gluconeogenic and transaminase enzyme activities in mouse liver. *Exp Gerontol* **38**, 267-278

45 DeLaBarre, B., Gross, S., Fang, C., Gao, Y., Jha, A., Jiang, F. *et al.* (2011) Full-length human glutaminase in complex with an allosteric inhibitor. *Biochemistry-Us* **50**, 10764-10770, doi:10.1021/bi201613d

46 Fontaine, M., Briand, G., Ser, N., Armelin, I., Rolland, M. O., Degand, P. *et al.* (1996) Metabolic studies in twin brothers with 2-methylacetoacetyl-CoA thiolase deficiency. *Clin Chim Acta* **255**, 67-83

47 Haapalainen, A. M., Merilainen, G., Pirila, P. L., Kondo, N., Fukao, T. & Wierenga, R. K. (2007) Crystallographic and kinetic studies of human mitochondrial acetoacetyl-CoA thiolase: the importance of potassium and chloride ions for its structure and function. *Biochemistry-Us* **46**, 4305-4321, doi:10.1021/bi6026192

48 Dombrovski, L., Min, J. R., Antoshenko, T., Wu, H., Loppnau, P., Edwards, A. M. *et al.* (RCSB, 2011).

49 Lin, W. D., Wang, C. H., Lee, C. C., Lai, C. C., Tsai, Y. & Tsai, F. J. (2007) Genetic mutation profile of isovaleric acidemia patients in Taiwan. *Mol Genet Metab* **90**, 134-139, doi:10.1016/j.ymgme.2006.08.011

50 Tiffany, K. A., Roberts, D. L., Wang, M., Paschke, R., Mohsen, A. W., Vockley, J. *et al.* (1997) Structure of human isovaleryl-CoA dehydrogenase at 2.6 A resolution: structural basis for substrate specificity. *Biochemistry-Us* **36**, 8455-8464, doi:10.1021/bi970422u

51 Civitarese, A. E., Carling, S., Heilbronn, L. K., Hulver, M. H., Ukropcova, B., Deutsch, W. A. *et al.* (2007) Calorie restriction increases muscle mitochondrial biogenesis in healthy humans. *Plos Med* **4**, 485-494, doi:Artn E76

Doi 10.1371/Journal.Pmed.0040076

52 Tibbetts, A. S. & Appling, D. R. (2010) Compartmentalization of Mammalian Folate-Mediated One-Carbon Metabolism. *Annu Rev Nutr* **30**, 57-81, doi:Doi 10.1146/Annurev.Nutr.012809.104810

53 Lopez-Lluch, G., Hunt, N., Jones, B., Zhu, M., Jamieson, H., Hilmer, S. *et al.* (2006) Calorie restriction induces mitochondrial biogenesis and bioenergetic efficiency. *Proc Natl Acad Sci U S A* **103**, 1768-1773, doi:Doi 10.1073/Pnas.0510452103

54 Plank, M., Wuttke, D., van Dam, S., Clarke, S. A. & de Magalhaes, J. P. (2012) A meta-analysis of caloric restriction gene expression profiles to infer common signatures and regulatory mechanisms. *Mol Biosyst* **8**, 1339-1349, doi:10.1039/c2mb05255e

55 Miller, B. F., Robinson, M. M., Bruss, M. D., Hellerstein, M. & Hamilton, K. L. (2012) A comprehensive assessment of mitochondrial protein synthesis and cellular proliferation with age and caloric restriction. *Aging Cell* **11**, 150-161, doi:10.1111/j.1474-9726.2011.00769.x

56 Richardson, D. R., Lane, D. J., Becker, E. M., Huang, M. L., Whitnall, M., Suryo Rahmanto, Y. *et al.* (2010) Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol. *Proc Natl Acad Sci U S A* **107**, 10775-10782, doi:10.1073/pnas.0912925107

57 Ngo, H. B., Kaiser, J. T. & Chan, D. C. (2011) The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat Struct Mol Biol* **18**, 1290-U1141, doi:Doi 10.1038/Nsmb.2159

58 Campbell, C. T., Kolesar, J. E. & Kaufman, B. A. (2012) Mitochondrial transcription factor A regulates mitochondrial transcription initiation, DNA packaging, and genome copy number. *Biochim Biophys Acta*, doi:10.1016/j.bbagrm.2012.03.002

59 Wang, Z., Cotney, J. & Shadel, G. S. (2007) Human mitochondrial ribosomal protein MRPL12 interacts directly with mitochondrial RNA polymerase to modulate mitochondrial gene expression. *J Biol Chem* **282**, 12610-12618, doi:10.1074/jbc.M700461200

60 Vedrenne, V., Galmiche, L., Chretien, D., de Lonlay, P., Munnich, A. & Rotig, A. (2012) Mutation in the mitochondrial translation elongation factor EFTs results in severe infantile liver failure. *J Hepatol* **56**, 294-297, doi:10.1016/j.jhep.2011.06.014

61 Jeppesen, M. G., Navratil, T., Spremulli, L. L. & Nyborg, J. (2005) Crystal structure of the bovine mitochondrial elongation factor Tu.Ts complex. *J Biol Chem* **280**, 5071-5081, doi:10.1074/jbc.M411782200

62 Finley, L. W., Carracedo, A., Lee, J., Souza, A., Egia, A., Zhang, J. *et al.* (2011) SIRT3 opposes reprogramming of cancer cell metabolism through HIF1alpha destabilization. *Cancer Cell* **19**, 416-428, doi:10.1016/j.ccr.2011.02.014

63 Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S.-E. *et al.* (2008) A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. *Cell* **134**, 112-123, doi:http://dx.doi.org/10.1016/j.cell.2008.06.016

64 Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. *et al.* (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature methods*

65 Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M. *et al.* (2004) Open Mass Spectrometry Search Algorithm. *Journal of Proteome Research* **3**, 958-964, doi:10.1021/pr0499491

66 Nesvizhskii, A. I. & Aebersold, R. (2005) Interpretation of Shotgun Proteomic Data. *Molecular & Cellular Proteomics* **4**, 1419-1440, doi:10.1074/mcp.R500012-MCP200

67 Kim, W., Bennett, Eric J., Huttlin, Edward L., Guo, A., Li, J., Possemato, A. *et al.* (2011) Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Molecular Cell* **44**, 325-340, doi:http://dx.doi.org/10.1016/j.molcel.2011.08.025

68 Storey, J. D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498, doi:10.1111/1467-9868.00346

69 Huang, D. W., Sherman, B. T. & Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13, doi:Doi 10.1093/Nar/Gkn923

70 Huang, D. W., Sherman, B. T. & Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:Doi 10.1038/Nprot.2008.211

Chapter 6

Quantification of mitochondrial acetylation dynamics highlights prominent sites of metabolic regulation

ASH designed research, performed mass spectrometry experiments, analyzed mass spectrometry data, and helped write the paper.

This chapter has been published:

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. (2013) Quantification of Mitochondrial Acetylation Dynamics Highlights Prominent Sites of Metabolic Regulation. *Journal of Biological Chemistry* **288**, 26209-26219, doi:10.1074/jbc.M113.483396

*co-first author

Abstract

Lysine acetylation is rapidly becoming established as a key post-translational modification (PTM) 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 to the mouse liver mitochondrial acetyl-proteome 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.

Introduction

Protein acetylation was first identified 50 years ago [64,65], 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 [66-68]. Metabolic enzymes, including many in mitochondria, are among the most heavily acetylated proteins [69-71]. 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 [72-78].

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 [79]. 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) [80]. The reasons for this disparity stem in part from both the relative youth of the field and from technical limitations of most previous 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 acetyl-proteome between two pairs of contrasting nutritional states [81]. First, because PTMs are a primary mechanism of regulation at acute time scales, we assessed changes in the mitochondrial acetyl-proteome between fasted and refed mice. Second, because chronic dysregulation of PTMs can be an important factor in disease, we compared the mitochondrial acetyl-proteome of obese (Leptin-deficient, $Leptin^{ob/ob}$) and lean ($Leptin^{+/+}$) C57/Bl6 (B6) mice. Our analyses

reveal that, amongst a background of largely static acetylation sites, 237 mitochondrial proteins have a significant change ($q \leq 0.1$, 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 Sirt3 targets [81]. Together, these analyses highlighted a select set of mitochondrial proteins as being likely targets for regulatory acetylation, including acetyl-CoA acetyltransferase 1 (Acat1), which possessed several sites of highly dynamic acetylation. By performing site-specific acetyl-lysine 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 coenzyme A (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. Our data is freely available through MitoMod, our online resource for quantitative mitochondrial PTMs (http://mitomod.biochem.wisc.edu/).

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 acetylpeptide enrichment [81] and our QuantMode method [82] for highly accurate quantitative proteomics with isobaric tags to profile mouse liver mitochondrial protein abundance and acetylation levels across the transition between fasting and refeeding. Eight mice were

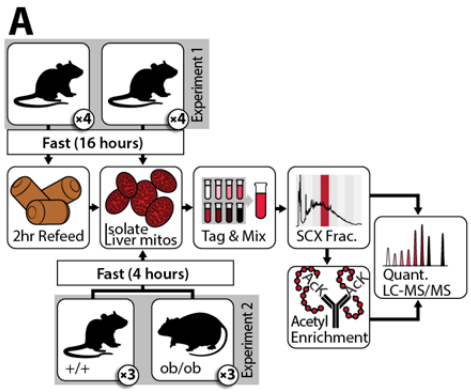
fasted overnight for 16 hours, after which half were allowed to feed *ad libitum* for 2 hours (**Figure 1A**), a time point featuring marked PTM changes with limited protein abundance changes [83]. We identified 1,915 unique acetylation isoforms, of which 1,796 were quantified (**Figure 1B**). Unsupervised 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 (**Figure 2A**), indicating that acetylation changes are a reproducible and distinguishing feature of the transition between fasting and refeeding [84,85]. Importantly, our quantitative approach reveals that approximately 10% (188) of identified acetyl isoforms exhibited a statistically significant change ($q \leq 0.1$) 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 (**Figure 2B**) [86]. Rat liver Mdh2 activity is increased upon feeding, [87] and previous studies have reported that Mdh2 activity can be increased either by acetylation (on K185, K301, K307, and K314) [88] or by Sirt3-dependent deacetylation (of K239) [81]. Here, we detect a large decrease in acetylation occupancy of Mdh2 K239 with little or no changes to the other sites. This indicates that the Sirt3-controlled acetylation site (K239) 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 changes in protein levels that accompany longer perturbations (**Figure 2C**). For example, despite unchanging protein levels among almost all enzymes involved in mitochondrial fatty acid oxidation (FAO), more than ninety distinct acetyl isoforms within

this pathway are altered significantly upon refeeding (**Figure 2D**). These include four acetylation sites on carnitine O-palmitoyltransferase 2 (Cpt2; K104, K544, K537, K453), an enzyme involved in the transport of fatty acids from the cytosol into the mitochondrial matrix for β-oxidation [89]. Similarly, energy production from amino acid degradation is robustly induced in the postprandial state [90], 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 K329 of glutaminase (Gsl2) decreases more than two-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 [91].

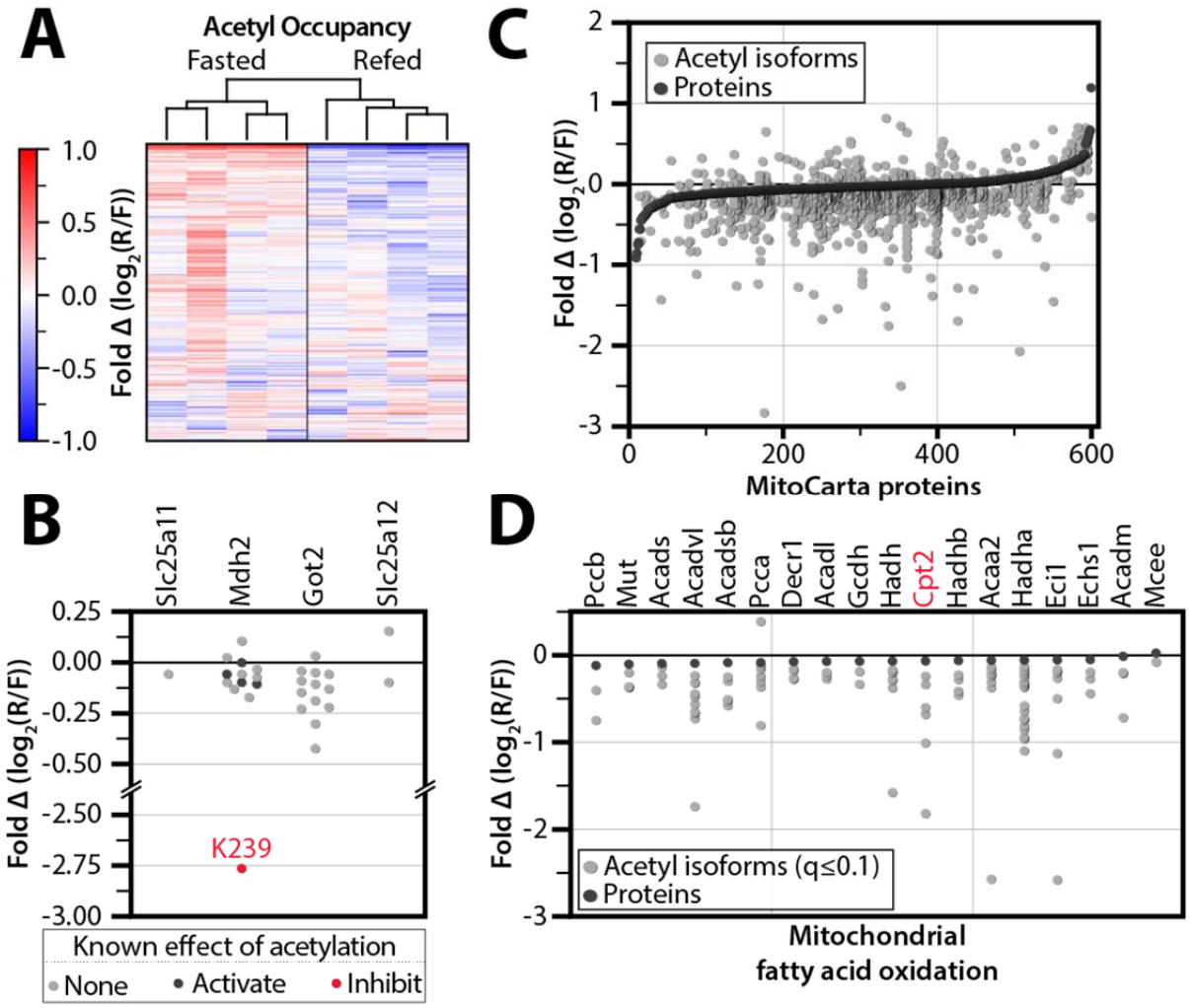
Figure 1. Schematic of workflow for quantitative proteomic and acetylomic analyses of mouse liver mitochondria. (A) In experiment one, eight lean mice were fasted for 16 hours and either sacrificed immediately or refed for two hours before being sacrificed. In experiment two, three obese mice and three lean mice were fasted for four hours before being sacrificed. In each case, liver mitochondria were enriched and analyzed by LC-MS/MS (see Methods). (B) Summary of protein and acetylation data. IDs, unique identifications at 1% FDR; Quant, measurements quantified with isotopic reporter ions in all mice assessed; Δ, measurements significantly changing with q≤ 0.1. See also Dataset S1.



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. Global analysis of acute protein and acetylation changes during the transition from fasting to refeeding. (A) Unsupervised average-linkage hierarchical clustering of 4 fasted and 4 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 less than -1 to greater than 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, while 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 ($q \leq 0.1$) acetyl isoforms are marked in gray.



