Multi-omic Single-Shot Technology for Integrated Proteome and Lipidome Analysis

Yuchen He
University of Wisconsin-Madison https://orcid.org/0000-0002-6096-4915
Edrees Rashan
University of Wisconsin–Madison
Vanessa Linke
University of Wisconsin https://orcid.org/0000-0002-1827-577X
Evgenia Shishkova
University of Wisconsin–Madison
Alexander Hebert
University of Wisconsin–Madison
Michael Westphall
University of Wisconsin–Madison
Dave Pagliarini
Morgridge Institute for Research https://orcid.org/0000-0002-0001-0087
Katherine Overmyer
Morgridge Institute for Research https://orcid.org/0000-0002-1929-1229
Joshua Coon (jcoon@chem.wisc.edu)
University of Wisconsin-Madison https://orcid.org/0000-0002-0004-8253

Article

Keywords: mass spectrometry, multi-omic single-shot technology, methodology

Posted Date: July 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-45548/v1

Version of Record: A version of this preprint was published at Analytical Chemistry on February 22nd, 2021. See the published version at https://doi.org/10.1021/acs.analchem.0c04764.

Multi-omic Single-Shot Technology for Integrated Proteome and Lipidome Analysis

Yuchen He[1], Edrees H. Rashan[2], Vanessa Linke[1], Evgenia Shishkova[1], Alexander S. Hebert[1], Michael S. Westphall[1], David J. Pagliarini[3], Katherine A. Overmyer[1,4], and Joshua J. Coon[1,3,5] *

[1] Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison WI, USA
[2] Department of Biochemistry, University of Wisconsin-Madison, Madison WI, USA
[3] Departments of Cell Biology and Physiology; Biochemistry and Molecular Biophysics; and Genetics, Washington University School of Medicine, St. Louis, MO, USA.
[4] Morgridge Institute for Research, Madison, WI, USA
[5] Department of Chemistry, University of Wisconsin-Madison, Madison WI, USA

* Correspondence: J.J.C. (jcoon@chem.wisc.edu)

Abstract

Mass spectrometry (MS) serves as the centerpiece technology for proteome, lipidome, and metabolome analysis. Despite the versatility of MS systems, fractured methodology drives nearly all MS laboratories to specialize in analysis of a single ome at the exclusion of the others. We describe a technology to achieve broad and deep coverage of multiple molecular classes simultaneously through multi-omic single-shot technology (MOST) requiring only one column, one LC-MS instrument and a simplified workflow.

Main

Myriad regulatory layers involving tens of thousands of biomolecules, including nucleic acids, proteins, lipids, and metabolites modulate the cellular processes that in turn govern complex organisms. Untangling these multi-faceted networks will require innovative technologies to globally monitor diverse classes of biomolecules. For example, gene expression profiles have failed to fill unresolved gaps in many biosynthesis pathways implicated in many human diseases. Mass spectrometry (MS) has propelled systems biology by offering access to the proteome, lipidome, and metabolome. For each of these classes, MS analysis is remarkably similar, consisting of chromatographic separation, mass measurement, and tandem MS. Despite this commonality, nearly all MS laboratories specialize in niche applications targeting a single ome.

This arrangement causes two major problems. First, the confinement to a single ome across published data impoverishes the MS community's understanding of their interplay; a search of publications from the last five years reveals that though multi-omic studies have tripled in that time, less than ~ 7.5% of MS-based omic studies incorporated multi-omic analysis in 2019. Second, the majority of existing multi-omic analyses are conducted across laboratories, potentially leading to noise introduced by sample heterogeneity, variable sample handling, and instrument variance that can then obscure otherwise strong biomolecular associations[1,2,3,4,5]. We conclude that an integrated workflow is needed to not only improve data quality but unlock more comprehensive coverage of the biological system.

Thus, in contrast to the fragmented single-ome status quo, we envision an integrated technology for the metabolome, lipidome, and proteome that allows these omes to be jointly extracted and prepared from a single sample; loaded onto and separated by a single LC column; and simultaneously mass analyzed as they elute on a single MS platform. We term this method multi-omic single-shot technology (MOST). In this study, we demonstrate proof of concept by performing an integrated analysis of proteins and lipids. To our knowledge, MOST is the first technology that integrates proteome and lipidome analysis in a single LC-MS run using a single reverse-phase (RP) column and a binary mobile phase system[6,7,8] – *i.e.*, without the need to maintain two dedicated LC-MS systems or work across multiple laboratories.

To test our hypothesis that peptides and lipids can be co-analyzed in a single LC-MS methodology, we first sequentially loaded a complex mixture of lipids and then peptides from a yeast cell lysate onto an RP LC column. Next, using a modified gradient lasting 90 minutes, we eluted first the peptides (0–60 mins) and then the lipids (60–90 mins) into a quadrupole Orbitrap hybrid MS (**Figure 1a, supplementary Figure 1a**). Given the hydrophobicity differences of peptide and lipids, these two molecule classes are easily separated temporally for