Quantitative analysis of mitochondrial acetylation dynamics due to chronic obesity

Previous studies have noted general changes to the acetyl-proteome during the onset of obesity [76,92]; 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 acetyl-proteome from three lean (*Leptin*$^{+/+}$) and three obese (*Leptin*$^{ob/ob}$) mice at 10 weeks of age, each fed standard chow diet *ad libitum* (**Figure 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 [93]. These obese mice, however, have only a mild, controlled elevation in blood glucose and model the pre-diabetic 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 ($q \leq 0.1$, **Figure 1B**).

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 two-hour refeeding (**Figure 1B, 3A-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 [83,94], and acetylation has been implicated in inhibiting all five complexes of OxPhos [92,95-98]; 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 to their lean counterparts (**Figure 3D, S1**). In our obese vs.

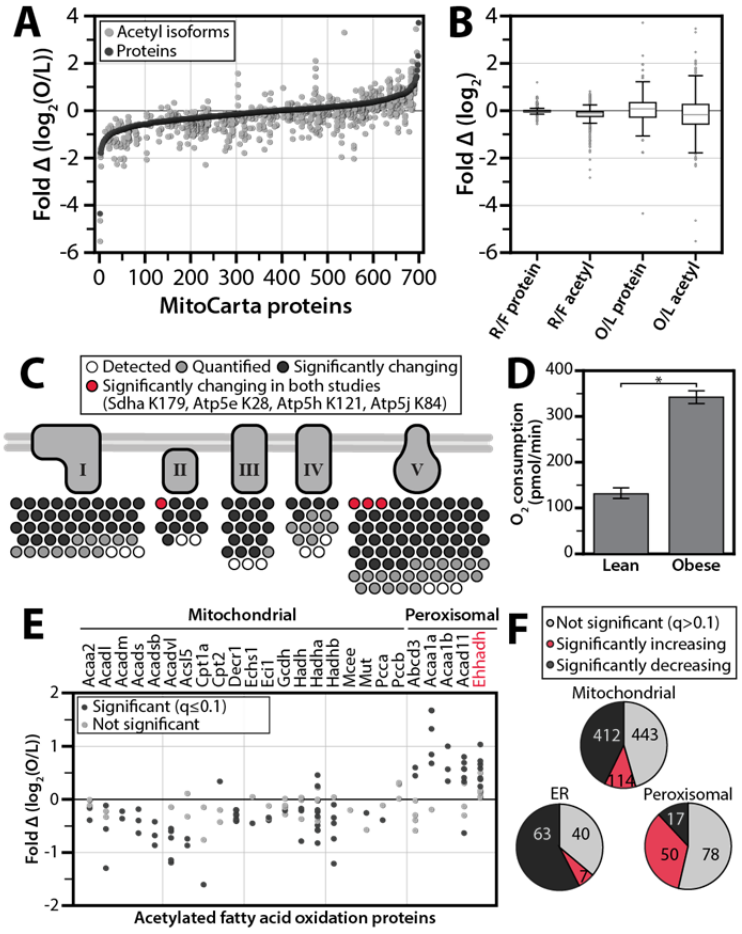
lean acetylome comparison, we identify 156 acetyl isoforms on OxPhos proteins (**Figure 3C**). Ninety-seven of these sites change significantly ($q \leq 0.1$) in acetyl occupancy, 91 of which decrease in the obese state. We therefore predict that a subset of these 91 sites (which we further prioritize 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 K42 on long chain acyl-CoA dehydrogenase (Acadl) is significantly decreased with obesity. This PTM has previously been shown to decrease enzymatic activity [11], consistent with elevated mitochondrial FAO in obese animals. Aminoadipate aminotransferase (Aadat), a protein involved in tryptophan degradation to kynurenine acid [99], has increased acetylation on K263, which is involved in binding the PLP cofactor [100,101]. We predict that acetylation of K263 reduces Aadat activity by preventing this enzyme-cofactor interaction. This decrease in activity could shift tryptophan degradation from the kynurenine-kynurenic acid pathway towards the kynurenine-NAD+ pathway, which has previously been linked to hypertension, diabetes, atherosclerosis, obesity, and immunodeficiency [102].

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 [22]). As such, we were also able to profile the acetyl-proteomes of other co-purifying organelles, including peroxisomes and endoplasmic reticulum (ER). Interestingly, FAO enzymes with evidence of mitochondrial/peroxisomal co-localization are generally increasing in relative occupancy with obesity (38 detected; 24 increase, 1 decreases, $q \leq 0.1$) (**Figure 3E**). In comparison, acetylation sites on mitochondria-specific FAO enzymes are mostly decreasing (89 detected; 3 increase, 49 decrease, $q \leq 0.1$). Among the peroxisomal enzymes with increasing acetylation is Ehhadh (enoyl-CoA, hydratase/3-hydroxyacyl CoA dehydrogenase), which is activated by this PTM [88]. We detect increasing acetylation on seven lysines, with the greatest change on K344, consistent with the known increase

peroxisomal fatty oxidation (*i.e.*, alpha-oxidation) in obesity [87,103]. Overall, acetylation increases on most peroxisomal proteins in the obese state, but decreases on most mitochondrial and ER proteins (**Figure 3F**).

Figure 3. Comparison of of 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 and lean 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 study. 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 ($q \leq 0.1$) in both studies are in red; those significantly changing in the obese vs. lean study only are in black; those isoforms that are not changing significantly are in in gray; isoforms which 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. * indicates significance (p-value = 3.9×10^{-14}). (E) Relative acetyl occupancy changes in fatty acid oxidation. Sites with significantly changing ($q \leq 0.1$) occupancy are marked in black, those that were quantified but not significantly changing in grey, 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).



Integration of disparate datasets 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 ($q \leq 0.1$) due to both refeeding and obesity (**Figure 4C**). 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 K179 in complex II, and Atp5e K28, Atp5h K121, and Atp5j K84 in complex V), further highlighting them as high-likelihood regulatory sites from amongst the hundreds of OxPhos sites identified (**Figure 3C, 4C**).

We reasoned that many *bona fide* regulatory acetylation sites would also be targets of Sirt3, the only well-established mitochondrial deacetylase [80]. 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 [81]. 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 (**Figure 4A-B**). Strikingly, putative Sirt3 targets represent over half of the acetyl isoforms that exhibited a significant change ($q \leq 0.1$) in relative acetyl occupancy in both studies despite accounting for less than 7% of all sites quantified. These results suggest a major, Sirt3-driven remodeling of the acetyl-proteome that is secondary to genetic obesity or to two hours of refeeding after an overnight fast. Measurements for Sirt3 protein using both quantitative MS (**Figure 4B, inset**) and immunoblotting (**Figure S2**) 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 HFD feeding [76], 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 (**Figure 4A-C and S2**). This is consistent with previous reports that alterations in Sirt3 activity are not always correlated with its protein abundance levels [92,104,105], 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 [106].

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 (BCAA) metabolism, fatty acid (FA) 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) [107,108]. 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, 9 of which are significantly changing ($q \leq 0.1$) in at least one comparison. Relative acetyl occupancy at three of these sites changed significantly in both analyses (K187, K260, and K265) (**Figure 5A**), the latter two of which are predicted Sirt3 targets. Another recent Sirt3 knockout study also identifies K260 and K265, but not K187, as Sirt3 targets [78]. 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), while acetylation on K187 increases due to obesity. These data led us to hypothesize that acetylation on K187, K260, and/or K265 alters the activity of Acat1. Moreover, as Sirt3 most often activates its target enzymes [80], we further hypothesized that deacetylation of K260 and K265 increases Acat1 activity in response to nutrient excess.

Figure 4. 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 vs obese/lean comparisons. Only isoforms that were quantified in both studies are represented. (B) Putative Sirt3-targets are shown in red, other MitoCarta proteins in black, and non-MitoCarta proteins in gray. *Inset*, Sirt3 protein abundance fold change by mass spectrometry. * indicates significance ($p \leq 0.05$). (C) Acetylation sites that are significantly changing in both studies ($q \leq 0.1$). Fold changes ($\log_2$) are shown in red (positive) and blue (negative) for refed/fasted (R/F) and obese/lean (O/L).

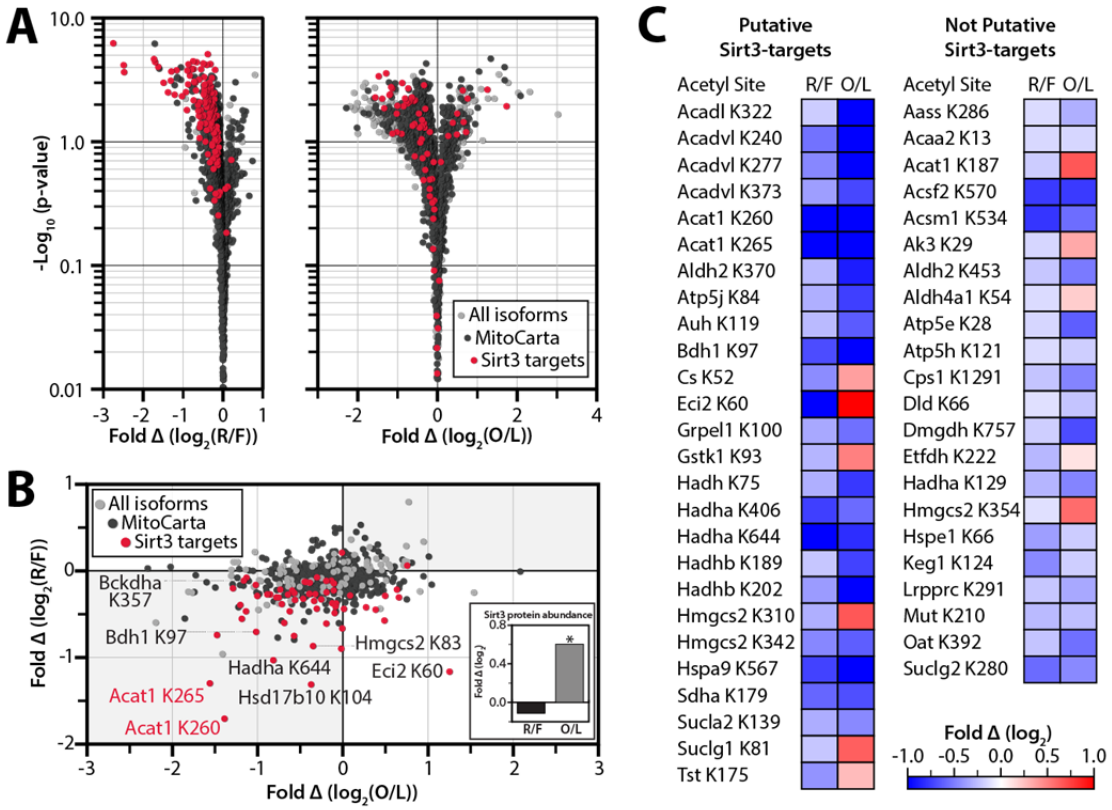
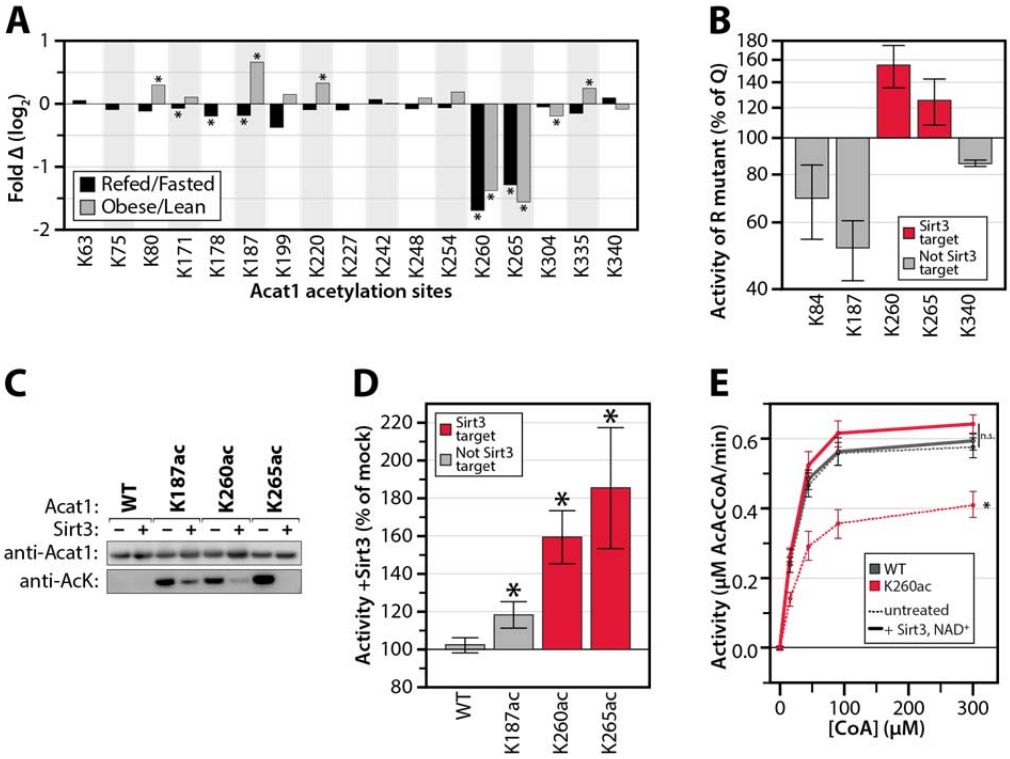


Figure 5. Reversible, Sirt3-regulated, acetylation of Acat1 K260 inhibits activity. (A) Relative acetyl occupancy fold change for 17 quantified sites on Acat1. * indicates q≤ 0.1. (B) *In vitro* enzyme activity assays of Acat1 mutants generated by mutating lysine to glutamine (Q, acetyl-mimic) or arginine (R, deacetyl-mimic). Data are expressed as the average activity of the R mutant as a percentage of that for the Q mutant, at 300 µM CoA (± SEM, N≥ 2). (C) Immunoblot for Acat1 and acetyl-lysine on samples of purified Acat1 protein used in kinetic assays. Proteins were treated with or without NAD^+ and Sirt3. Panel C is representative of samples used in panel D. (D-F) *In vitro* enzyme activity assays of Acat1 with site-specific incorporation of acetyl-lysine at the indicated residues, with or without deacetylase (Sirt3) treatment. (D) Data are expressed as the average activity after Sirt3-treatment as a percent of mock-treated enzyme, at 300 µM CoA (± SEM, N≥ 3; * indicates a significant difference from mock treatment, p< 0.05). (E) Aggregated enzyme kinetic analysis (N=4; * 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 (± SEM, N≥ 3, * indicates a significant difference from mock treatment, p< 0.05).



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

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 (K84, K187, K260, K265, K340) to glutamine to mimic the charge state of an acetylated lysine, or to arginine to mimic a deacetylated state [109]. These sites were selected based on the structure of human ACAT1 [110] and on the dynamic changes seen in both datasets. 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 to the respective glutamine mutant (K260Q, K265Q), suggesting that acetylation would decrease activity (**Figure 5B**). In contrast, K187R led to a loss of activity compared to K187Q. Interestingly, mutation of the other acetylation sites did not significantly change the activity of Acat1, further

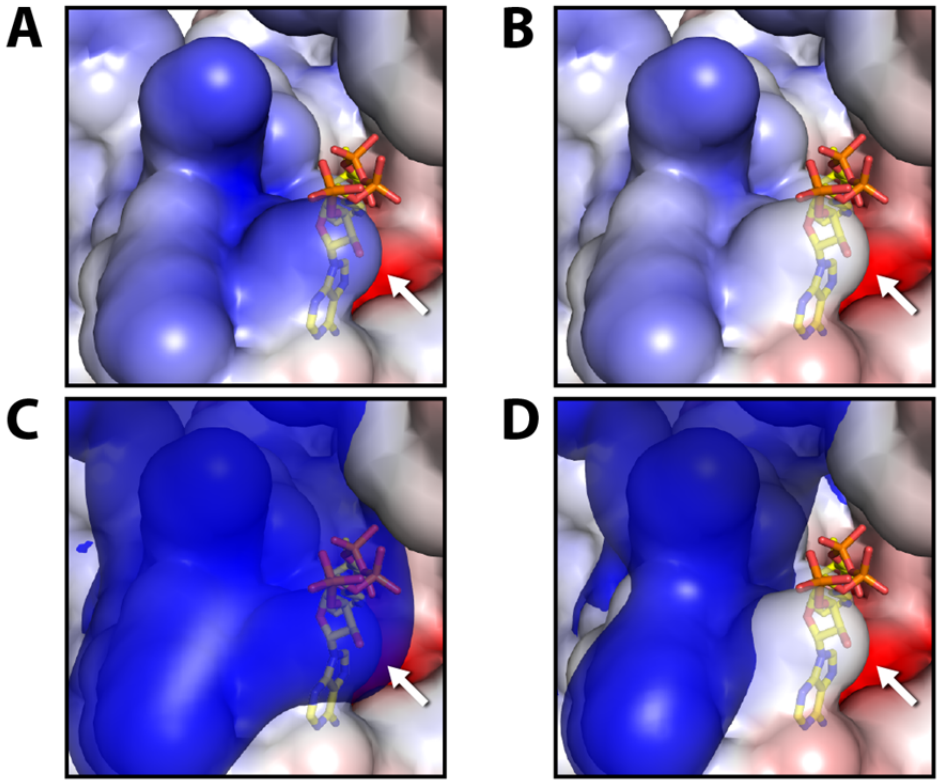
suggesting that K187, K260, K265 may be the most important acetylation sites for regulating the activity of this enzyme.