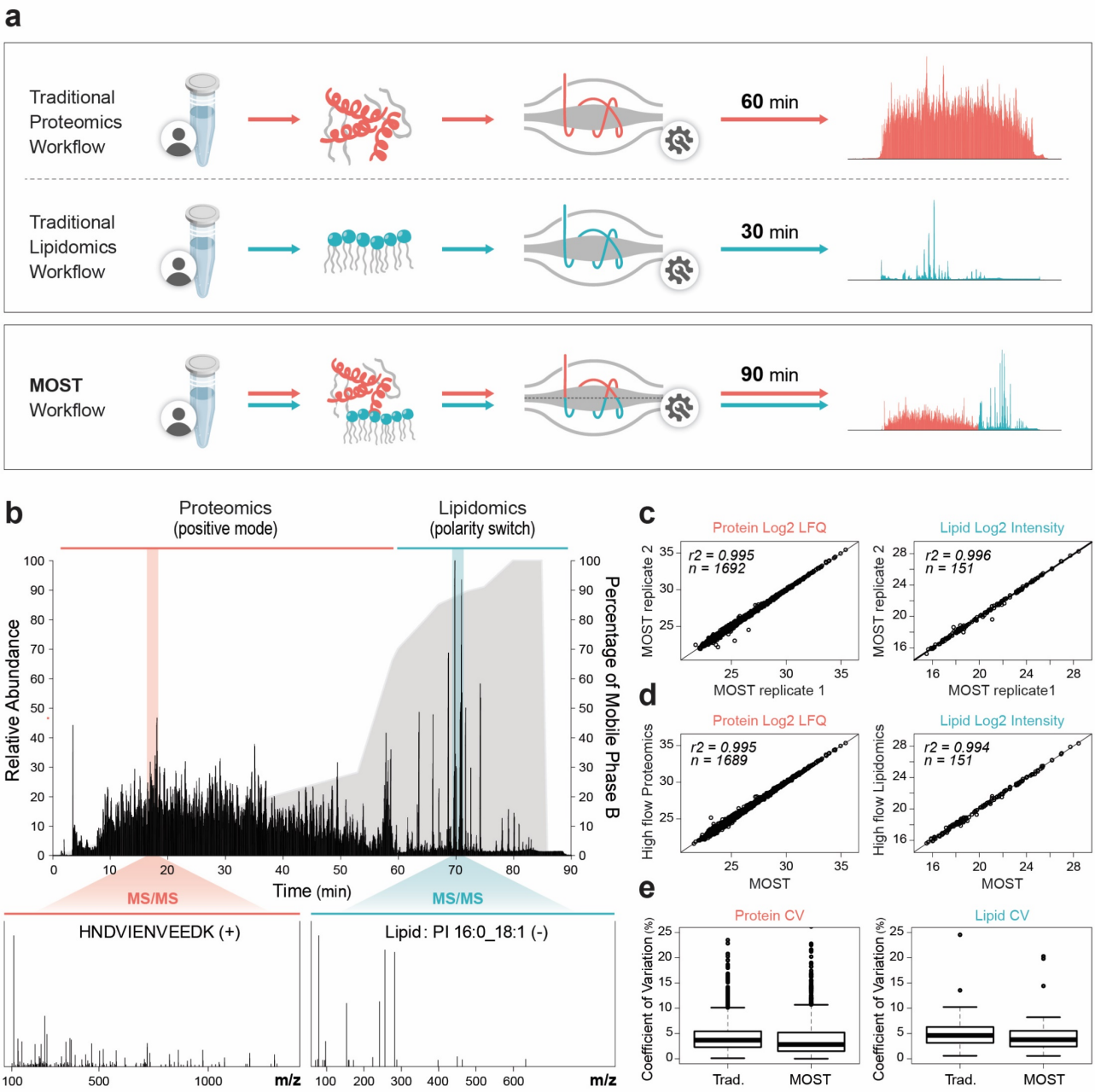


Figure 1. Performance characteristics of MOST. (a) Diagram showing comparison between traditional workflow and MOST workflow. (b) An example of total ion current (TIC) chromatogram of MOST, with the first 60 min for proteomic analysis and last 30 min for lipidomics analysis, along with two examples of MS/MS spectra of a peptide and a lipid. (c) Scatter plot of two MOST runs showing good reproducibility of quantification of proteins and lipids. (d) Scatter plot of a MOST run versus a high flow proteomics/lipidomics run showing no interference on protein/lipid quantification from each other. (e) Boxplot comparing coefficient of variation (CV) of proteins and lipids between traditional platforms and MOST (n=3).

durations we control using our MOST gradient (**Figure 1b**). During peptide elution, the mass spectrometer was operated in positive-ion mode and peptides were selected for tandem MS using a data-dependent top-ten method. Then, when mostly lipids begin to elute at approximately 60 minutes, the mass spectrometer is instructed to begin collecting in polarity-switching mode, now collecting with a data-dependent top-two method. Note best performance was achieved with a gradient comprising 90% isopropanol and 10% acetonitrile, a gradient commonly used for lipid separations[9, 10]. Further, the separations were performed using a higher flow system (60 µL per min); high flow systems are typical of lipidomics but have recently been developed for robust, high throughput proteomics[11, 12, 13].

Encouraged by these initial proof-of-concept results, we next sought to optimize the balance between peptide and lipid identifications and instrumentation demands. First, to achieve an optimal peptide injection amount under

high flow conditions, we assessed serial dilutions (**Supplementary Figure 1b**) and determined that an optimal load of 20 µg of peptides produced > 18,000 unique peptide identifications corresponding to > 2,200 proteins. In that same experiment, we typically detect ~ 150 lipids, a number comparable to the amount observed when we perform lipidomics separately on yeast (**Supplementary Figure 1c**). For both omes, we analyzed approximately equivalent amounts of starting material. Using MOST for mammalian cell line analysis, we identified and quantified > 2,600 proteins and > 500 lipids from a HAP1 cell line. These results were achieved without the addition of dimethyl sulfoxide (DMSO)[13,14]. Although DMSO enhances ionization and increases identification, in our case those benefits were short-term, as it rapidly caused instrument fouling. Additionally, in contrast to other high flow proteomics, we have added ammonium formate to the mobile phase. Ammonium salt is a common additive to lipidomics analyses to improve neutral lipid detection. Here, its addition led to a slight reduction in peptide identifications but a substantial increase in lipid identifications compared to no-ammonium analyses: 5.8% fewer unique peptides and 3.3% fewer protein groups as compared to an 11.6% gain in lipid identifications. It also led to the detection of key lipid classes, including cardiolipins (CLs) and triglycerides (TGs) (**Supplementary Figure 1d-e**). Finally, we tested two types of RP columns, C18 BEH and C18 CSH. The BEH column led to more identifications in both proteomic and lipidomic analyses: 1.3% more unique peptides and proteins, 14.9% more lipids (**Supplementary Figure 1f-g**).

MOST showed great reproducibility as measured by quantification and retention time (RT) stability (**Figure 1c; Supplementary Figure 2a**). Importantly, protein quantification (R^2 = 0.995) and peptide RT (R^2 = 1) were not affected by the presence of lipids on the column. Similarly, lipid quantification (R^2 = 0.994) and lipid RT (R^2 = 0.996) were identical (**Figure 1d; Supplementary Figure 2b**). MOST and our traditional platforms achieved similar reproducibility (3-4% median CV) (**Figure 1e**). In addition, comparable results were obtained for mass error (0.40 *vs.* 0.41 absolute median ppm for peptides and 2.73 *vs.* 2.72 absolute median ppm for lipids) and identifications (< 10 identified proteins or lipids differed between single-ome high flow platforms and our newly developed MOST platform) (**Supplementary Figure 2c-e**).

As the objective of MOST is to deliver high throughput multi-omic measurements, we next applied our system to study the yeast lipid metabolism – an enterprise that benefits from quantitative measurement of both proteins and lipids. We examined 20 *Saccharomyces cerevisiae* strains with mitochondria- and lipid-related gene deletions (Δ*gene*) in biological triplicate. Using MOST, we identified and quantified a total of 2,842 proteins and 325 lipids among 20 *S. cerevisiae* strains from a single LC-MS method. Before exploring biological findings, we sought to compare these results with traditional platforms, *i.e.* nano flow capillary LC-MS analysis (proteomics) and high flow LC-MS analysis (lipidomics) (**Supplementary Figure 3**). Obviously, the state-of-the-art capillary LC-MS methods outperform high flow approaches in terms of identifications (4,182 protein groups for these samples). However, we observed other comparable quantitative figures of merit, *e.g.*, capability of measuring fold change (up to three orders of magnitude) and coefficient of variation (CV) (6–7% median CV). Lipidomic measurements using the traditional platform provided nearly as many lipid identifications as MOST (287 *vs.* 325 lipids). Again, capability of measuring fold change (up to three orders of magnitude) and CV (15–16% median CV) were comparable. Highlighting the high level of agreement between these datasets, the distribution of gene strains in principal component analysis (PCA) followed the same pattern in both platforms. Finally, by every one of these metrics in this expanded comparison, there is no difference in performance of MOST when both lipids and peptides are present on the column as compared to when they are loaded individually.

Having validated the quality of multi-omic data generated by MOST, we next aimed to assess what it can reveal about the biological system of interest, particularly perturbations to lipid biosynthesis pathways. The Δ*gene* strains studied here can be divided into three categories based on their target pathway: coenzyme Q_6 (CoQ_6) biosynthesis, cardiolipin (CL) biosynthesis, and those of uncharacterized function. We discovered that on average 647 molecules (575 proteins and 72 lipids, 20.4% of total measured) were perturbed with each gene deletion as compared to wild-type. For example, the gene products of *tam41* and *taz1* function in the early and late stage of CL biosynthesis, respectively. Accordingly, we observed significant changes in both protein and lipid abundances in Δ*tam41* and Δ*taz1*, which results an ultimate reduction of CL levels (**Figure 2a**). Similarly, for CoQ_6 biosynthesis-related Δ*gene* strains, we observed expected dramatic changes in CoQ_6 biosynthesis

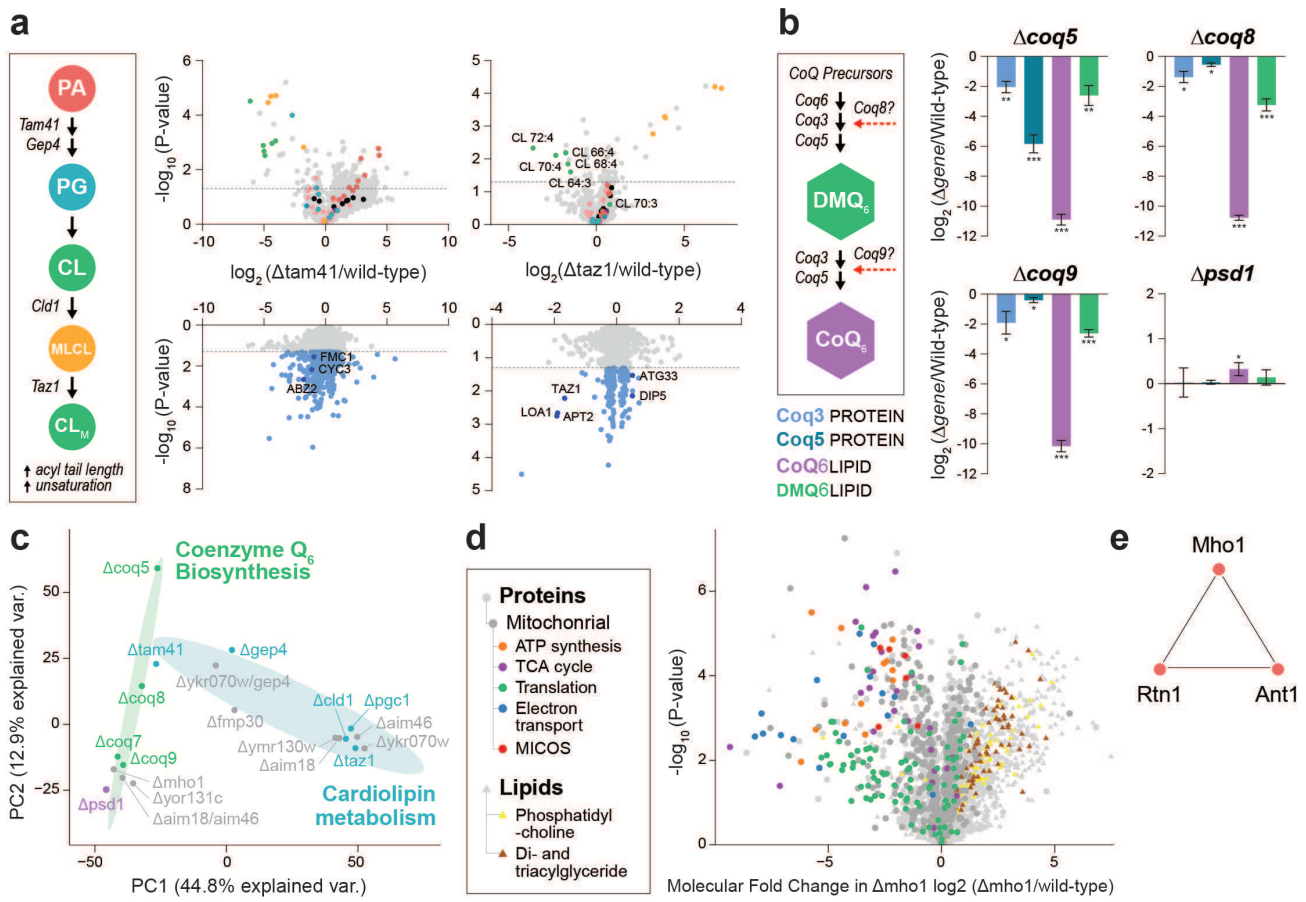


Figure 2. Biological study by MOST. (a) Schematic pathway of cardiolipin biosynthsis and volcano plots across related gene knockout strains. PA, phosphatidic acid. PG, phosphoglycerol. CL, cardiolipin. MLCL, monolysocardiolipin. (b) Schematic pathway of Co-enzyme Q6 biosynthsis and molecule abundance across related gene knockout strains. *, p-value < 0.05. **, p-value < 0.01. ***, p-value < 0.001. (c) Principal component analysis (PCA) of Δgene strains. (d) Volcano plot showing average fold-change in molecule abundances (mean log2[Δmho1/wild-type]) versus statistical significance, showing select functional groups (GO terms and lipid class) significantly (Benjamini-Hochberg adjusted p-value < 0.05) enriched in either upregulated or downregulated molecules. (e) Nearest neighbor covariance network for Mho1.

proteins, the lipid CoQ_6 itself, and its upstream intermediates (**Figure 2b**). To our knowledge, this is the first time such multi-omic data has been acquired simultaneously.

Serving the multi-omic nature of our MS profiles, we determined pairwise covariance between proteins and lipids to infer gene function (**Supplementary Figure 4**). Examination of correlations across all proteins and lipids revealed numerous molecular relationships of this multi-omic covariance network, visualized as nodes for molecules and edges for correlations. After applying strict correlation thresholds (Pearson |r| ≥ 0. 8, Benjamini-Hochberg adjusted *p*-value < 0.001), a striking 477,090 edges between 2,618 nodes remained. For example, loss of *mho1* leads to a severe respiratory deficiency, decreased abundance of prominent mitochondrial proteins, and increased abundance of di- and triacylglycerides (**Figure 2c-d**). we observed that the poorly characterized yeast protein Mho1 correlates with Rtn1 and Ant1 (**Figure 2e**), which are important mediators of peroxisomal biogenesis and metabolic activity[15,16,17]. The mammalian homolog of Mho1, MEMO1, is a copper-binding enzyme that produces O_2^- and regulates microtubule stability during cell migration[18,19]. One hypothesis drawn from our covariance analysis is that Mho1 exerts a similar enzymatic activity in yeast and produces ROS that signals to and controls peroxisome homeostasis. While further experimentation is required to verify whether a functional connection exists between Mho1 and peroxisomal organization, we present MOST as exciting new MS approach that can deliver multi-omic data expeditiously and thereby accelerate biochemical discovery.

We have described a method for integrated, fast multi-omics that enables the co-analysis of lipids and peptides. Leveraging this method to analyze a yeast gene deletion collection, we have demonstrated its ability to detect coordinated multi-omic perturbations. Next we wondered whether co-processing and co-analyzing proteins and lipids might afford increased ability to detect co-regulation across omes. To do this, we compared correlations generated by the different LC-MS methods: traditional methods for proteins (nano-flow) and lipids (high-flow); single-ome high-flow LC; and MOST. In doing so we obtained a comparable density of protein-lipid correlation for MOST compared to both the separate high flow and the traditional platforms (**Supplementary Figure 4f**). In certain cases, the simultaneous analysis of peptides and lipids afforded by MOST improved the ability to detect coregulated proteins and lipids, here exemplified by the well-known coenzyme-Q biosynthetic pathway (as shown in the nearest neighbor molecule covariance networks in **Supplementary Figure 4g-h**).

In conclusion, the MOST method we describe here provides a framework to offer integrated multi-omics data acquisition in a simple and robust manner. As compared to traditional fractured methodology, this method saves considerable time through the co-analysis of lipids and peptides. Further, we demonstrate that the method generates data of equivalent or higher quality to traditional approaches while using an easy-to-maintain high-flow LC system. That is, with a single LC-MS system, laboratories can now create proteomic and lipidomic data without having to dedicate a system or a specialist to either of these technologies. In its future iteration, we envision expanding the platform to nano-flow LC[20], which will greatly reduce the required sample amounts while boosting proteomic depth, as well as integrating metabolite co-analysis for a more comprehensive biological survey. Here we envision incorporating the use of ion mobility and intelligent mass spectrometer data acquisition to quickly direct MS data acquisition between co-eluting metabolites and peptides. Moreover, we anticipate that future optimization of sample preparation will allow researchers to comprehensively extract and process all three omes from a single sample in a single vial. Such a platform will represent an ideal technology for studies constricted by limited sample amount, namely clinical research and, ultimately, single cell analyses.

Methods

Materials and reagents. Acetonitrile HPLC grade (ACN; Fisher Scientific, cat. no. A955-4), 2-propanol HPLC grade (IPA; Fisher Scientific, cat. no. A461-4), methanol HPLC grade (Fisher Scientific, cat. no. A454SK-4), water HPLC grade (Sigma-Aldrich, cat. no. 270733-4 L), tert-Butyl methyl ether (MTBE; Sigma-Aldrich, cat. no. 443808-1L), ammonium formate (Sigma-Aldrich, cat. no. 516961), formic acid (FA; Thermo Scientific, cat. no. 28905), 2-Chloroacetamide (CAA; Sigma-Aldrich, cat. no. C0267-500G), tris(2-carboxyethyl)phosphine hydrochloride (TCEP; VWR, cat. no. VWRVK831-10G), strata-X 33 um Polymeric Reversed Phase 60 mg/3 mL Tubes (Phenomenex, cat. no 8B-S100-UBJ), sequencing-grade modified trypsin (Promega, cat. no. V5111 or V5113), lysyl endopeptidase (Wako Chemicals, cat. no 129-02541), trifluoroacetic acid (TFA; Thermo Scientific, cat. no. 28904), tris base (EP154-1, cat. no. EP154-1), Guanidine hydrochloride (GnHCl; Sigma-Aldrich, cat. no. G3272-1kg), Urea (Sigma-Aldrich, cat. no. U5378-1kg), Stainless Steel Beads, 5 mm (QIAGEN, cat. no. 69989).