While 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 acetyl-lysine incorporation — a technique that allows for the production of a protein with nearly full acetylation occupancy on a specific lysine [111,112]. Acat1 variants were generated and purified that each possessed a single acetylated lysine at position 187, 260, or 265 (K187ac, K260ac, K265ac). Acetyl-lysine 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 (**Figure 5C, S3**). Interestingly, *in vitro* Sirt3 treatment more readily removed the acetyl groups on K260 and K265 than that on K187, consistent with the prediction that K260ac and K265ac are Sirt3 targets *in vivo*.

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 (**Figure 5D-F, S4**). 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 concentrations of substrate and determined that acetylation decreases the V_{max} of the enzyme while increasing its K_m for CoA (**Figure 5F**). As physiological levels of CoA in liver are markedly higher than the Acat1 K_m [113], these alterations in V_{max} will likely alter metabolic flux *in vivo*.

Consistent with the hypothesis that regulatory residues are likely to be conserved throughout evolution [79,114], Acat1 K260 and K265 are highly invariant and correspond to lysines 263 and 268 in the human ACAT1 ortholog (K263_Hs and K268_Hs). Recently, Haapalainen and colleagues solved the crystal structure of human ACAT1 in complex with its CoA substrate [110]. 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 (PDB # 2IBW) (**Figure 6A**), we modeled a neutralized lysine residue at K263_Hs to simulate the effect of acetylation (**Figure 6B**). We then calculated electrostatic isosurfaces extending through space at +1 kT/e with all lysines fully charged (**Figure 6C**), or with K263_Hs neutralized (**Figure 6D**). 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 K260 (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 MnSOD [115,116]. 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 6. Electrostatic modeling of acetylation effects using the crystal structure of *Homo sapiens* ACAT1 (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, while 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 overlayed with an electrostatic isosurface extending through space at +1kT/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.



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, while 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 [80]. Likewise, there is no known acetyltransferase within this organelle, although emerging evidence suggests that GCN5-L1 may be a regulator of a yet to be characterized protein acetyltransferase complex [117]. Further confounding this issue is the fact that even at low concentrations of acetyl-CoA, acetylation can occur non-enzymatically [118,119]. Coupled with the knowledge that mitochondrial enzymes tend to become acetylated following nutritional changes that produce an increase in acetyl-CoA [120], 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 non-enzymatic and, thus, likely non-regulatory. 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" [121].

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 [81], 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 (**Figure** 4C). While 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 [79,122]. 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 acetyl-lysine incorporation, *in vitro* enzymatic analyses and molecular modeling, we demonstrated that acetylation of K260 and K265 inhibits Acat1 activity by disrupting CoA binding.

Acat1 is found at the crossroads of BCAA degradation, ketogenesis, and FA 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 [123], Hmgcs2 acetylation [124] and phosphorylation [83], and Acadl acetylation [11]). Interestingly, Acat1 K265 was also recently identified as a prominent site of reversible succinylation, further suggesting that this is an unusually important site of post-translational regulation [125]. Based on our work here, we predict that this modification would likewise dampen Acat1 activity.

While each of the three pathways mentioned above occurs in the liver, they are not simultaneously induced. For example, BCAA degradation is robustly induced by refeeding, as assessed by Bckdha phosphorylation status [90 83] but FA oxidation and ketogenesis are inhibited in that same state [107]. Conversely, FA oxidation and ketogenesis are highly active in the obese state [126]. Thus, it seems reasonable to speculate that the overabundance of PTMs on ketogenesis, FAO and BCAA enzymes noted above — including those on Acat1 — operate combinatorially to help properly calibrate the individual

fluxes through these interrelated but distinct pathways. Furthermore, while 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 datasets 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 [120]. 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 K60, Gstk1 K93, Cs K52, and Tst K175) 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. Our integrated datasets are freely available through our MitoMod website (http://mitomod.biochem.wisc.edu/) — a resource that includes our previous work on mitochondrial phosphorylation under similar metabolic states [83].

Experimental Procedures

Animal models

Breeding, sacrificing, and tissue harvesting of mice were described previously for the fasting vs. refeeding study [83] and the obese/lean study [127]. Briefly, male mice were bred and housed in an environmentally controlled facility with a 12-hour light-dark cycle (6am-6pm light cycle) and provided water and standard rodent chow (Purina number 5008). For the fasting vs. refeeding study, chow was removed for 16 hours (5pm-9am), after which half of the animals were sacrificed while the other half were allowed to feed *ad libitum* an additional 2 hours before being sacrificed. Liver tissue was dissected and immediately enriched for mitochondria using differential centrifugation [83]. For the obese vs. lean study, chow was removed for 4 hours (8am-noon) 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 [81]. Briefly, mitochondrial pellets were re-solubilized, the proteins reduced and alkylated, digested enzymatically, and labeled with isobaric tags (6-plex TMT or 8-plex iTRAQ). Samples were mixed and subjected to strong cation exchange chromatography (SCX), followed by immunoprecipitation with pan-acetyl lysine 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 [82] with HCD fragmentation. Our custom software COMPASS [128] was utilized to search the MS/MS spectra against a concatenated target-decoy mouse database (UniProt), to trim the identified peptides (1% FDR), to normalize the intensities of isobaric tag reporter ions, to group all peptides to parsimonious protein groups (1% FDR), 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

Oxygen consumption rate was measured using an XF-96 Extracellular Flux Analyzer (Seahorse Bioscience) with methods based on previous work [91]. 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 minutes, 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. See the supplemental methods for further details.

Biochemical assessment of Acat1 activity

Protein Purification

Recombinant Acat1 (aa31-424) with a C-terminal hexa-histidine tag was over-expressed in *E. coli* and purified using metal affinity resin based on previous methods [110,129]. For site-specific acetyl-lysine 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 acetyl-lysine and 20 mM nicotinamide [111,112,130]. 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 OD600 of 0.5 to 0.7, then supplemented with 10 mM acetyllysine and 20 mM nicotinamide for 30 minutes before the addition of 0.3 mM IPTG. 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 acetyl-lysine. Sirt3 was purified as previously described [131].

Acat1 activity assays

Activity was assessed as described previously [110,132]. Briefly, the reaction mixture contained 50 mM Tris HCl (pH 8.1), 20 mM $MgCl_2$, 10 μM acetoacetyl-CoA, 40 mM KCl, 8 ng 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 minutes 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 five minutes while the reaction velocity curve remained linear over the first 10 minutes. K_m and V_{max} calculations were obtained using the Enzkin package in Matlab and SigmaPlot 12.

Sirt3 treatments

Sirt3 treatments were performed at 37°C for 30 minutes 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 acetyl-lysine 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 2IBW [110] 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 APBS [133] 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.

Acknowledgements

We thank AJ Bureta for assistance with **Figure** illustrations. This work was supported by a Searle Scholars Award, a Shaw Scientist Award, a Glenn Foundation Award and NIH grants RC1DK086410, U01GM094622 and R01DK098672 (to D.J.P.), a NSF graduate fellowship DGE-1256259 and NIH training grant 5T32GM007215-37 (to A.J.S.), and by NIH grants R01GM080148 (to J.J. Coon), DK066369 and DK058037 (to A.D.A.), AG038679 and GM065386 (to J.M.D.), T32DK007665 and T32GM008692 MSTP (to B.J.F.), F32DK091049 (to P.A.G.), and an AHA grant 12PRE839 (to J.J. Carson), and a NSF graduate fellowship and NIH training grant 5T32GM08349 (to K.E.D).

References

1 Colman, R. J., Anderson, R. M., Johnson, S. C., Kastman, E. K., Kosmatka, K. J., Beasley, T. M., Allison, D. B., Cruzen, C., Simmons, H. A., Kemnitz, J. W. & Weindruch, R. (2009) Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science* **325**, 201-204, doi:10.1126/science.1173635

2 Mattison, J. A., Roth, G. S., Beasley, T. M., Tilmont, E. M., Handy, A. M., Herbert, R. L., Longo, D. L., Allison, D. B., Young, J. E., Bryant, M., Barnard, D., Ward, W. F., Qi, W., Ingram, D. K. & de Cabo, R. (2012) Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study. *Nature*, doi:10.1038/nature11432

3 Weindruch, R., Walford, R. L., Fligiel, S. & Guthrie, D. (1986) The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. *J Nutr* **116**, 641-654

4 Koubova, J. & Guarente, L. (2003) How does calorie restriction work? *Genes Dev* **17**, 313-321, doi:10.1101/gad.1052903

5 Sohal, R. S. & Weindruch, R. (1996) Oxidative stress, caloric restriction, and aging. *Science* **273**, 59-63

6 Anderson, R. M. & Weindruch, R. (2012) The caloric restriction paradigm: implications for healthy human aging. *Am J Hum Biol* **24**, 101-106, doi:10.1002/ajhb.22243

7 Wallace, D. C. (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* **39**, 359-407, doi:10.1146/annurev.genet.39.110304.095751

8 Lin, S. J., Defossez, P. A. & Guarente, L. (2000) Requirement of NAD and SIR2 for life-span extension by calorie restriction in Saccharomyces cerevisiae. *Science* **289**, 2126-2128

9 Someya, S., Yu, W., Hallows, W. C., Xu, J., Vann, J. M., Leeuwenburgh, C., Tanokura, M., Denu, J. M. & Prolla, T. A. (2010) Sirt3 mediates reduction of oxidative damage and prevention of age-related hearing loss under caloric restriction. *Cell* **143**, 802-812, doi:10.1016/j.cell.2010.10.002

10 Yu, W., Dittenhafer-Reed, K. E. & Denu, J. M. (2012) SIRT3 protein deacetylates isocitrate dehydrogenase 2 (IDH2) and regulates mitochondrial redox status. *J Biol Chem* **287**, 14078-14086, doi:10.1074/jbc.M112.355206

11 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., Jr., Alt, F. W., Kahn, C. R. & Verdin, E. (2010) SIRT3 regulates mitochondrial fatty-acid oxidation by reversible enzyme deacetylation. *Nature* **464**, 121-125, doi:10.1038/nature08778

12 Lombard, D. B., Alt, F. W., Cheng, H. L., Bunkenborg, J., Streeper, R. S., Mostoslavsky, R., Kim, J., Yancopoulos, G., Valenzuela, D., Murphy, A., Yang, Y., Chen, Y., Hirschey, M. D., Bronson, R. T., Haigis, M., Guarente, L. P., Farese, R. V., Jr., Weissman, S., Verdin, E. & Schwer, B. (2007) Mammalian Sir2 homolog SIRT3 regulates global mitochondrial lysine acetylation. *Mol Cell Biol* **27**, 8807-8814, doi:10.1128/MCB.01636-07

13 Hallows, W. C., Lee, S. & Denu, J. M. (2006) Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. *Proc Natl Acad Sci U S A* **103**, 10230-10235, doi:10.1073/pnas.0604392103

14 Schwer, B., Bunkenborg, J., Verdin, R. O., Andersen, J. S. & Verdin, E. (2006) Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-CoA synthetase 2. *Proc Natl Acad Sci U S A* **103**, 10224-10229, doi:10.1073/pnas.0603968103

15 Hallows, W. C., Yu, W., Smith, B. C., Devries, M. K., Ellinger, J. J., Someya, S., Shortreed, M. R., Prolla, T., Markley, J. L., Smith, L. M., Zhao, S., Guan, K. L. & Denu, J. M. (2011) Sirt3 promotes the urea cycle and fatty acid oxidation during dietary restriction. *Mol Cell* **41**, 139-149, doi:10.1016/j.molcel.2011.01.002

16 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. (2010) SIRT3 deacetylates mitochondrial 3-hydroxy-3-methylglutaryl CoA synthase 2 and regulates ketone body production. *Cell Metab* **12**, 654-661, doi:10.1016/j.cmet.2010.11.003

17 Qiu, X., Brown, K., Hirschey, M. D., Verdin, E. & Chen, D. (2010) Calorie restriction reduces oxidative stress by SIRT3-mediated SOD2 activation. *Cell Metab* **12**, 662-667, doi:10.1016/j.cmet.2010.11.015

18 Schlicker, C., Gertz, M., Papatheodorou, P., Kachholz, B., Becker, C. F. & Steegborn, C. (2008) Substrates and regulation mechanisms for the human mitochondrial sirtuins Sirt3 and Sirt5. *J Mol Biol* **382**, 790-801, doi:10.1016/j.jmb.2008.07.048

19 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. (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* **23**, 607-618, doi:10.1016/j.molcel.2006.06.026

20 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. (2010) Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000-1004, doi:10.1126/science.1179689

21 Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V. & Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840, doi:10.1126/science.1175371

22 Pagliarini, D. J., Calvo, S. E., 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. (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-123, doi:10.1016/j.cell.2008.06.016

23 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. (2012) Proteomic Analysis of Lysine Acetylation Sites in Rat Tissues Reveals Organ Specificity and Subcellular Patterns. *Cell Rep*, doi:10.1016/j.celrep.2012.07.006

24 Henriksen, P., Wagner, S. A., Weinert, B. T., Sharma, S., Bacinskaja, G., Rehman, M., Juffer, A. H., Walther, T. C., Lisby, M. & Choudhary, C. (2012) Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in Saccharomyces cerevisiae. *Mol Cell Proteomics*, doi:10.1074/mcp.M112.017251

25 Chen, Y., Zhao, W., Yang, J. S., Cheng, Z., Luo, H., Lu, Z., Tan, M., Gu, W. & Zhao, Y. (2012) Quantitative acetylome analysis reveals the roles of SIRT1 in regulating diverse substrates and cellular pathways. *Mol Cell Proteomics*, doi:10.1074/mcp.M112.019547

26 Fritz, K. S., Galligan, J. J., Hirschey, M. D., Verdin, E. & Petersen, D. R. (2012) Mitochondrial acetylome analysis in a mouse model of alcohol-induced liver injury utilizing SIRT3 knockout mice. *J Proteome Res* **11**, 1633-1643, doi:10.1021/pr2008384

27 Kendrick, A. A., Choudhury, M., Rahman, S. M., McCurdy, C. E., Friederich, M., Van Hove, J. L., Watson, P. A., Birdsey, N., Bao, J., Gius, D., Sack, M. N., Jing, E., Kahn, C. R., Friedman, J. E. & Jonscher, K. R. (2011) Fatty liver is associated with reduced SIRT3 activity and mitochondrial protein hyperacetylation. *Biochem J* **433**, 505-514, doi:10.1042/BJ20100791

28 Law, I. K., Liu, L., Xu, A., Lam, K. S., Vanhoutte, P. M., Che, C. M., Leung, P. T. & Wang, Y. (2009) Identification and characterization of proteins interacting with SIRT1 and SIRT3: implications in the anti-aging and metabolic effects of sirtuins. *Proteomics* **9**, 2444-2456, doi:10.1002/pmic.200800738

29 Schwer, B., Eckersdorff, M., Li, Y., Silva, J. C., Fermin, D., Kurtev, M. V., Giallourakis, C., Comb, M. J., Alt, F. W. & Lombard, D. B. (2009) Calorie restriction alters mitochondrial protein acetylation. *Aging Cell* **8**, 604-606, doi:10.1111/j.1474-9726.2009.00503.x

30 Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. & Coon, J. J. (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Meth* **8**, 933-935, doi:http://www.nature.com/nmeth/journal/v8/n11/abs/nmeth.1716.html#supplementary-information

31 Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. (2011) COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* **11**, 1064-1074, doi:10.1002/pmic.201000616

32 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. (2011) Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Meth* **8**, 821-827, doi:http://www.nature.com/nmeth/journal/v8/n10/abs/nmeth.1699.html#supplementary-information

33 Wu, R., Dephoure, N., Haas, W., Huttlin, E. L., Zhai, B., Sowa, M. E. & Gygi, S. P. (2011) Correct Interpretation of Comprehensive Phosphorylation Dynamics Requires Normalization by Protein Expression Changes. *Molecular & Cellular Proteomics* **10**, doi:10.1074/mcp.M111.009654

34 Guan, K. L., Yu, W., Lin, Y., Xiong, Y. & Zhao, S. (2010) Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides. *Nat Protoc* **5**, 1583-1595, doi:10.1038/nprot.2010.117

35 Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences* **104**, 5860-5865, doi:10.1073/pnas.0608638104

36 Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistcal Learning*. (Springer, 2009).