Yeast cultures. The parental (WT) *Saccharomyces cerevisiae* strain for this study was the haploid MATa W303. Single deletion (Δgene) derivatives of W303 were generated using a KanMX or NatMX (only for Δ*coq8*) deletion cassette. All gene deletions were confirmed by PCR.

Yeast from a -80°C glycerol stock were streaked onto plates of YPD agar containing the appropriate antibiotic selection (G418 or Nourseothricin). Plates were incubated at 30°C for two days and subsequently stored at 4°C. For each batch of samples (each batch containing 1 WT culture and 19 Δgene cultures), individual colonies of yeast were picked into 3 mL of liquid YPD medium and incubated (30°C, 230 RPM, 16-20 hrs). Cell density was estimated from optical density (OD600). YPGD media (100 mL at ambient temperature in a sterile 250 mL Erlenmeyer flask) was inoculated with 2.5 x 106 cells and incubated (30°C, 220 RPM). After 25 hrs of incubation, a time point that corresponds to early respiration growth, 1 x 10^8 cells from each 100 mL culture were harvested by centrifugation (3000 g, 3 min, 4°C). Yeast cell pellets were flash frozen in liquid nitrogen and stored at -80°C.

Sample preparation for mass spectrometry. Frozen cell pellets were thawed on ice and mixed with 250 µL of methanol, 750 µL of methyl tert-butyl ether (MTBE), and 200 µL of water. The samples were vortexed for 10 s

and sonicated for 5 min. Phase separation was completed after centrifugation (12,000 g, 5 min, 4 °C). 200 µL of the upper hydrophobic layer was aliquoted into a glass insert amber autosampler vial, dried, and reconstituted in 100 µL of ACN/IPA/H_2O (65:30:5, v/v/v).

To the lower hydrophilic layer, 200 µL of 6M GnHCl and 100 mM tris (pH = 8.0) was added. The samples were boiled at 100 °C for 5 min, rested at room temperature for 5 min, were boiled at 100 °C for 5 min again. Proteins were precipitated by addition with methanol to 90%, vortex for ~ 10 s, and centrifugation at 12,000 g for 5 min. The supernatant was discard, and the resulting protein pellet was resuspended in 8M urea, 10 mM tris(2-carboxyethyl)phosphine (TCEP), 40 mM chloroacetamide (CAA), and 100 mM tris (pH = 8.0). The samples were digested with lysyl endopeptidase (1:50, enzyme/protein) for 4 hrs. The samples were diluted to 2M urea with 50 mM tris and digested with trypsin (1:50, enzyme/protein) overnight.

The pH was adjusted to ~1 by addition of trifluoroacetic acid (TFA), followed by centrifugation at 9,000 g for 5 min. Samples were desalted using Phenomenex Strata-X Polymeric Reversed Phase columns, which were equilibrated with one column volume of 100% acetonitrile (ACN), followed by 0.2% formic acid. Acidified samples were loaded on column, followed by washing with three column volumes of 0.2% formic acid. Peptides were eluted off the column by the addition of 900 µL 80% ACN 0.1% TFA. After being dried, peptides were reconstituted in 0.2% formic acid. Peptide concentration was measured using a quantitative colorimetric peptide assay (Thermo).

MOST LC-MS. LC–MS analysis was performed on a Waters C18 reverse-phase BEH column (150 mm × 1.0 mm × 2.1 µm particle size) at 50 °C and 60 µL/min flow rate. Mobile phase A consisted of 0.2% formic acid in H_2O. Mobile phase B consisted of 0.2% formic acid and 5 mM ammonium formate in IPA/ACN (90:10, v/v). 5 µL of lipids were injected followed by the injection of 20 µg of peptides at 0% of mobile phase B. For the gradient, mobile phase B was held at 0% for 1 min, increased to 28% over 52 min, reached to 70% at 60 min, increased to 100% over 20 min where it was held for 5 min. The column was then reequilibrated for 14 min before the next injection. The LC system was coupled to a Q Exactive HF mass spectrometer by a HESI II heated ESI source (Thermo). Sheath gas was set to 30 units, and auxiliary gas to 6 units, and the spray voltage was set to 4.5 kV. Capillary temperature and aux gas heater temperature were set to 275 °C and 300 °C, respectively. A 34 gauge spray needle was used.

Two scan functions were used to achieve optimal data acquisition for peptides and lipids. For peptides, only positive mode was used. MS1 (MS scan of precursor ions without fragmentation) data were acquired from 0–60 min at a resolution of 60,000 with the AGC target set to 3×10^6, mass range to 300–1350 Th, and maximum injection time to 50 ms. MS2 (MS scan of product ions after fragmentation) data were acquired at a resolution of 30,000 with the AGC target set to 1×10^5, maximum injection time to 60 ms, and loop count to 10 (top10). For lipids, polarity switch was used. MS1 data were acquired from 60–90 min at a resolution of 30,000 with the AGC target set to 1×10^6, mass range to 200–1600 Th, and maximum injection time to 100 ms. MS2 data were acquired at a resolution of 30,000 with the AGC target set to 1×10^5, maximum injection time to 50 ms, and loop count to 2 (top2).

Proteomics nLC-MS (traditional methodology). nLC–MS analysis was performed on an inhouse prepared C18 reverse-phase BEH column (300 mm × 75 µm × 1.7 µm particle size) at 55 °C and 0.315 µL/min flow rate. Mobile phase A consisted of 0.2% formic acid in H_2O. Mobile phase B consisted of 0.2% formic acid in 80% ACN. 2 µg of peptides were loaded onto the column for 4 min. Mobile phase B increases to 10% in the first 4 min then to 55% B at 54 min, 100% B at 55 min, followed by a 4 min wash at 100% B and a 11 min re-equilibration at 0%B.

Eluting peptide cations were converted to gas-phase ions by electrospray ionization and analyzed on a Thermo Orbitrap Eclipse. The spray voltage was set to 2.0 kV in positive ion mode. Ion transfer tube temperature was set to 275 °C. RF level was set to 30%. Survey scans of peptide precursors from 300 to 1350 m/z were performed at 240K resolution (at 200 m/z) with a 250% of normalized AGC target and 50 ms of maximum injection time. Tandem MS was performed by isolation at 0.5 Th with the quadrupole, HCD fragmentation with normalized collision energy of 25, and rapid scan MS analysis in the ion trap. The MS2 normalized AGC target was set to 300% and the max injection time was 14 ms. Only those precursors with charge state 2-5 were sampled for MS2.

The dynamic exclusion duration was set to 10 s with a 25 ppm tolerance around the selected precursor and its isotopes. Cycle time was set to 1 s.

Lipidomics LC-MS (traditional methodology). Sample analysis was performed on an Acquity CSH C18 column held at 50 °C (100 mm x 2.1 mm x 1.7 µm particle size; Waters) using a Vanquish Binary Pump (400 µL/min flow rate; Thermo Scientific). Mobile phase A consisted of 10 mM ammonium acetate in ACN:H_2O (70:30, v/v) containing 250 µL/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA:ACN (90:10, v/v) with the same additives. Mobile phase B was initially held at 2% for 2 min and then increased to 30% over 3 min. Mobile phase B was further increased to 50% over 1 min, then raised to 85% over 14 min, and finally raised to 99% over 1 min and held at 99 % for 7 min. The column was re-equilibrated with mobile phase B at 2% for 1.75 min before the next injection. 10 µL of extract was injected by a Vanquish Split Sampler HT autosampler (Thermo Scientific).

The LC system was coupled to a Q Exactive Orbitrap HF mass spectrometer through a heated electrospray ionization (HESI II) source (Thermo Scientific). Source conditions were as follow: HESI II and capillary temperature at 350 °C, sheath gas flow rate at 25 units, aux gas flow rate at 15 units, sweep gas flow rate at 5 units, spray voltage at |3.5 kV| for both positive and negative modes, and S-lens RF at 90.0 units. The MS was operated in a polarity switching mode acquiring positive and negative full MS and MS2 spectra (Top2) within the same injection. Acquisition parameters for full MS scans in both modes were 30,000 resolution, 1×10^6 automatic gain control (AGC) target, 100 ms ion accumulation time (max IT), and 200 to 1600 m/z scan range. MS2 scans in both modes were then performed at 30,000 resolution, 1×10^5 AGC target, 50 ms max IT, 1.0 m/z isolation window, stepped normalized collision energy (NCE) at 20, 30, 40, and 10.0 s dynamic exclusion.

Data processing. For proteomics, raw data files were processed by MaxQuant[21] (Version 1.5.8.3). Searches were performed against a target-decoy [22] database of reviewed yeast proteins plus isoforms (UniProt, downloaded July 23, 2019) using the Andromeda [23] search algorithm. Searches were performed using a precursor search tolerance of 4.5 p.p.m. and a product mass tolerance of 0.35 Da. Specified search parameters included fixed modification for carbamidomethylation of cysteine residues and a variable modification for the oxidation of methionine and protein N-terminal acetylation, and a maximum of two missed tryptic cleavages. A 1% peptide spectrum match (PSM) false discovery rate (FDR) and a 1% protein FDR was applied according to the target-decoy method. Proteins were identified using at least one peptide (razor + unique). Proteins were quantified using MaxLFQ with a label-free quantification (LFQ) minimum ratio count of 2. LFQ intensities were calculated using the match between runs feature, and MS/MS spectra were not required for LFQ comparisons. Missing values were imputed from normal distribution by Perseus (Version 1.6.0.7), where mean and s.d. equivalent to that of the lowest 1% of measured LFQ intensities was generated. Missing values were filled in with values drawn from this distribution at random. Replicate protein LFQ values from corresponding Δgene or WT strains were pooled, log2 transformed, and averaged (mean log2[strain], n = 3). Average Δgene LFQ intensities were normalized against their appropriate WT control (mean log2[Δgene/WT], n = 3) and a 2-tailed t-test (homostatic) was performed to obtain P values.

For lipidomics, raw data files were processed using Compound Discoverer 2.1 (Thermo Scientific) and Lipidex[24], an in-house-developed software suite. All peaks with a 60.1 min to 89 min retention time and 100 Da to 5000 Da MS1 precursor mass were aggregated into distinct chromatographic profiles (i.e., compound groups) using a 10-ppm mass and 0.5 min retention time tolerance. Profiles not reaching a minimum peak intensity of 5×10^5, a maximum peak-width of 0.75, a signal-to-noise (S/N) ratio of 1.5, and a 5-fold intensity increase over blanks were excluded from further processing. MS/MS spectra were searched against an in-silico generated lipid spectral library containing 35,000 unique molecular compositions representing 48 distinct lipid classes. Spectral matches with a dot product score greater than 500 and a reverse dot product score greater than 700 were retained for further analysis. Lipid MS/MS spectra which contained no significant interference (<75 %) from co-eluting isobaric lipids, eluted within a 3.5 median absolute retention time deviation (M.A.D. RT) of each other, and found within at least 2 processed files were then identified at the individual fatty acid substituent level of structural resolution. If individual fatty acid substituents were unresolved, then identifications were made with the sum of the fatty acid substituents. Replicate lipid intensities from corresponding Δgene or WT strains were pooled, log2 transformed, and averaged (mean log2[strain], n = 3). Average Δgene lipid intensities were normalized

against their appropriate WT control (mean log2[Δgene/WT], n = 3) and a 2-tailed t-test (homostatic) was performed to obtain P values.

Molecule covariance network analysis. Regression analysis was conducted using fold change measurements from all Δgene strains for both molecules in the pair. Pearson's regression analysis was performed to obtain correlation coefficients (r). From these test statistics, P values were calculated using a two-sided Student's t-test. All P values were corrected for multiple hypothesis testing (Benjamini-Hochberg) and correlations where $|r| \geq 0.8$ and $P < 0.001$ were reported. All pairs of covariant molecules are visualized as networks generated using the Gephi open graph visualization platform (version 0.9.2). Complete network layouts were generated using the Fruchterman–Reingold graph-drawing algorithm with area set to 10,000 and gravity set to 30. Where applicable, nearest neighbor covariance networks for certain molecules are visualized through igraph() function[25] in R statistical environment[26].

Acknowledgements

We are grateful for support from NIH P41 GM108538 (JJC) and NIH R35GM131795 (DJP)

Author contributions

YH, ASH, MSW, VL, KO, ES, and JJC designed the MOST method. ER and DJP designed the yeast knock-out study. YH prepared and analyzed samples. YH and ER preformed data analysis. All authors contributed to writing and editing the manuscript.

Conflict of Interest

The authors have no competing interests.

Reference

[1] Nakayasu, ES et al. MPLEx: a Robust and Universal Protocol for Single-Sample Integrative Proteomic, Metabolomic, and Lipidomic Analyses. mSystems **1**(3) (2016).

[2] Zhou, Y et al. Integrated Proteomics and Lipidomics Investigation of the Mechanism Underlying the Neuroprotective Effect of N-benzylhexadecanamide. Molecules, **23**(11), 2929 (2018).

[3] Stefely, JA et al. Mitochondrial Protein Functions Elucidated by Multi-Omic Mass Spectrometry Profiling. Nature Biotechnology **34**, 1191–1197 (2016).

[4] Blum, BC et al. Single-platform 'multi-omic' profiling: unified mass spectrometry and computational workflows for integrative proteomics–metabolomics analysis. Mol. Omics 14, 307-319 (2018).

[5] Misra, B et al. Integrated omics: tools, advances and future approaches. Journal of Molecular Endocrinology **62**(1), R21-R45(2019).

[6] Li, Y. et al. A Novel Approach to the Simultaneous Extraction and Non-Targeted Analysis of the Small Molecules Metabolome and Lipidome Using 96-Well Solid Phase Extraction Plates with Column-Switching Technology, J. Chromatogr. A, **1409** , 277-281 (2015).

[7] Wang, S et al. Simultaneous Metabolomics and Lipidomics Analysis Based on Novel Heart-Cutting Two-Dimensional Liquid Chromatography-Mass Spectrometry, Anal. Chim. Acta, **966**, 34-40 (2017).

[8] Schwaiger, M et al. Merging metabolomics and lipidomics into one analytical run. *Analyst* **144**, 220 (2019).