37 Smith, B. C., Settles, B., Hallows, W. C., Craven, M. W. & Denu, J. M. (2011) SIRT3 substrate specificity determined by peptide arrays and machine learning. *ACS Chem Biol* **6**, 146-157, doi:10.1021/cb100218d

38 Jin, L., Wei, W., Jiang, Y., Peng, H., Cai, J., Mao, C., Dai, H., Choy, W., Bemis, J. E., Jirousek, M. R., Milne, J. C., Westphal, C. H. & Perni, R. B. (2009) Crystal structures of human SIRT3 displaying substrate-induced conformational changes. *J Biol Chem* **284**, 24394-24405, doi:10.1074/jbc.M109.014928

39 Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202, doi:10.1006/jmbi.1999.3091

40 Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D. & Morris, Q. (2010) The GeneMANIA prediction server: biological network

integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-220, doi:10.1093/nar/gkq537

41 Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431-432, doi:10.1093/bioinformatics/btq675

42 Bathe, O. F., Shaykhutdinov, R., Kopciuk, K., Weljie, A. M., McKay, A., Sutherland, F. R., Dixon, E., Dunse, N., Sotiropoulos, D. & Vogel, H. J. (2011) Feasibility of identifying pancreatic cancer based on serum metabolomics. *Cancer Epidemiol Biomarkers Prev* **20**, 140-147, doi:10.1158/1055-9965.EPI-10-0712

43 Janes, K. A. & Yaffe, M. B. (2006) Data-driven modelling of signal-transduction networks. *Nat Rev Mol Cell Biol* **7**, 820-828, doi:10.1038/nrm2041

44 Hagopian, K., Ramsey, J. J. & Weindruch, R. (2003) Caloric restriction increases gluconeogenic and transaminase enzyme activities in mouse liver. *Exp Gerontol* **38**, 267-278

45 DeLaBarre, B., Gross, S., Fang, C., Gao, Y., Jha, A., Jiang, F., Song, J. J., Wei, W. & Hurov, J. B. (2011) Full-length human glutaminase in complex with an allosteric inhibitor. *Biochemistry-Us* **50**, 10764-10770, doi:10.1021/bi201613d

46 Fontaine, M., Briand, G., Ser, N., Armelin, I., Rolland, M. O., Degand, P. & Vamecq, J. (1996) Metabolic studies in twin brothers with 2-methylacetoacetyl-CoA thiolase deficiency. *Clin Chim Acta* **255**, 67-83

47 Haapalainen, A. M., Merilainen, G., Pirila, P. L., Kondo, N., Fukao, T. & Wierenga, R. K. (2007) Crystallographic and kinetic studies of human mitochondrial acetoacetyl-CoA thiolase: the importance of potassium and chloride ions for its structure and function. *Biochemistry-Us* **46**, 4305-4321, doi:10.1021/bi6026192

48 Dombrovski, L., Min, J. R., Antoshenko, T., Wu, H., Loppnau, P., Edwards, A. M., Arrowsmith, C. H., Bochkarve, A. & Plotnikov, A. N. (RCSB, 2011).

49 Lin, W. D., Wang, C. H., Lee, C. C., Lai, C. C., Tsai, Y. & Tsai, F. J. (2007) Genetic mutation profile of isovaleric acidemia patients in Taiwan. *Mol Genet Metab* **90**, 134-139, doi:10.1016/j.ymgme.2006.08.011

50 Tiffany, K. A., Roberts, D. L., Wang, M., Paschke, R., Mohsen, A. W., Vockley, J. & Kim, J. J. (1997) Structure of human isovaleryl-CoA dehydrogenase at 2.6 A resolution: structural basis for substrate specificity. *Biochemistry-Us* **36**, 8455-8464, doi:10.1021/bi970422u

51 Civitarese, A. E., Carling, S., Heilbronn, L. K., Hulver, M. H., Ukropcova, B., Deutsch, W. A., Smith, S. R. & Ravussin, E. (2007) Calorie restriction increases muscle mitochondrial biogenesis in healthy humans. *Plos Med* **4**, 485-494, doi:Artn E76
Doi 10.1371/Journal.Pmed.0040076

52 Tibbetts, A. S. & Appling, D. R. (2010) Compartmentalization of Mammalian Folate-Mediated One-Carbon Metabolism. *Annu Rev Nutr* **30**, 57-81, doi:Doi 10.1146/Annurev.Nutr.012809.104810

53 Lopez-Lluch, G., Hunt, N., Jones, B., Zhu, M., Jamieson, H., Hilmer, S., Cascajo, M. V., Allard, J., Ingram, D. K., Navas, P. & de Cabo, R. (2006) Calorie restriction induces mitochondrial biogenesis and bioenergetic efficiency. *Proc Natl Acad Sci U S A* **103**, 1768-1773, doi:Doi 10.1073/Pnas.0510452103

54 Plank, M., Wuttke, D., van Dam, S., Clarke, S. A. & de Magalhaes, J. P. (2012) A meta-analysis of caloric restriction gene expression profiles to infer common signatures and regulatory mechanisms. *Mol Biosyst* **8**, 1339-1349, doi:10.1039/c2mb05255e

55 Miller, B. F., Robinson, M. M., Bruss, M. D., Hellerstein, M. & Hamilton, K. L. (2012) A comprehensive assessment of mitochondrial protein synthesis and cellular proliferation with age and caloric restriction. *Aging Cell* **11**, 150-161, doi:10.1111/j.1474-9726.2011.00769.x

56 Richardson, D. R., Lane, D. J., Becker, E. M., Huang, M. L., Whitnall, M., Suryo Rahmanto, Y., Sheftel, A. D. & Ponka, P. (2010) Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol. *Proc Natl Acad Sci U S A* **107**, 10775-10782, doi:10.1073/pnas.0912925107

57 Ngo, H. B., Kaiser, J. T. & Chan, D. C. (2011) The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat Struct Mol Biol* **18**, 1290-U1141, doi:Doi 10.1038/Nsmb.2159

58 Campbell, C. T., Kolesar, J. E. & Kaufman, B. A. (2012) Mitochondrial transcription factor A regulates mitochondrial transcription initiation, DNA packaging, and genome copy number. *Biochim Biophys Acta*, doi:10.1016/j.bbagrm.2012.03.002

59 Wang, Z., Cotney, J. & Shadel, G. S. (2007) Human mitochondrial ribosomal protein MRPL12 interacts directly with mitochondrial RNA polymerase to modulate mitochondrial gene expression. *J Biol Chem* **282**, 12610-12618, doi:10.1074/jbc.M700461200

60 Vedrenne, V., Galmiche, L., Chretien, D., de Lonlay, P., Munnich, A. & Rotig, A. (2012) Mutation in the mitochondrial translation elongation factor EFTs results in severe infantile liver failure. *J Hepatol* **56**, 294-297, doi:10.1016/j.jhep.2011.06.014

61 Jeppesen, M. G., Navratil, T., Spremulli, L. L. & Nyborg, J. (2005) Crystal structure of the bovine mitochondrial elongation factor Tu.Ts complex. *J Biol Chem* **280**, 5071-5081, doi:10.1074/jbc.M411782200

62 Finley, L. W., Carracedo, A., Lee, J., Souza, A., Egia, A., Zhang, J., Teruya-Feldstein, J., Moreira, P. I., Cardoso, S. M., Clish, C. B., Pandolfi, P. P. & Haigis, M. C. (2011) SIRT3 opposes reprogramming of cancer cell metabolism through HIF1alpha destabilization. *Cancer Cell* **19**, 416-428, doi:10.1016/j.ccr.2011.02.014

63 Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**, 786-787, doi:10.1038/nmeth1109-786

64 Allfrey, V. G., Faulkner, R. & Mirsky, A. E. (1964) 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

65 Phillips, D. M. (1963) The presence of acetyl groups of histones. *Biochem J* **87**, 258-263

66 Gnad, F., Forner, F., Zielinska, D. F., Birney, E., Gunawardena, J. & Mann, M. (2010) Evolutionary constraints of phosphorylation in eukaryotes, prokaryotes, and mitochondria. *Molecular & cellular proteomics : MCP* **9**, 2642-2653, doi:10.1074/mcp.M110.001594

67 Norvell, A. & McMahon, S. B. (2010) Cell biology. Rise of the rival. *Science* **327**, 964-965, doi:10.1126/science.1187159

68 Guan, K.-L. & Xiong, Y. (2011) Regulation of intermediary metabolism by protein acetylation. *Trends in biochemical sciences* **36**, 108-116, doi:10.1016/j.tibs.2010.09.003

69 Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H. & Xiao, L. (2006) Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey. *Molecular cell* **23**, 607-618, doi:10.1016/j.molcel.2006.06.026

70 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. (2012) Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns. *Cell reports* **2**, 419-431, doi:10.1016/j.celrep.2012.07.006

71 Anderson, K. A. & Hirschey, M. D. (2012) Mitochondrial protein acetylation regulates metabolism. *Essays in biochemistry* **52**, 23-35, doi:10.1042/bse0520023

72 Vadvalkar, S. S., Baily, C. N., Matsuzaki, S., West, M., Tesiram, Y. A. & Humphries, K. M. (2012) Metabolic inflexibility and protein lysine acetylation in heart mitochondria of a chronic model of Type 1 diabetes. *The Biochemical journal* **449**, 253-261, doi:10.1042/BJ20121038

73 Lombard, D. B., Tishkoff, D. X. & Bao, J. (2011) Mitochondrial sirtuins in the regulation of mitochondrial activity and metabolic adaptation. *Handbook of experimental pharmacology* **206**, 163-188, doi:10.1007/978-3-642-21631-2_8

74 Vaitheesvaran, B., Yang, L., Hartil, K., Glaser, S., Yazulla, S., Bruce, J. E. & Kurland, I. J. (2012) Peripheral Effects of FAAH Deficiency on Fuel and Energy Homeostasis: Role of Dysregulated Lysine Acetylation. *PLoS ONE* **7**, e33717, doi:10.1371/journal.pone.0033717.t003

75 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. (2010) SIRT3 Is a Mitochondria-Localized Tumor Suppressor Required for Maintenance of Mitochondrial Integrity and Metabolism during Stress. *Cancer Cell* **17**, 41-52, doi:10.1016/j.ccr.2009.11.023

76 Hirschey, M. D., Lu, G., 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, J., Robert V, Kahn, C. R. & Verdin, E. (2011) SIRT3 Deficiency and Mitochondrial Protein Hyperacetylation Accelerate the Development of the Metabolic Syndrome. *Molecular cell* **44**, 177-190, doi:10.1016/j.molcel.2011.07.019

77 Pereira, C. V., Lebiedzinska, M., Wieckowski, M. R. & Oliveira, P. J. (2012) Regulation and protection of mitochondrial physiology by sirtuins. *MITOCH* **12**, 66-76, doi:10.1016/j.mito.2011.07.003

78 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. (2013) Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways. *Proc Natl Acad Sci U S A*, doi:10.1073/pnas.1302961110

79 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. (2012) Systematic functional prioritization of protein posttranslational modifications. *Cell* **150**, 413-425, doi:10.1016/j.cell.2012.05.036

80 He, W., Newman, J. C., Wang, M. Z., Ho, L. & Verdin, E. (2012) Mitochondrial sirtuins: regulators of protein acylation and metabolism. *Trends in Endocrinology & Metabolism* **23**, 467-476, doi:10.1016/j.tem.2012.07.004

81 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. (2013) Calorie Restriction and SIRT3 Trigger Global Reprogramming of the Mitochondrial Protein Acetylome. *Molecular cell* **49**, 186-199, doi:10.1016/j.molcel.2012.10.024

82 Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. & Coon, J. J. (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature Methods* **8**, 933-935, doi:10.1038/nmeth.1716

83 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. (2012) A Quantitative Map of the Liver Mitochondrial Phosphoproteome Reveals Posttranslational Control of Ketogenesis. *Cell Metabolism* **16**, 672-683, doi:10.1016/j.cmet.2012.10.004

84 de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. (2004) Open source clustering software. *Bioinformatics* **20**, 1453-1454, doi:10.1093/bioinformatics/bth078

85 Saldanha, A. J. (2004) Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248, doi:10.1093/bioinformatics/bth349

86 Locasale, J. W. & Cantley, L. C. (2011) Metabolic Flux and the Regulation of Mammalian Cell Growth. *Cell Metabolism* **14**, 443-451, doi:10.1016/j.cmet.2011.07.014

87 Baez-Ruiz, A., Escobar, C., Aguilar-Roblero, R., Vazquez-Martinez, O. & Diaz-Munoz, M. (2005) Metabolic adaptations of liver mitochondria during restricted feeding schedules. *American journal of physiology. Gastrointestinal and liver physiology* **289**, G1015-1023, doi:10.1152/ajpgi.00488.2004

88 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. (2010) Regulation of Cellular Metabolism by Protein Lysine Acetylation. *Science (New York, NY)* **327**, 1000-1004, doi:10.1126/science.1179689

89 Longo, N., Amat di San Filippo, C. & Pasquali, M. (2006) Disorders of carnitine transport and the carnitine cycle. *American journal of medical genetics. Part C, Seminars in medical genetics* **142C**, 77-85, doi:10.1002/ajmg.c.30087

90 Adeva, M. M., Calviño, J., Souto, G. & Donapetry, C. (2011) Insulin resistance and the metabolism of branched-chain amino acids in humans. *Amino Acids* **43**, 171-181, doi:10.1007/s00726-011-1088-7

91 Rogers, G. W., Brand, M. D., Petrosyan, S., Ashok, D., Elorza, A. A., Ferrick, D. A. & Murphy, A. N. (2011) High throughput microplate respiratory measurements using minimal quantities of isolated mitochondria. *PLoS One* **6**, e21746, doi:10.1371/journal.pone.0021746

92 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. (2011) Fatty liver is associated with reduced SIRT3 activity and mitochondrial protein hyperacetylation. *The Biochemical journal* **433**, 505-514, doi:10.1210/jc.2003-030202

93 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. (2008) A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome research* **18**, 706-716, doi:10.1101/gr.074914.107

94 Buchner, D. A., Yazbek, S. N., Solinas, P., Burrage, L. C., Morgan, M. G., Hoppel, C. L. & Nadeau, J. H. (2009) Increased Mitochondrial Oxidative Phosphorylation in the Liver Is Associated With Obesity and Insulin Resistance. *Obesity* **19**, 917-924, doi:10.1038/oby.2010.214

95 Finley, L. W. S., Haas, W., Desquiret-Dumas, V., Wallace, D. C., Procaccio, V., Gygi, S. P. & Haigis, M. C. (2011) Succinate Dehydrogenase Is a Direct Target of Sirtuin 3 Deacetylase Activity. *PLoS ONE* **6**, e23295, doi:10.1371/journal.pone.0023295.g004

96 Ahn, B.-H., Kim, H.-S., Song, S., Lee, I. H., Liu, J., Vassilopoulos, A., Deng, C.-X. & Finkel, T. (2008) 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, doi:10.1073/pnas.0803790105

97 Bao, J., Scott, I., Lu, Z., Pang, L., Dimond, C. C., Gius, D. & Sack, M. N. (2010) SIRT3 is regulated by nutrient excess and modulates hepatic susceptibility to lipotoxicity. *Free Radical Biology and Medicine* **49**, 1230-1237, doi:10.1016/j.freeradbiomed.2010.07.009

98 Cimen, H., Han, M.-J., Yang, Y., Tong, Q., Koc, H. & Koc, E. C. (2010) Regulation of Succinate Dehydrogenase Activity by SIRT3 in Mammalian Mitochondria. *Biochemistry* **49**, 304-311, doi:10.1021/bi901627u

99 Han, Q., Cai, T., Tagle, D. A., Robinson, H. & Li, J. (2008) Substrate specificity and structure of human aminoadipate aminotransferase/kynurenine aminotransferase II. *Bioscience reports* **28**, 205-215, doi:10.1042/BSR20080085

100 Rossi, F., Garavaglia, S., Montalbano, V., Walsh, M. A. & Rizzi, M. (2008) Crystal structure of human kynurenine aminotransferase II, a drug target for the treatment of schizophrenia. *The Journal of biological chemistry* **283**, 3559-3566, doi:10.1074/jbc.M707925200

101 Han, Q., Robinson, H. & Li, J. (2008) Crystal structure of human kynurenine aminotransferase II. *The Journal of biological chemistry* **283**, 3567-3573, doi:10.1074/jbc.M708358200

102 Oxenkrug, G. F. (2010) Metabolic syndrome, age-associated neuroendocrine disorders, and dysregulation of tryptophan-kynurenine metabolism. *Annals of the New York Academy of Sciences* **1199**, 1-14, doi:10.1111/j.1749-6632.2009.05356.x

103 Noland, R. C., Woodlief, T. L., Whitfield, B. R., Manning, S. M., Evans, J. R., Dudek, R. W., Lust, R. M. & Cortright, R. N. (2007) Peroxisomal-mitochondrial oxidation in a rodent model of obesity-associated insulin resistance. *Am J Physiol Endocrinol Metab* **293**, E986-E1001, doi:10.1152/ajpendo.00399.2006

104 Stein, L. R. & Imai, S. (2012) The dynamic regulation of NAD metabolism in mitochondria. *Trends in endocrinology and metabolism: TEM* **23**, 420-428, doi:10.1016/j.tem.2012.06.005

105 Feldman, J. L., Dittenhafer-Reed, K. E. & Denu, J. M. (2012) Sirtuin Catalysis and Regulation. *Journal of Biological Chemistry* **287**, 42419-42427, doi:10.1074/jbc.R112.378877

106 Williamson, D. H., Lund, P. & Krebs, H. A. (1967) The redox state of free nicotinamide-adenine dinucleotide in the cytoplasm and mitochondria of rat liver. *Biochem J* **103**, 514-527

107 McGarry, J. D. & Foster, D. W. (1980) Regulation of hepatic fatty acid oxidation and ketone body production. *Annual Review of Biochemistry* **49**, 395-420, doi:10.1146/annurev.bi.49.070180.002143

108 Korman, S. H. (2006) Inborn errors of isoleucine degradation: a review. *Molecular genetics and metabolism* **89**, 289-299, doi:10.1016/j.ymgme.2006.07.010

109 Megee, P. C., Morgan, B. A., Mittman, B. A. & Smith, M. M. (1990) Genetic analysis of histone H4: essential role of lysines subject to reversible acetylation. *Science (New York, NY)* **247**, 841-845

110 Haapalainen, A. M., Meriläinen, G., Pirilä, P. L., Kondo, N., Fukao, T. & Wierenga, R. K. (2007) 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, doi:10.1021/bi6026192

111 Neumann, H., Peak-Chew, S. Y. & Chin, J. W. (2008) Genetically encoding Nε-acetyllysine in recombinant proteins. *Nature chemical biology* **4**, 232-234, doi:10.1038/nchembio.73

112 Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D. & Chin, J. W. (2009) A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Molecular cell* **36**, 153-163, doi:10.1016/j.molcel.2009.07.027

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

114 Boekhorst, J., van Breukelen, B., Heck, A. J. & Snel, B. (2008) Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biology* **9**, R144, doi:10.1186/gb-2008-9-10-r144

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

116 Tao, R., Coleman, M. C., Pennington, J. D., Ozden, O., Park, S.-H., Jiang, H., Kim, H.-S., Flynn, C. R., Hill, S., McDonald, W. H., Olivier, A. K., Spitz, D. R. & Gius, D. (2010) Sirt3-Mediated Deacetylation of Evolutionarily Conserved Lysine 122 Regulates MnSOD Activity in Response to Stress. *Molecular cell* **40**, 893-904, doi:10.1016/j.molcel.2010.12.013

117 Scott, I., Webster, B. R., Li, J. H. & Sack, M. N. (2012) Identification of a molecular component of the mitochondrial acetyltransferase programme: a novel role for GCN5L1. *Biochem J* **443**, 655-661, doi:10.1042/BJ20120118

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

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

120 Newman, J. C., He, W. & Verdin, E. (2012) Mitochondrial Protein Acylation and Intermediary Metabolism: Regulation by Sirtuins and Implications for Metabolic Disease. *Journal of Biological Chemistry* **287**, 42436-42443, doi:10.1074/jbc.R112.404863

121 Covian, R. & Balaban, R. S. (2012) Cardiac mitochondrial matrix and respiratory complex protein phosphorylation. *American journal of physiology. Heart and circulatory physiology* **303**, H940-966, doi:10.1152/ajpheart.00077.2012

122 Xiong, Y. & Guan, K. L. (2012) Mechanistic insights into the regulation of metabolic enzymes by acetylation. *The Journal of cell biology* **198**, 155-164, doi:10.1083/jcb.201202056

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

124 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. (2010) SIRT3 deacetylates mitochondrial 3-hydroxy-3-methylglutaryl CoA synthase 2 and regulates ketone body production. *Cell Metab* **12**, 654-661, doi:10.1016/j.cmet.2010.11.003

125 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. (2013) SIRT5-Mediated Lysine Desuccinylation Impacts Diverse Metabolic Pathways. *Molecular cell* **50**, 919-930

126 Newgard, C. B. (2012) Interplay between Lipids and Branched-Chain Amino Acids in Development of Insulin Resistance. *Cell Metabolism* **15**, 606-614, doi:10.1016/j.cmet.2012.01.024

127 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. (2009) Obesity and genetics regulate microRNAs in islets, liver, and adipose of diabetic mice. *Mammalian Genome* **20**, 476-485, doi:10.1007/s00335-009-9217-2

128 Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. (2011) COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *PROTEOMICS* **11**, 1064-1074, doi:10.1002/pmic.201000616

129 Blommel, P. G., Becker, K. J., Duvnjak, P. & Fox, B. G. (2007) Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnology progress* **23**, 585-598, doi:10.1021/bp070011x

130 Yu, W., Dittenhafer-Reed, K. E. & Denu, J. M. (2012) SIRT3 protein deacetylates isocitrate dehydrogenase 2 (IDH2) and regulates mitochondrial redox status. *The Journal of biological chemistry* **287**, 14078-14086, doi:10.1074/jbc.M112.355206

131 Hallows, W. C., Lee, S. & Denu, J. M. (2006) Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. *Proc Natl Acad Sci U S A* **103**, 10230-10235, doi:10.1073/pnas.0604392103

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

133 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041, doi:10.1073/pnas.181342398

Chapter 7

The One Hour Yeast Proteome

ASH designed research, performed experiments, analyzed data, and wrote the paper.

This chapter has been published:

Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S. & Coon, J. J. (2013) The One Hour Yeast Proteome. *Mol. Cell. Proteomics*, doi:10.1074/mcp.M113.034769

*co-first author

Abstract

We describe the comprehensive analysis of the yeast proteome in just over one hour of optimized analysis. We achieve this expedited proteome characterization with improved sample preparation, chromatographic separations, and by using a new Orbitrap hybrid mass spectrometer equipped with a mass filter, a collision cell, a high-field Orbitrap analyzer, and, finally, a dual cell linear ion trap analyzer (Q-OT-qIT, Orbitrap Fusion). This system offers high MS^2 acquisition speed of 20 Hz and detects up to 19 peptide sequences within a single second of operation. Over a 1.3 hour chromatographic method, the Q-OT-qIT hybrid collected an average of 13,447 MS^1 and 80,460 MS^2 scans (per run) to produce 43,400 ($\overline{x}$) peptide spectral matches and 34,255 ($\overline{x}$) peptides with unique amino acid sequences (1% FDR). On average, each one hour analysis achieved detection of 3,977 proteins (1% FDR). We conclude that further improvements in mass spectrometer scan rate could render comprehensive analysis of the human proteome within a few hours.

Introduction

The ability to measure differences in protein expression has become key to understanding biological phenomena[1,2]. Owing to cost, speed, and accessibility, transcriptomic analysis is often used as a proteomic proxy[3,4]. That said, mRNA is a genetic intermediary and cannot inform on the myriad of post-translational regulation processes[5-7]. For the past decade considerable effort has been invested in maturing proteomic technology to deliver information at a rate and cost commensurate to transcriptomic technologies.

Historically yeast, with its 6,600 open reading frames, has been the preferred proteomic technology test-bed[8]. In 2003, Weissmann and colleagues measured approximate expression levels of each yeast gene using either GFP or TAP tags[9]. This seminal work established that ~ 4,500 proteins are expressed during log-phase yeast growth. Subsequent mass spectrometry-based studies have confirmed this early estimate[9-12]. With this knowledge, we hereby define comprehensive proteome analysis as an experiment that detects ~ 90% of the expressed proteome ($\geq$ 4,000 proteins for yeast). Note others have used the term "nearly-complete" for this purpose; we posit that comprehensive has identical meaning (*i.e.*, including many, most, or all things)[13].

Initial MS-based proteomic analyses of yeast, each identifying up to a few hundred proteins, were conducted using a variety of separation and MS technologies[14-16]. Yates and co-workers reported the first large-scale yeast proteome study in 2001 with the identification of 1,483 proteins following ~ 68 hours of mass spectral analysis, *i.e.*, 0.4 proteins were identified per minute[17]. Their method – two dimensional chromatography coupled with tandem mass spectrometry – has provided a template for large-scale protein analysis for the past decade[18-20]. By incorporating an offline first dimension of separation with more extensive fractionation (80 *vs.* 15) Gygi *et al.* expanded upon this work in 2003[21]. That said, the modest increase in identified proteins (1,504) required 135 hours of analysis, reducing the protein per minute count to 0.2. Armed with a faster hybrid mass spectrometer capable of accurate mass measurement,

Mann and colleagues achieved detection of 2,003 yeast proteins in an impressive 48 hours (0.7 proteins/minute) in 2006[22].

From these three pioneering studies we begin to see the impact of mass spectrometer acquisition rate on the depth and rate of proteome analysis. The most recent application of such technology to the yeast proteome, however, The Mann work utilized a hybrid linear ion trap-ion cyclotron resonance Fourier transform instrument (LTQ-FT) that delivered MS^2 scans at a rate of ~ 650 ms[23]. The earlier studies, *i.e.*, Yates and Gygi, relied upon the considerably slower scanning (1-3 seconds/scan) three-dimensional ion trap technology. In 2008, using the novel Orbitrap hybrid mass spectrometer, Mann and colleagues reported on the first comprehensive analysis of the yeast proteome by identifying nearly 4,000 proteins[10]. Extensive fractionation (24) and triplicate analysis of each fraction rendered the study a considerable time investment at ~ 144 analysis hours (0.5 proteins/minute). In 2010 our group achieved similar comprehensive analysis, but improved sequence coverage, using by use of fractionation and multiple proteases[24]. That work, however, required even longer analysis time (0.2 proteins/min).

And that was the state-of-the-art as recently as three years ago. Doubtless we, the proteomics community, had achieved one momentous goal – comprehensive coverage of the yeast proteome. Still, obtaining this depth was not routine as it mandated days of MS analysis and a considerable amount of expert labor. In 2012, with new, even faster scanning, quadrupole-Orbitrap technology (Q-OT, Q-Exactive), Mann and colleagues dispatched the concept of fractionation, improved the quality of sample preparation, and placed emphasis on higher quality online separations[25]. With their streamlined method they achieved detection of just over 3,900 yeast proteins following four hours of MS analysis. Even more impressive this strategy translated to the identification of 16.3 proteins per minute – a 33-fold improvement over the next best comprehensive study. This success was a remarkable achievement and illustrates that comprehensive proteomic technology can indeed be executed in a time efficient manner.

Time-of-flight hybrid systems, of course, can deliver very high MS^2 acquisition rates, up to 100 Hz in some reports. In 2011, Muddiman and colleagues reported yeast proteome analysis using a

quadrupole-TOF system (*i.e.*, TripleTOF) operating at a much lower rate (20 Hz) MS^2 scan rate[26]. Even at this reduced rate, only 16% of the spectra were mapped to unique sequences and 1,112 unique proteins identified. Due to reduced MS^2 spectral quality (*i.e.,* low signal-to-noise, S/N), even fewer unique peptide identifications were achieved at higher MS^2 acquisition rates. Other studies using TOF technologies report similar results [27,28]. For maximal proteome depth, we conclude that increased scan speed must not come at the cost of reduced spectral quality. Recently, a new Orbitrap hybrid mass spectrometer having a mass filter, a collision cell, a high-field Orbitrap analyzer, and, finally, a dual cell linear ion trap analyzer was described (Q-OT-qIT, Orbitrap Fusion)[29,30]. This system offers high MS^2 acquisition speed of 20 Hz – double that of the Q-OT system used by Mann and colleagues. We postulated that this fresh system, with its fast scan rate, could provide comprehensive proteome analysis in record time. To maximize performance we developed an optimized cellular lysis approach, employed trypsin digestion, and used dimethyl sulfoxide (DMSO, 5%) as an LC additive to increase abundance of acidic peptides and unify charge state[31,32]. Using this novel system we report the comprehensive analysis of the yeast proteome (4,002 @ 1% FDR) following 1.3 hours of nLC-MS^2 analysis (70 minute gradient). These experiments delivered an extraordinary 67 proteins per minute and demonstrate that complete analysis of the yeast proteome can be routinely performed in approximately one hour.