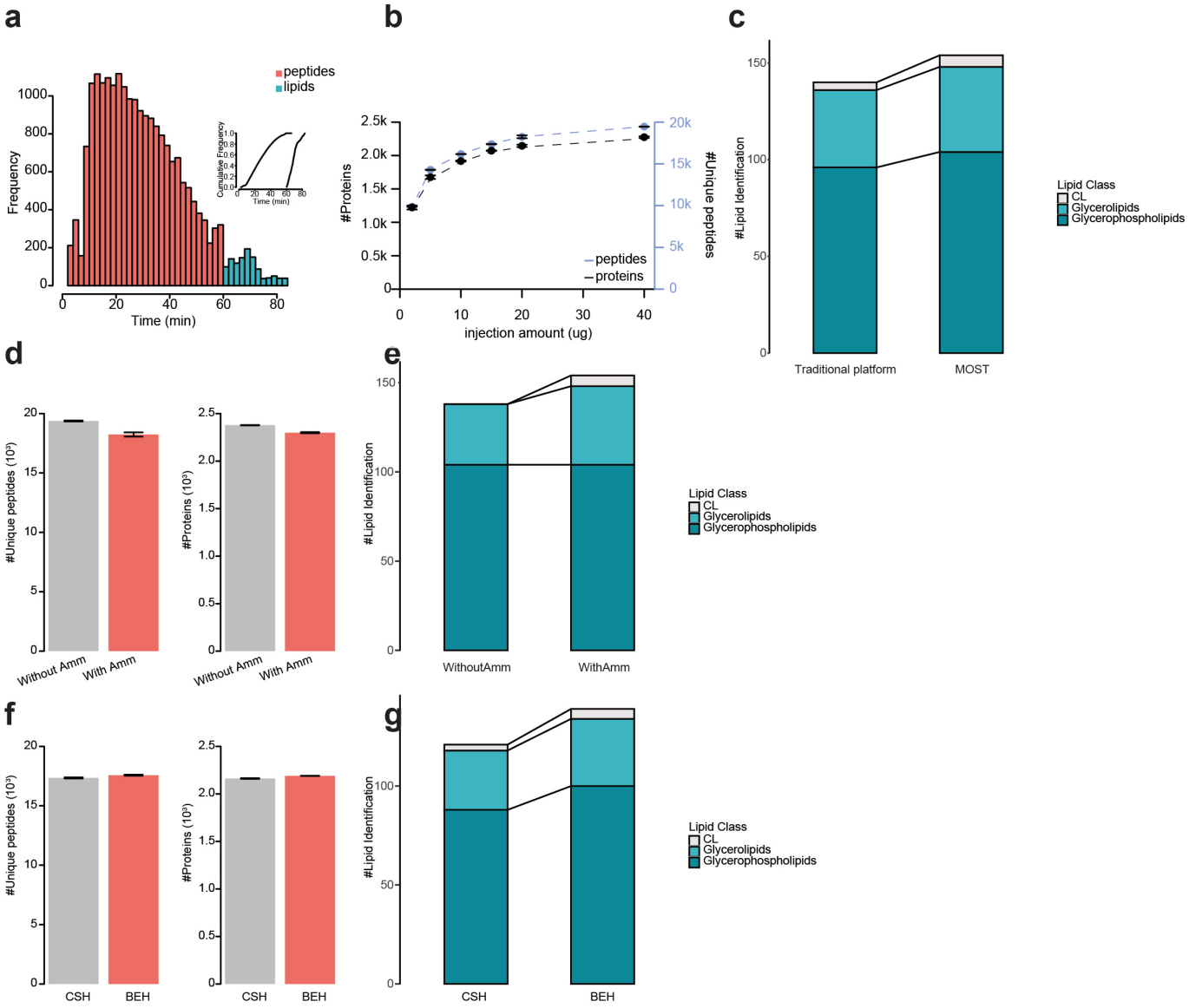
[9] Cajka, T & Fiehn, O. Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *Trends Analyt Chem*. **61**:192–206 (2014).

[10] Ulmer, CZ et al. A Robust Lipidomics Workflow for Mammalian Cells, Plasma, and Tissue Using Liquid-Chromatography High-Resolution Tandem Mass Spectrometry. Methods Mol Biol. 1609:91–106 (2017).

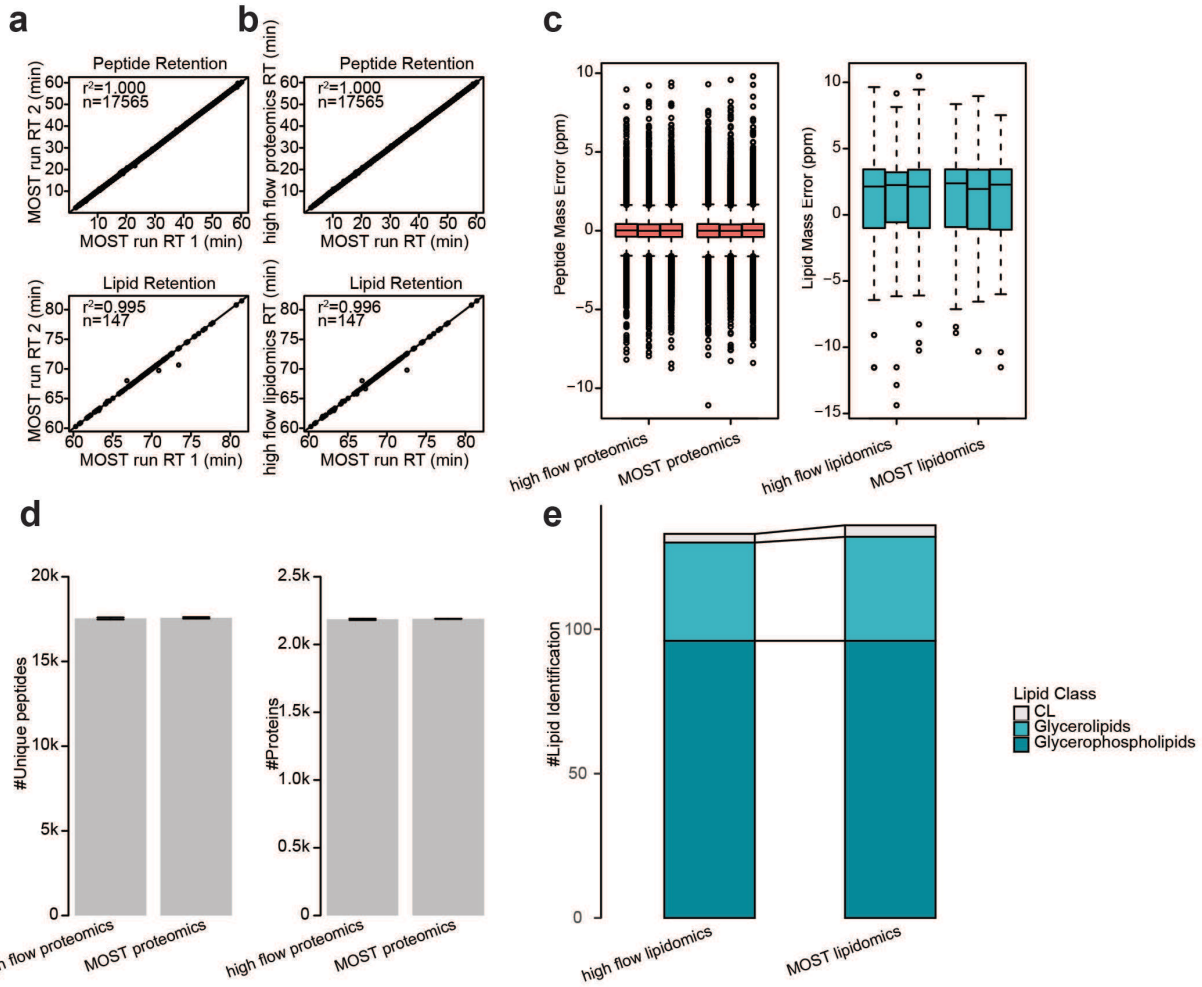
[11] Fernández-Niño, SMG et al. Standard flow liquid chromatography for shotgun proteomics in bioenergy research. *Front. Bioeng. Biotechnol.* **3**(44), 1-7 (2015).

[12] Yin, X et al. Plasma Proteomics for Epidemiology Increasing Throughput With Standard-Flow Rates. *Circ Cardiovasc Genet.* **10**, e001808 (2017).