Results

Considerable gains in the depth and rate of proteomic analysis have been realized over the past decade (*vide supra*). These improved results stem from routine use of high mass accuracy and resolution, but also from a steady increase in MS^2 acquisition rate. In the decade spanning the seminal Yates publication in 2001 and the single-shot proteome work of Mann *et al.* in 2012, MS^2 sampling rates rose from ~ 0.75 Hz to nearly 10 Hz[17,25]. Here we report on an even newer generation of mass spectrometer that comprises a mass resolving quadrupole, Orbitrap, collision cell, and linear ion trap (Q-OT-qIT, Fusion, **Figure 1**)[29,30]. In this system, MS acquisition rate is not only boosted by the presence of a very

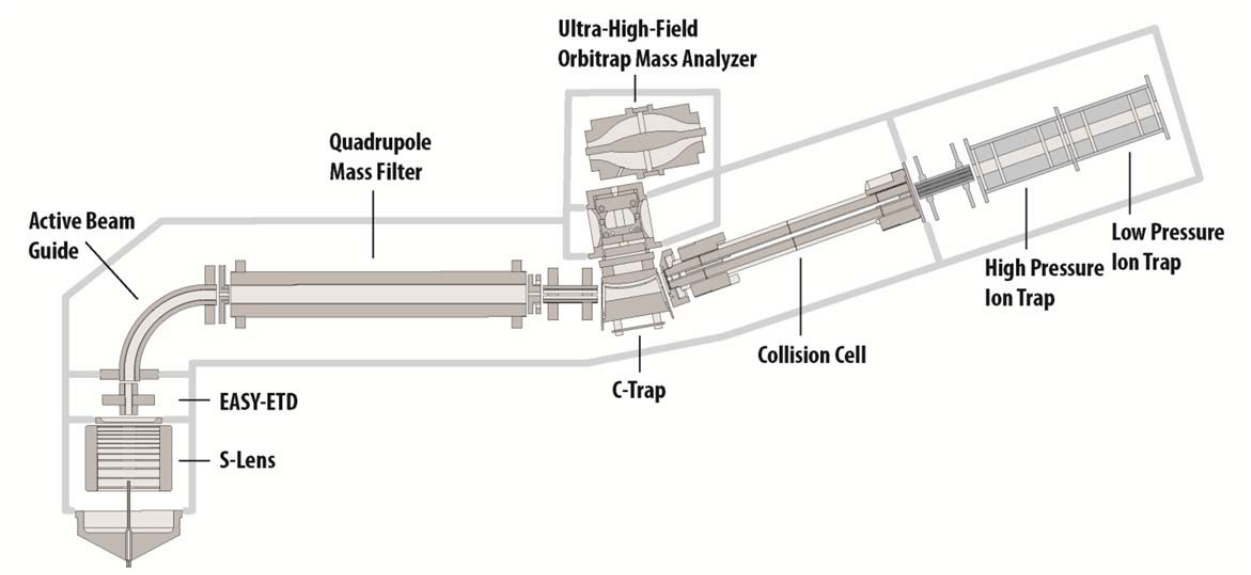
fast scanning dual cell linear ion trap, but also by a control environment having multiple, independent processing units. The new system is equipped with a sophisticated control system that parallelizes the processes of ion injection, precursor isolation, fragmentation, and mass analysis to achieve a ~2X boost in acquisition rates. We reasoned that this Q-OT-qIT configuration, with its 20 Hz MS^2 acquisition rate, could afford a considerable gain for rapid, whole proteome analysis.

The Q-OT-qIT system

To test this hypothesis we began by performing a parametric evaluation using a complex mixture of yeast tryptic peptides eluted into the system over a 70 minute gradient. We examined several settings including collisional activation mode (*i.e.*, HCD or trap CAD), MS^1 resolution, collision energy, maximum inject time, and dynamic exclusion settings. Briefly, we found that MS^2 analysis using HCD followed by ion trap mass analysis (low-res HCD; 80,626 MS^2 events with 33,127 unique PSMs) generated more identifications compared to ion trap CAD with ion trap mass analysis (CAD; 75,973 MS^2 events with 31,820 unique PSMs). This is not surprising as HCD tends to offer more random backbone fragmentation and, with the Q-OT-qIT geometry, can be accomplished slightly faster. Operation of the system with an MS^1 resolving power setting of 60,000 (@ *m/z* 200) afforded a 20% increase in detected unique peptides over 15,000 resolving power. We conclude the boosted resolving power elevates precursor signal-to-noise (S/N) ratios, allowing for improved selection of low abundance precursors, and can potentially separate otherwise unresolved precursors so that multiple MS^2 events can be acquired. Of course, in this scenario such closely spaced precursors would be co-isolated (0.7 *m/z* isolation width); however, the selected precursor *m/z* annotated in the MS^2 scan would be different and would facilitate identification from a chimeric MS^2 scan. Increasing MS^1 resolving power above 60,000 did not provide any apparent benefit for increased identifications. 35 msec was the optimal maximum injection time. Decreasing the maximum injection time to 30 msec, or increasing it to 45 msec caused a 10% decrease in peptide identifications. We found only slight variations in peptide identifications between dynamic

exclusion settings of 30, 45, and 60 seconds. Quadrupole isolation widths from 0.5 to 1.5 *m/z* were examined, the best results were achieved at a value of 0.7 *m/z*. MS^1 and MS^2 automatic gain control (AGC) target values of 500,000 and 7,000, respectively, produced the maximum number of peptide identifications.

Figure 1. Schematic of the Q-OT-qIT hybrid mass spectrometer (Fusion). The system differs from previous generations of quadrupole ion trap/Orbitrap hybrids by introduction of a resolving quadrupole mass filter and rearrangement of the geometry to place the linear ion traps to the rear of the collision cell. The reconfigured geometry relieves the linear ion trap of two of its former functions – precursor ion isolation and dissociation. The consequence is both improved and faster operation.



Lysis, chromatography, and additives

Yeast cell lysis is a critical step in achieving comprehensive proteome detection and must be executed with care. Detergents, such as SDS, or bead beating are typical approaches for yeast lysis[33-35]. The SDS method, used in the Mann yeast studies, has produced excellent results, but requires removal of the detergent prior to MS sampling. The bead beating method mixes glass beads and yeast cells in a buffer slurry which is shaken for three, 1-4 minute cycles, at 30 hz. This approach, however, can be too gentle to sufficiently lyse the yeast and, in our hands, does not efficiently extract all proteins. We aimed to avoid use of detergents and investigated a more vigorous bead beating procedure. By simply extending the number of cycles to eight (4 minutes each) we achieved considerably improved results. Finally, we note increased identifications when lysates were not cleared of insoluble material. Zubarev and colleagues recently reported similar findings for mammalian cell culture samples[36].

Previous single-shot yeast proteome analyses utilized long capillary LC columns (50 cm) and lengthy gradients (240 minutes).[36-38] We aimed to achieve comparable or better coverage using a much shorter gradient (70 minutes). We found that capillary LC columns of 30 cm packed with 1.7 μm BEH particles (Waters Corporation) operating at flow rates of 350-375 nL/minute offered consistent elution across the one hour gradient. To accommodate this flow rate, a home-built column heater was maintained at a temperature of 60°C throughout the separation. Sample was loaded directly onto the column to avoid losses.

Recent work by Kuster *et al.* described that addition of ~5% DMSO to the mobile phase solvents boosted precursor S/N, providing up to a 20% increase in protein identifications[31,32]. We tested addition of DMSO to our chromatography solvents by comparing the number of yeast peptide and protein identifications obtained over our 70 minute gradient either with or without DMSO. In our hands the presence of DMSO increased the average precursor signal, from ~ 2.8×10^7 to 4.8×10^7 (arbitrary units) and increased the total ion current by 170%. This amplified signal afforded a 9% increase in unique

peptide identifications and 5% more proteins. We conclude that DMSO can indeed improve performance, with no apparent downside, and included it for all subsequent experiments.

Whole Yeast Proteome Analysis

To test our supposition that the faster scanning Q-OT-qIT hybrid could deliver comprehensive yeast proteome analysis within ~ one hour, we sequentially analyzed trypsin-digested, yeast cell lysate in quintuplicate. Each replicate began by loading ~1.4 µg of sample and followed by analysis over a 70 minute gradient. Allowing for sample loading, column washing, and equilibration, the five consecutive analyses consumed ~8 hours; however, actual instrument acquisition times were ~1.3 hours per experiment. As anticipated, the Q-OT-qIT hybrid posted a considerable number of scans: on average 13,447 MS^1 and 80,460 MS^2 events per run. As a point of reference, state-of-the-art analysis in 2003, by Gygi *et al.*, recorded 162,000 MS^2 events following 135 hours of MS operation. The Q-OT-qIT hybrid delivered this number of scans in two and a half hours! Next, we analyzed the yeast sample on the most recently introduced quadrupole linear ion trap Orbitrap system (*i.e.*, qIT-OT or Orbitrap Elite) using the identical chromatographic conditions[39]. That mass spectrometer produced only about a quarter of the MS^1 scans, as compared to the Q-OT-qIT (3,635), and half the MS^2 events (39,447).

Figure 2 presents a series of MS^2 scans acquired by the Q-OT-qIT MS over a one second period. In this example, 22 precursors were selected for MS^2 analysis from MS^1 scan #59,211. All 22 product ion spectra were acquired, individually, within one second and are presented in the lower portion of **Figure 2**. 19 of these 22 scans were mapped to sequence (1% FDR). Upon database searching, each one hour experiment (Q-OT-qIT system) produced 43,400 ($\bar{x}$) peptide spectral matches (PSMs) and 34,255 ($\bar{x}$) peptides with unique amino acid sequences (1% FDR, **Table 1**). Batched analysis of the five experiments yielded 47,624 unique peptides. In each analysis, over half of the 80,460 MS^2 scans were mapped to sequence (54%). Despite the swift Q-OT-qIT scan rate (~20 Hz), we conclude the system routinely delivers spectra of high quality.

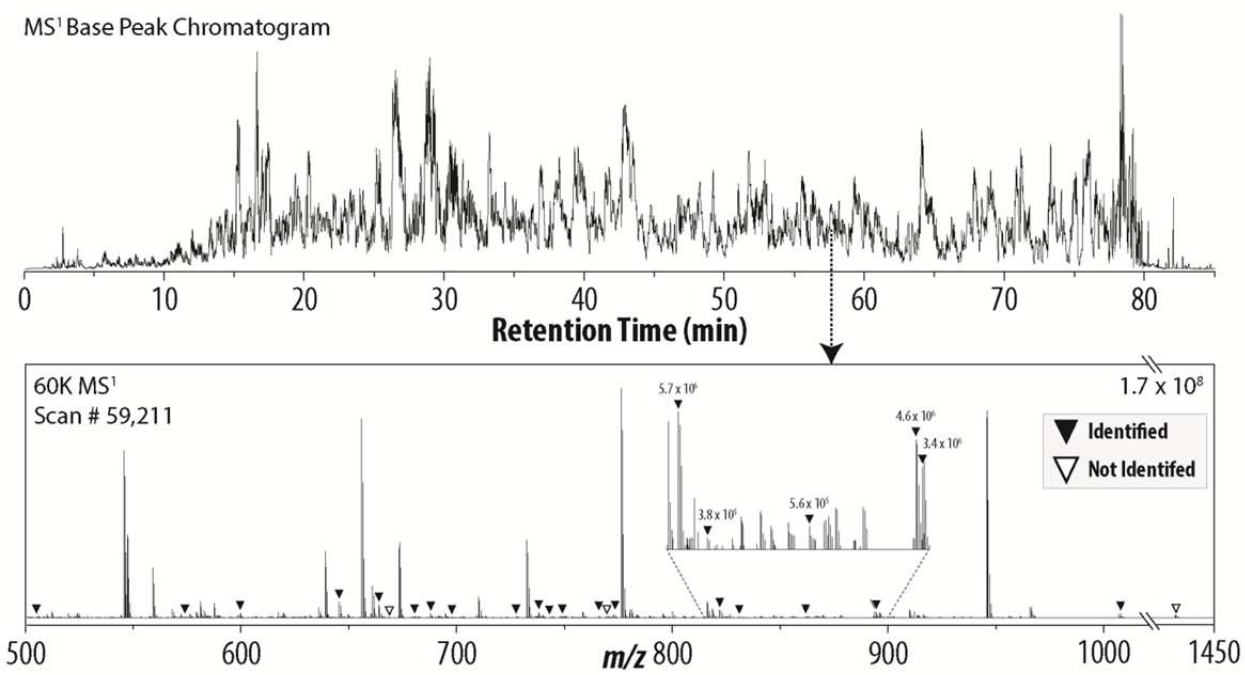
On average, each of the quintuplicate analyses achieved detection of 3,977 proteins (1% FDR), 13.5% (538, $\bar{x}$) of which originate from single peptides (**Figure 3**, panel A). Combination of the data reduces single peptide proteins to 460 while deepening coverage to 4,395 protein groups. 3,643 of these proteins (83%) were present in all five one hour experiments and 3,853 (88%) were found in four of five (**Figure 3**, panel B). Median sequence coverage was 18.4% and 23.7% for the individual and combined experiments, respectively (**Figure 3**, panel C), with a median of seven identified peptide sequences per protein. Yeast contains approximately 800 dubious open reading frames (ORFs). These ORFs, which are believed not to encode a corresponding protein, are commonly used to verify the FDR of proteomic datasets. For the one hour experiments, between 3 and 8 dubious ORFs were identified per dataset (**Table 1**), confirming these data are indeed well below the 1% FDR threshold.

To directly contrast the performance of the Q-OT-qIT hybrid to the most recent comprehensive yeast analysis we analyzed the same samples using a 240 minute gradient. This longer method mimics the 2012 study of Mann and colleagues (*vide infra*). With the extended gradient conditions the Q-OT-qIT system identified 46,381 unique peptide sequences corresponding to 4,392 protein groups (1% FDR), providing a median sequence coverage of 24.1% (average sequence coverage = 28.0%).

To estimate the dynamic range of the one hour experiments we compared our mass spectrometry-based identifications to those mapped by either tandem affinity (TAP) or green fluorescent protein (GFP) tagging experiments. From a single one hour experiment we identify 89% of proteins for which there is abundance data, including 73% of proteins present at less than 125 copies/cell (**Figure 3**, panel D)[9]. We also note detection of 886 proteins lacking abundance data. Next, we benchmarked the dynamic range of our global analysis to a recent multiple reaction monitoring (MRM) study. There, Aebersold *et al*. targeted 152 yeast proteins, spanning the full concentration range including several proteins never observed in public proteomic datasets, using synthetic peptides and triple quadrupole MRM technology[40]. A single one hour Q-OT-qIT experiment, batched analysis of all 5 one hour experiments, and our four hour analysis netted 122, 133, and 132 of the 152 Aebersold targets, respectively. The Aebersold work

detected 137 of these protein targets likely following dozens of MRM experiments. We conclude that our one hour method, using the Q-OT-qIT hybrid, provides dynamic range and sensitivity comparable to state-of-the-art MRM studies, but with whole proteome depth. We also note that the system, with its quadrupole mass filter, offers considerable promise for parallel reaction monitoring (PRM)[41,42].

Figure 2. Overview of Q-OT-qIT scan cycle. At a retention time of 57.88 minutes scan #59,211, an MS^1, was acquired and presented several spectral features for MS^2 analysis. Triangles indicate the 22 precursors that were selected for subsequent MS^2 sampling – all of which were acquired within one second of scan #59,211. 19 of these 22 MS^2 spectra were subsequently mapped to sequence.

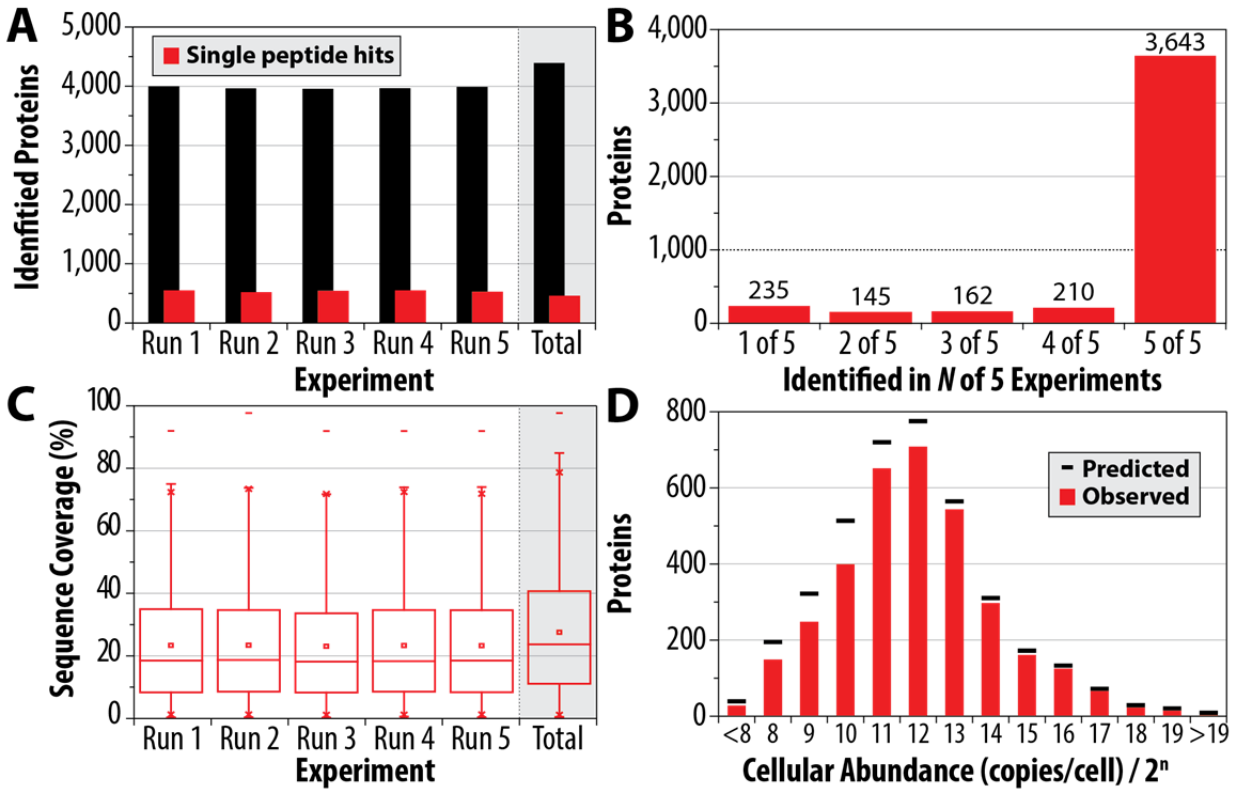


Scan #	RT	Identified Sequence	MS/MS Spectra
59,212	57.883	ALINNVSGYVIK	
59,213	57.885	YVVSFDELEAR	
59,214	57.885	IMINCFNECIDYAK	
59,215	57.886	SPIILQTSNGGAAYFAGK	
59,216	57.887	IVNEPTAAAIAYGLDK	
59,217	57.888	FELTGIPPAPR	
59,218	57.889	*n/a*	
59,219	57.889	GAFGTVVQAYLTQDKK	
59,220	57.890	ELQGELNFIEER	
59,221	57.891	LPLVGGHEGAGVVVGMEGENVK	
59,222	57.891	GIVIGSGAMSPLECR	
59,223	57.892	SPTGIVLMNMGGPSK	
59,224	57.893	*n/a*	
59,225	57.894	FQLEVTTPTAAAVGK	
59,226	57.894	AEVSSMPTLIFYK	
59,227	57.895	DAQVMIDLVNDK	
59,228	57.896	LNADFQKPWFATVNGQAR	
59,229	57.896	GAFGTVVQAYLTQDKK	
59,230	57.897	FDEKPDYLFLAR	
59,231	57.898	FLNNDLSAQFVK	
59,232	57.898	*n/a*	
59,233	57.899	SPQGLPYPIQIDK	

Table 1. Summary of identification results for the quintuplicate one hour yeast proteome experiments using the Q-OT-qIT mass spectrometer. Note SGD stems from the Saccharomyces Genome Database (www.yeastgenome.org).

Experiment	PSMs	Peptides	Proteins	SGD verified	SGD un-characterized	SGD dubious
1	43,423	34,535	4,002	3,630	337	8
2	43,622	34,495	3,966	3,608	331	7
3	42,339	33,450	3,959	3,595	334	3
4	43,326	34,347	3,968	3,602	337	8
5	43,343	34,449	3,991	3,623	341	4
Total	**216,256**	**47,624**	**4,395**	**3,976**	**381**	**16**

Figure 3. Performance metrics for one hour analysis, performed in quintuplicate, of a yeast trypic digest using the Q-OT-qIT hybrid. On average, 3,977 yeast proteins were identified in each experiment (1% FDR) with only 13.5% (538, $\bar{x}$) originating from single peptide identifications (panel A). 4,395 proteins were detected across all experiments – 3,643 of which were present in all five one hour experiments (panel B). Panel C presents the median sequence coverage for the individual and combined experiments. Panel D displays the overlap in identified proteins vs. known expression level information derived from published tagging experiments.



Discussion

The data presented above provides a deep view of the yeast proteome. Equally important is that this depth is achieved within an unprecedented time-scale. To understand how these gains were realized we plotted the number of identified peptide sequences per second as a function of elution time (**Figure 4**, panel A) for both the Q-OT-qIT (red) and the previous generation qIT-OT (black) hybrids. Remarkably, this plot reveals that the Q-OT-qIT hybrid (Orbitrap Fusion) routinely identifies ~ 8 peptides per second with occasions where up to 19 peptide sequences are detected in a single one second window. These stunning metrics are approximately double that achieved by the qIT-OT (Orbitrap Elite). Further, the speed of the Q-OT-qIT allows for deeper MS^2 sampling of the MS^1. The mean precursor depth sampled by the Q-OT-qIT is the 349^{th} most abundant *m/z* peak in the MS^1 scan while the qIT-OT system achieves an average depth of only 202 (**Figure 4**, panel B). The Q-OT-qIT frequently samples *m/z* peaks that are the 800^{th} or weaker in intensity rank.

To further examine the effect of Q-OT-qIT scan speed we plotted the cumulative number of unique peptide sequence identifications as a function of retention time (**Figure 4**, panel C, I/L ambiguity removed). Almost linearly across the 70 minute gradient and wash period the Q-OT-qIT (solid red) accumulates unique peptide identifications at a rate considerably faster than the qIT-OT (solid black) – 8.3 vs. 4.3 unique sequences per second, respectively (linear fit between 10 and 80 min). Use of a four hour gradient on this same system slightly reduces the slope (3.7 unique sequences per second, linear fit between 10 and 200 minutes, dotted red), but allows for considerably deeper analysis and again outpaces the qIT-OT system (2.3 unique sequences per second, dotted black). For reference we plotted these same data for the 2012 Mann *et al.* study using the Q-OT hybrid (Q-Exactive, dotted blue). That system posts the shallowest slope (1.7 unique sequences per second) and approximately half the number of unique peptide sequences as compared to the Q-OT-qIT (Orbitrap Fusion) for the same analysis time. Panel D of **Figure 4** compares unique protein identification rates for the same datasets. Again the Q-OT-qIT system

is the top performer, even when comparing the one hour analysis (solid red) to four hour experiments on either of the other systems (dotted black and blue).

Here we described new mass spectrometer technology that is capable of achieving comprehensive yeast proteome coverage within an unprecedented time-scale. Doubtless over the past decade many improvements to sample preparation, chromatography, and MS hardware have contributed to making this achievement possible. Among all these, we attribute increased mass spectrometer scan speed as the primary reason for the acceleration in proteome analysis speed and depth. **Figure 5** illustrates the pace of protein identifications for several large-scale yeast proteomic analyses as a function of the mass spectrometer MS^2 scan rate. Note the rapid ascent in protein identification rates scales correlates with increasing MS^2 scan rate.

The correlation depicted in **Figure 5** was somewhat surprising to us as we expected that ionization suppression of lower abundance peptides would become increasingly dominant as complex peptide mixtures comprising whole proteomes are separated over shorter gradients – *i.e.*, from four to one hour[43,44]. In other words, as the separation duration of the online chromatography is compressed, increased co-elution must occur. With increased co-elution one might expect that, regardless of the MS speed or sensitivity, ionization suppression would prevent a considerable fraction of peptides from becoming gas-phase ions – a requisite for MS detection. The results shown here refute this hypothesis and confirm that further improvements in MS sensitivity and speed will continue to reduce whole proteome analysis time, most likely to less than one hour for relatively simple proteomes like yeast.

Finally, we conclude that comprehensive analysis of mammalian proteomes within several hours is now within our technical reach. Consider that recent estimates suggest between 10,000 and 12,000 proteins are expressed at any given time for human cells in culture[45-47]. That is only approximately three to four times the complexity of the yeast proteome. Thus, our current efforts are aimed at achieving comprehensive coverage of mammalian system within just a few hours of analysis. Looking forward, one more doubling of MS^2 acquisition rate, *i.e.*, from 20 to 40 Hz, has potential to deliver detection of the

whole human proteome in just one to two hours. And, given the history and rate of MS innovation, such capability is likely only a few years away.

Figure 4. Analytical metrics of yeast proteome analysis using the Q-OT-qIT (Fusion) as compared to qIT-OT (Orbitrap Elite) and Q-OT (Q-Exactive) hybrids. The Q-OT-qIT (panel A, red) achieves identification of up to 19 peptides per second as compared to 10 with the qIT-OT system (panel A, black). Peak depth is likewise considerably higher on account of the faster MS^2 scanning rate of the Q-OT-qIT system (panel B). Panel C plots the pace of unique yeast peptide identifications for the three instruments. For the one hour analysis, the Q-OT-qIT posts almost twice as many unique peptide identifications as compared to the qIT-OT. Similar data, except for unique proteins, is shown in Panel D.

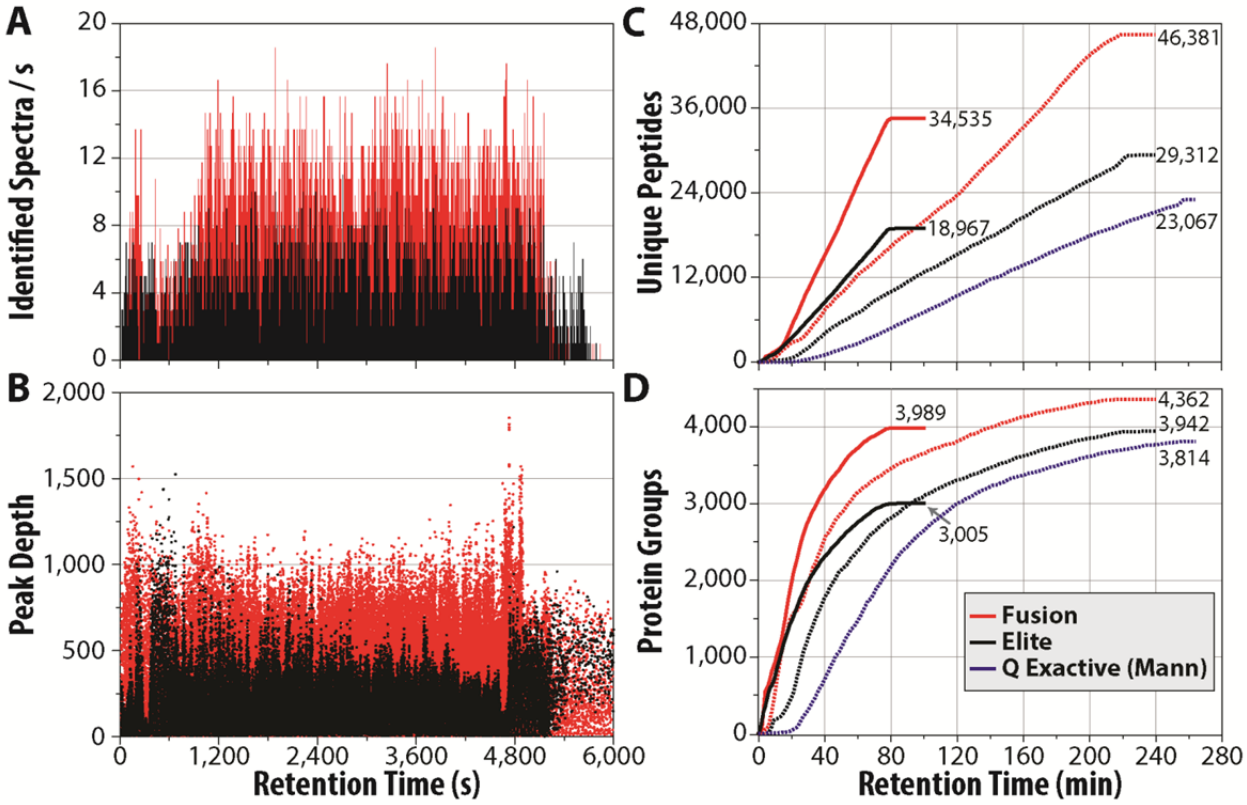
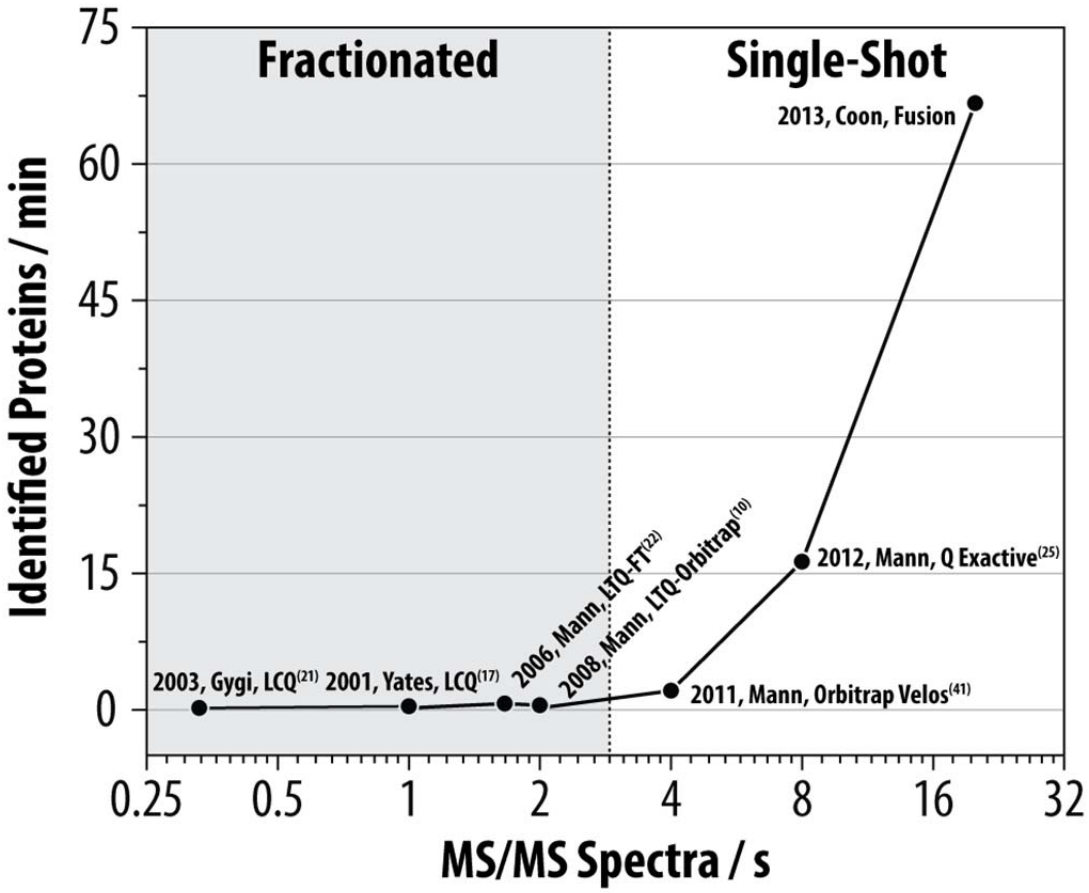


Figure 5. Rate of protein identifications as a function of mass spectrometer scan rate for large-scale yeast proteome analyses over the past decade.