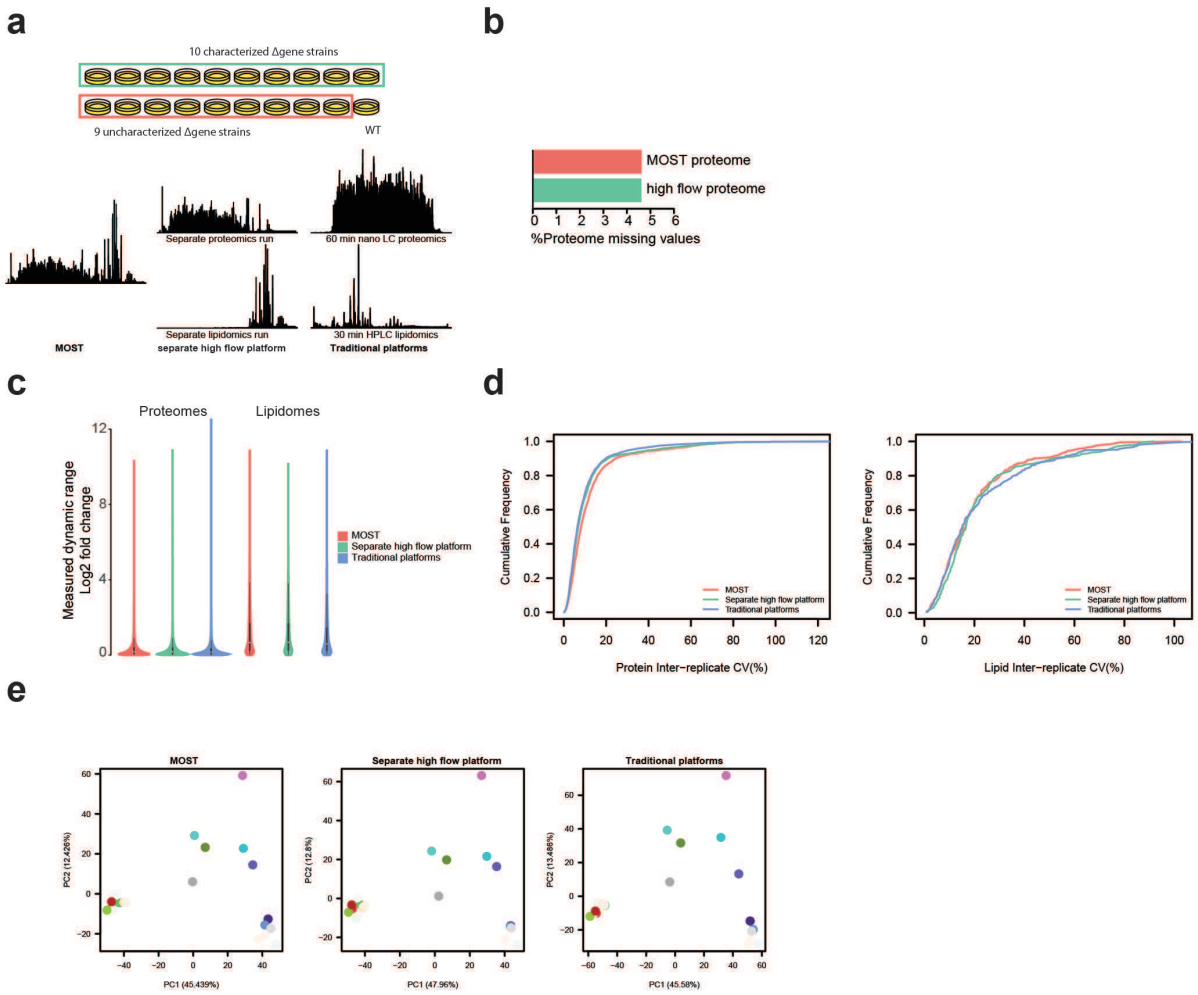
[13] Lenčo, J et al. Conventional-Flow Liquid Chromatography−Mass Spectrometry for Exploratory Bottom-Up Proteomic Analyses. *Anal. Chem.* **90**, 5381−5389 (2018).
[14] Distler U et al. Enhancing Sensitivity of Microflow-Based Bottom-Up Proteomics through Postcolumn Solvent Addition. *Anal. Chem.* **91**, 7510−7515 (2019).
[15] De Craene JO, Coleman J, Estrada de Martin P, et al. Rtn1p is involved in structuring the cortical endoplasmic reticulum. *Mol Biol Cell*. **17**(7):3009‐3020 (2006).
[16] Mast FD, Jamakhandi A, Saleem RA, et al. Peroxins Pex30 and Pex29 Dynamically Associate with Reticulons to Regulate Peroxisome Biogenesis from the Endoplasmic Reticulum. *J Biol Chem*. **291**(30):15408‐15427 (2016).
[17] David C, Koch J, Oeljeklaus S, et al. A combined approach of quantitative interaction proteomics and live-cell imaging reveals a regulatory role for endoplasmic reticulum (ER) reticulon homology proteins in peroxisome biogenesis. *Mol Cell Proteomics*. **12**(9):2408‐2425 (2013).
[18] Marone R, Hess D, Dankort D, et al. Memo mediates ErbB2-driven cell motility. *Nat Cell Biol*. **6**, 515–522 (2004).
[19] MacDonald G, Nalvarte I, Smirnova T, et al. Memo is a copper-dependent redox protein with an essential role in migration and metastasis. *Sci Signal*. **7**(329):ra56 (2014).
[20] Danne-Rasche, N et al. Nano-LC/NSI MS Refines Lipidomics by Enhancing Lipid Coverage, Measurement Sensitivity, and Linear Dynamic Range. *Anal. Chem.* **90**, 8093−8101 (2018).
[21] Cox, J & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. **26**, 1367–1372 (2008).
[22] Elias, JE & Gygi, SP Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods **4**, 207–214 (2007).
[23] Cox, J et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. **10**, 1794–1805 (2011).
[24] Hutchins PJ, Russell JD, Coon JJ. LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. Cell Systems **6**, 1-5 (2018).
[25] Csardi, G & Nepusz, T. The igraph software package for complex network research. InterJournal, Complex Systems, 1695 (2006).
[26] R Core Team. R: A Language and Environment for Statistical Computing. (2017). https://www.R-project.org/



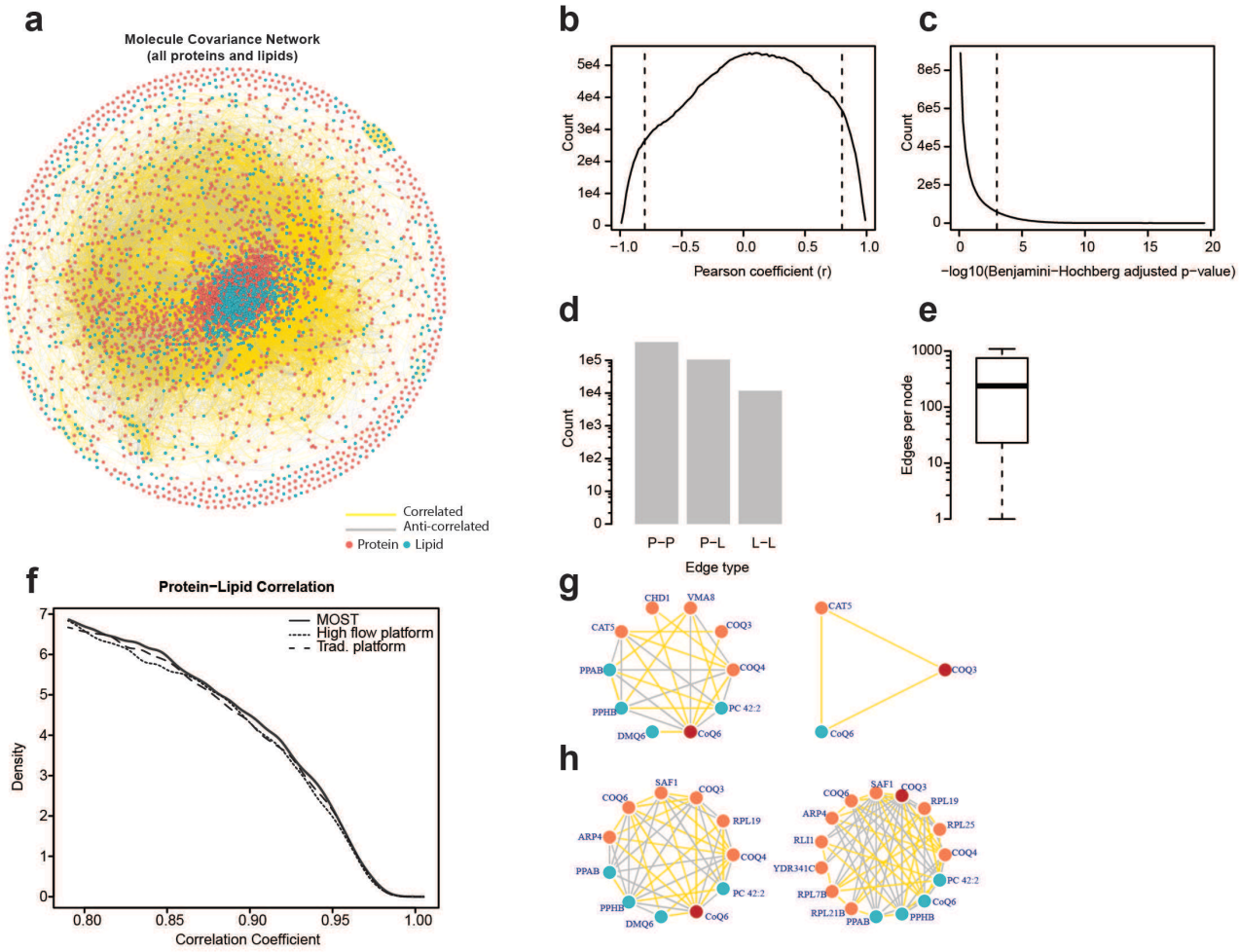
Supplementary Figure 1. Key parameters for method development. (a) Number of peptides and lipids identified versus time, showing the elution of peptides and lipids were separate. (b) 20 ug of peptides were loaded onto column after comparison of a range of injection amount. (c) Stacked barplot showing lipid identification difference between traditional platform and MOST. (d) Barplot showing ammonium reduced peptide and protein identification slightly. (e) Stacked barplot showing ammonium boosted lipid identification. (f) Barplot showing peptide and protein identification was slightly more for BEH column than CSH column. (g) Stacked barplot showing lipid identification was more for BEH column.