Experimental procedures

Yeast Culture and Lysis

Saccharomyces cerevisiae strain BY4741 was grown in yeast extract peptone dextrose media (YPD) (1% yeast extract, 2% peptone, 2% dextrose). Four liters of media was divided between four-two liter flasks and inoculated with a starter culture (OD_{600} = 2.58). Cells were allowed to propagate for ~12 generations (20 hours) to an OD_{600} ~2 (average of 2.18). The cells were harvested by centrifugation at 5,000 rpm for 5 minutes, supernatant decanted, re-suspended in chilled NanoPure water and all pellets were pooled together. The cells were washed two more times and centrifuged for the final pelleting at 5,000 rpm for 10 min. A pellet corresponding to 5% of the total cells grown, was resuspended in lysis buffer composed of 50mM Tris pH8, 8M urea, 75mM sodium chloride, 100mM sodium butyrate, protease (Roche) and phosphatase inhibitor tablet (Roche). Yeast cells were lysed by glass bead milling (Retsch). Briefly, 2 mL of acid washed glass beads were combined with 2.5 mL of resuspended yeast cells in a stainless steel container and shaken 8 times at 30 hz for 4 minutes with a 1 min rest in between.

Protein Digestion

Lysate protein concentration was measured by BCA (Thermo Pierce). Protein was reduced by addition of 5 mM dithiothreitol and incubated for 45 minutes at 55 °C. The mixture was cooled to room temperature, followed by alkylation of free thiols by addition of 15 mM iodoacetamide in the dark for 30 minutes. The alkylation reaction was quenched with 5 mM dithiothreitol. Urea concentration was diluted to 1.5 M with 50 mM Tris pH 8.0. Proteolytic digestion was performed by addition of Trypsin (Promega), 1:50 enzyme to protein ratio, and incubated at ambient temperature overnight. An additional 1:50 bolus of trypsin was added in the morning and incubated at ambient temperature for 1 hour. The digestion was quenched by addition of TFA and desalted over a tC18 Sep-Pak (Waters).

nLC-MS^2 analysis

Reversed phase columns were prepared in-house. Briefly, a 75 µm-360 µm inner-outer diameter bare-fused silica capillary, with a laser pulled electrospray tip, was packed with 1.7 µm diameter, 130 Å pore size, Bridged Ethylene Hybrid C18 particles (Waters) to a final length of 35 cm. The column was installed on a nanoAcquity UPLC (Waters) using a stainless steel ultra-high pressure union formatted for 360 µm outer diameter columns (IDEX) and heated to 60 °C for all runs. Mobile phase buffer A was composed of water, 0.2% formic acid, and 5% DMSO. Mobile phase B was composed of acetonitrile, 0.2% formic acid, and 5% DMSO. Samples were loaded onto the column for 12 min at 0.35 µL/min. Mobile phase B increases to 4% in the first 0.1 min then to 12% B at 32 min, 22% B at 60 min, and 30% B at 70 min, followed by a 5 min wash at 70% B and a 20 min re-equilibration at 0%B.

Eluting peptide cations were converted to gas-phase ions by electrospray ionization and analyzed on a Thermo Orbitrap Fusion (Q-OT-qIT, Thermo). Survey scans of peptide precursors from 300 to 1,500 *m/z* were performed at 60K resolution (at 200 *m/z*) with a 5 x 10^5 ion count target. Tandem MS was performed by isolation at 0.7 Th with the quadrupole, HCD fragmentation with normalized collision energy of 30, and rapid scan MS analysis in the ion trap. The MS^2 ion count target was set to 10^4 and the max injection time was 35 ms. Only those precursors with charge state 2-6 were sampled for MS^2. The dynamic exclusion duration was set to 45 seconds with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 5 second cycles, meaning the instrument would continuously perform MS^2 events until the list of non-excluded precursors diminishes to zero or 5 seconds, whichever is shorter. Elite runs were performed with Survey scans of peptide precursors from 300 to 1,500 *m/z* 60K resolution (at 200 *m/z*) with a 1 x 10^6 ion count target. Tandem MS was performed by isolation at 1.8 Th with the ion-trap, CAD fragmentation with normalized collision energy of 35, and rapid scan MS analysis in the ion trap. The data dependent top 20 precursors were selected for MS^2. MS^2 ion count target was set to 5 x 10^3 and the

max injection time was 125 ms. Only those precursors with charge state +2 or higher were sampled for MS^2. The dynamic exclusion duration was set to 40 seconds with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on.

Data Analysis

The raw data was processed using Proteome Discoverer (version 1.4.0.288, Thermo Fischer Scientific). MS^2 spectra were searched with SEQUEST engine against a database of 6,632 yeast ORFs (www.yeastgenome.com, February 3, 2011)[48]. Peptides were generated from a tryptic digestion with up to two missed cleavages, carbamidomethylation of cysteines as fixed modifications, and oxidation of methionines and protein N-terminal acetylation as variable modifications. Precursor mass tolerance was 20 ppm and product ions were searched at 0.35 Da tolerances. Peptide spectral matches (PSM) were validated using percolator based on q-values at a 1% false discovery rate (FDR)[49]. With proteome Discoverer, peptide identifications were grouped into proteins according to the law of parsimony and filtered to 1% FDR[50]. For cumulative protein group identification, PSMs passing the FDR were exported to a text file and processed by a modified version of Protein Hoarder (version 2.4.1)[51]. The PSMs were iteratively processed in successive one minute windows and grouped into proteins using the law of parsimony at a 1% FDR.

Acknowledgements

We thank Graeme McAlister, Steve Gygi, Jae Schwartz, John Syka, Jens Griep-Raming, Vlad Zabrouskov, Mike Senko, and Jesse Canterbury for helpful discussions. We are grateful to Anna Merrill for yeast production and for critical reading of the manuscript. We thank Audrey Gasch for assistance with yeast growth. This work was supported by the National Institutes of Health (R01 GM080148) and the National Science Foundation (0701846). A.L.R. gratefully acknowledges support from a National Institutes of Health-funded Genomic Sciences Training Program (5T32HG002760).

References

1 Walther, T. C. & Mann, M. (2010) Mass spectrometry–based proteomics in cell biology. *The Journal of Cell Biology* **190**, 491-500, doi:10.1083/jcb.201004052

2 Mallick, P. & Kuster, B. (2010) Proteomics: a pragmatic perspective. *Nat. Biotechnol.* **28**, 695-709, doi:10.1038/nbt.1658

3 Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) QUANTITATIVE MONITORING OF GENE-EXPRESSION PATTERNS WITH A COMPLEMENTARY-DNA MICROARRAY. *Science* **270**, 467-470, doi:10.1126/science.270.5235.467

4 DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686, doi:10.1126/science.278.5338.680

5 Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. (1999) Correlation between Protein and mRNA Abundance in Yeast. *Molecular and Cellular Biology* **19**, 1720-1730

6 Grimsrud, P. A., Swaney, D. L., Wenger, C. D., Beauchene, N. A. & Coon, J. J. (2010) Phosphoproteomics for the Masses. *ACS Chemical Biology* **5**, 105-119, doi:10.1021/cb900277e

7 Hebert, Alexander S., Dittenhafer-Reed, Kristin E., Yu, W., Bailey, Derek J., Selen, Ebru S., Boersma, Melissa D., Carson, Joshua J., Tonelli, M., Balloon, Allison J., Higbee, Alan J., Westphall, Michael S., Pagliarini, David J., Prolla, Tomas A., Assadi-Porter, F., Roy, S., Denu, John M. & Coon, Joshua J. (2013) Calorie Restriction and SIRT3 Trigger Global Reprogramming of the Mitochondrial Protein Acetylome. *Molecular Cell* **49**, 186-199, doi:http://dx.doi.org/10.1016/j.molcel.2012.10.024

8 Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., Simison, M., Weng, S. & Wong, E. D. (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700-D705, doi:10.1093/nar/gkr1029

9 Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. & Weissman, J. S. (2003) Global analysis of protein expression in yeast. *Nature* **425**, 737-741, doi:http://www.nature.com/nature/journal/v425/n6959/suppinfo/nature02046_S1.html

10 de Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F., Walther, T. C. & Mann, M. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251-1254, doi:http://www.nature.com/nature/journal/v455/n7217/suppinfo/nature07341_S1.html

11 Wu, R., Dephoure, N., Haas, W., Huttlin, E. L., Zhai, B., Sowa, M. E. & Gygi, S. P. (2011) Correct Interpretation of Comprehensive Phosphorylation Dynamics Requires Normalization by Protein Expression Changes. *Molecular & Cellular Proteomics* **10**, doi:10.1074/mcp.M111.009654

12 Webb, K. J., Xu, T., Park, S. K. & Yates, J. R. (2013) Modified MuDPIT Separation Identified 4488 Proteins in a System-wide Analysis of Quiescence in Yeast. *Journal of Proteome Research* **12**, 2177-2184, doi:10.1021/pr400027m

13 *Merriam-Webster's dictionary*

14 Figeys, D., Ducret, A., Yates, J. R. & Aebersold, R. (1996) Protein identification by solid phase microextraction[mdash]capillary zone electrophoresis[mdash]microelectrospray[mdash]tandem mass spectrometry. *Nat Biotech* **14**, 1579-1583

15 Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. & Mann, M. (1996) Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proceedings of the National Academy of Sciences* **93**, 14440-14445

16 Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. & Yates, J. R. (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotech* **17**, 676-682

17 Washburn, M. P., Wolters, D. & Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotech* **19**, 242-247, doi:http://www.nature.com/nbt/journal/v19/n3/suppinfo/nbt0301_242_S1.html

18 Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J. X., Cohn, M. A., Cantley, L. C. & Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12130-12135, doi:10.1073/pnas.0404720101

19 Zanivan, S., Gnad, F., Wickstrom, S. A., Geiger, T., Macek, B., Cox, J., Fassler, R. & Mann, M. (2008) Solid Tumor Proteome and Phosphoproteome Analysis by High Resolution Mass Spectrometry. *Journal of Proteome Research* **7**, 5314-5326, doi:10.1021/pr800599n

20 Ledvina, A. R., Beauchene, N. A., McAlister, G. C., Syka, J. E. P., Schwartz, J. C., Griep-Raming, J., Westphall, M. S. & Coon, J. J. (2010) Activated-Ion Electron Transfer Dissociation Improves the Ability of Electron Transfer Dissociation to Identify Peptides in a Complex Mixture. *Analytical Chemistry* **82**, 10068-10074, doi:10.1021/ac1020358

21 Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. & Gygi, S. P. (2002) Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC−MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. *Journal of Proteome Research* **2**, 43-50, doi:10.1021/pr025556v

22 de Godoy, L. M. F., Olsen, J. V., de Souza, G. A., Li, G. Q., Mortensen, P. & Mann, M. (2006) Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **7**, doi:R50

23 Syka, J. E. P., Marto, J. A., Bai, D. L., Horning, S., Senko, M. W., Schwartz, J. C., Ueberheide, B., Garcia, B., Busby, S., Muratore, T., Shabanowitz, J. & Hunt, D. F. (2004) Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *Journal of Proteome Research* **3**, 621-626, doi:10.1021/pr0499794

24 Swaney, D. L., Wenger, C. D. & Coon, J. J. (2010) Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *Journal of Proteome Research* **9**, 1323-1329, doi:10.1021/pr900863u

25 Nagaraj, N., Alexander Kulak, N., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O. & Mann, M. (2012) System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Molecular & Cellular Proteomics* **11**, doi:10.1074/mcp.M111.013722

26 Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M. & Muddiman, D. C. (2011) Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600). *Analytical Chemistry* **83**, 5442-5446, doi:10.1021/ac200812d

27 Cristobal, A., Hennrich, M. L., Giansanti, P., Goerdayal, S. S., Heck, A. J. R. & Mohammed, S. (2012) In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. *Analyst* **137**, 3541-3548, doi:10.1039/c2an35445d

28 Geromanos, S., Hughes, C., Ciavarini, S., Vissers, J. C. & Langridge, J. (2012) Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples. *Anal Bioanal Chem* **404**, 1127-1139, doi:10.1007/s00216-012-6197-y

29 Senko, M. W., Remes, P., Song, Q., Canterbury, J., Blethrow, J., Zabrouskov, V., Lange, O. & Makarov, A. (2013) Improving data dependent MSn Performance with a Multitasking Mass Spectrometer. *Proceedings of the 61ST ASMS Conference on Mass spectrometry*

30 Senko, M., Remes, P., Canterbury, J., Mathur, R., Song, Q., Eliuk, S., Mullen, C., Earley, L., Hardman, H., Blethrow, J., Bui, H., Specht, A., Lange, O., Denisov, E., Makarov, A., Horning, S. & Zabrouskov, V. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improves Proteome Coverage and Peptide Identification Rates. *Analytical Chemistry*, doi:pending

31 Meyer, J. & A. Komives, E. (2012) Charge State Coalescence During Electrospray Ionization Improves Peptide Identification by Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **23**, 1390-1399, doi:10.1007/s13361-012-0404-0

32 Hahne, H., Pachl, F., Ruprecht, B., Maier, S. K., Klaeger, S., Helm, D., Medard, G., Wilm, M., Lemeer, S. & Kuster, B. (2013) DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Meth* **10**, 989-991, doi:10.1038/nmeth.2610

33 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat Meth* **6**, 359-362, doi:http://www.nature.com/nmeth/journal/v6/n5/suppinfo/nmeth.1322_S1.html

34 Dephoure, N. & Gygi, S. P. (2012) Hyperplexing: A Method for Higher-Order Multiplexed Quantitative Proteomics Provides a Map of the Dynamic Response to Rapamycin in Yeast. *Sci. Signal.* **5**, rs2-, doi:10.1126/scisignal.2002548

35 Hebert, A. S., Merrill, A. E., Bailey, D. J., Still, A. J., Westphall, M. S., Strieter, E. R., Pagliarini, D. J. & Coon, J. J. (2013) Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Meth* **10**, 332-334, doi:http://www.nature.com/nmeth/journal/v10/n4/abs/nmeth.2378.html#supplementary-information

36 Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J. & Zubarev, R. A. (2013) Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Molecular & Cellular Proteomics*, doi:10.1074/mcp.O113.028787

37 Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Frohlich, F., Cox, J. & Mann, M. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol Cell Proteomics* **10**, M110 003699, doi:10.1074/mcp.M110.003699

38 Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O. & Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol Cell Proteomics* **11**, M111 013722, doi:10.1074/mcp.M111.013722

39 Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Müller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J.-J., Cox, J., Horning, S., Mann, M. & Makarov, A. (2012) Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Molecular & Cellular Proteomics* **11**, doi:10.1074/mcp.O111.013698

40 Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. (2009) Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. *Cell* **138**, 795-806, doi:http://dx.doi.org/10.1016/j.cell.2009.05.051

41 Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. (2012) Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & Cellular Proteomics*, doi:10.1074/mcp.O112.020131

42 Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T. & Domon, B. (2012) Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics*, doi:10.1074/mcp.O112.019802

43 King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C. & Olah, T. (2000) Mechanistic investigation of ionization suppression in electrospray ionization. *J. Am. Soc. Mass Spectrom.* **11**, 942-950, doi:10.1016/s1044-0305(00)00163-x

44 Annesley, T. M. (2003) Ion suppression in mass spectrometry. *Clin. Chem.* **49**, 1041-1044, doi:10.1373/49.7.1041

45 Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J. & Aebersold, R. (2011) The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, doi:http://www.nature.com/msb/journal/v7/n1/suppinfo/msb201182_S1.html

46 Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S. & Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**, doi:http://www.nature.com/msb/journal/v7/n1/suppinfo/msb201181_S1.html

47 Mann, M., Kulak, Nils A., Nagaraj, N. & Cox, J. (2013) The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. *Molecular Cell* **49**, 583-590, doi:http://dx.doi.org/10.1016/j.molcel.2013.01.029

48 Eng, J. K., McCormack, A. L. & Yates Iii, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976-989, doi:http://dx.doi.org/10.1016/1044-0305(94)80016-2

49 Brosch, M., Yu, L., Hubbard, T. & Choudhary, J. (2009) Accurate and Sensitive Peptide Identification with Mascot Percolator. *Journal of Proteome Research* **8**, 3176-3181, doi:10.1021/pr800982s

50 Nesvizhskii, A. I. & Aebersold, R. (2005) Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Molecular & Cellular Proteomics* **4**, 1419-1440, doi:10.1074/mcp.R500012-MCP200

51 Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. (2011) COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *PROTEOMICS* **11**, 1064-1074, doi:10.1002/pmic.201000616