Supplementary Figure 2. Additional performance characteristics of MOST. (a) Scatter plot of two MOST runs showing good reproducibility of retention time of peptides and lipids. (b) Scatter plot of a MOST run versus a high flow proteomics/lipidomics run showing no interference on protein/lipid quantification from each other. (c) Boxplot showing comparable mass error between MOST run versus a high flow proteomics/lipidomics run. (d) Barplot showing peptide and protein identification was similar in absence and presence of lipids. (e) Stacked barplot showing lipid identification was similar in absence and presence of peptides.



Supplementary Figure 3. Mass spectrometry analysis metrics and quality assessment for biological study. (a) Overview of MOST, separate high flow platforms and traditional platforms, for comparison of performance. (b) Identical levels of proteome missing values were achieved for MOST and normal high flow proteomics. (c) Violin plots depicting the range of fold changes in molecule abundance (log2[Δgene/WT]) across all molecule classes. (d) Cumulative density plot showing the inter-replicate reproducibility of protein and lipid quantification. (e) PCA plot showing grouping and separation of 20 different conditions. For separate high flow platform, high flow proteomic LC/MS analysis and high flow lipidomic LC/MS analysis were performed separately on the same instrument. For traditional platforms, nano flow proteomic LC/MS analysis and high flow lipidomic LC/MS analysis were performed separately on the distinct instruments.



Supplementary Figure 4. Features of MOST multi-omic molecule covariance networks. (a) Network of all covariant molecules observed in MOST dataset (|r| ≥ 0.8, Benjamini-Hochberg adjusted p-value < 0.001; two-sided Student's t-test). (b) Distribution of calculated Pearson coefficients for all pairwise molecule covariance comparisons (r cutoff at ±0.8 used throughout the study is indicated). (c) Distribution of Benjamini-Hochberg p-values from all pairwise molecule comparisons (p-value cutoff at 0.001 used throughout the study is indicated). (d) Bar chart indicating number of protein–protein (P–P), protein–lipid (P–L), and lipid–lipid (L–L) edges. (e) Box plots indicating the number of edges per node. (f) Density plot showing the density of protein-lipid correlation coefficient. The same samples were analyzed by different platforms for comparison of methods. (g) Nearest neighbor molecule covariance network for CoQ6 (Ubiquinone) and Coq3 gene, generated by high flow platform. (h) Nearest neighbor molecule covariance network for CoQ6 (Ubiquinone) and Coq3 gene, generated by MOST.

Figures

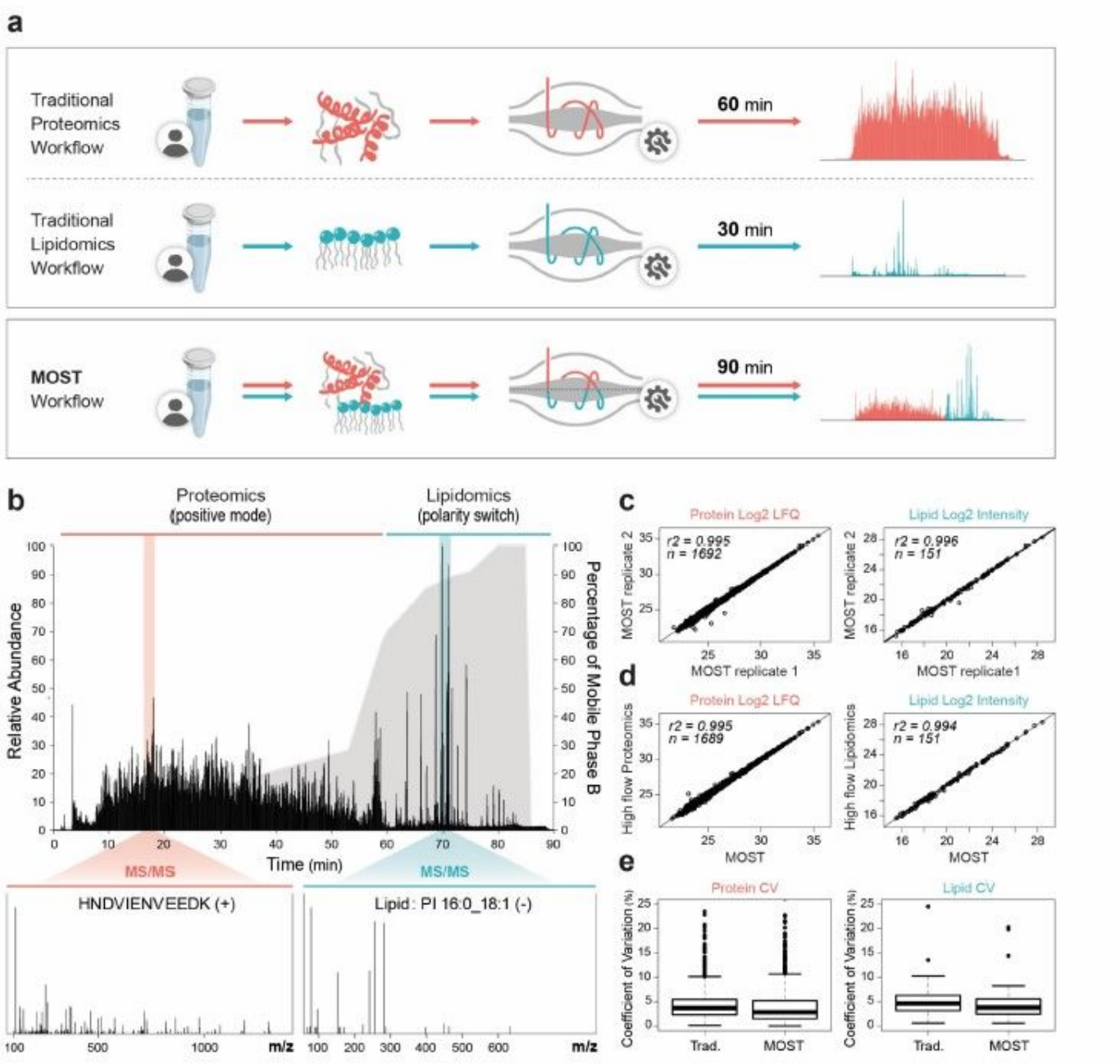


Figure 1

Performance characteristics of MOST. (a) Diagram showing comparison between traditional workflow and MOST workflow. (b) An example of total ion current (TIC) chromatogram of MOST, with the first 60 min for proteomic analysis and last 30 min for lipidomics analysis, along with two examples of MS/MS spectra of a peptide and a lipid. (c) Scatter plot of two MOST runs showing good reproducibility of quantification of proteins and lipids. (d) Scatter plot of a MOST run versus a high flow proteomics/lipidomics run showing no interference on protein/lipid quantification from each other. (e) Boxplot comparing coefficient of variation (CV) of proteins and lipids between traditional platforms and MOST (n=3).

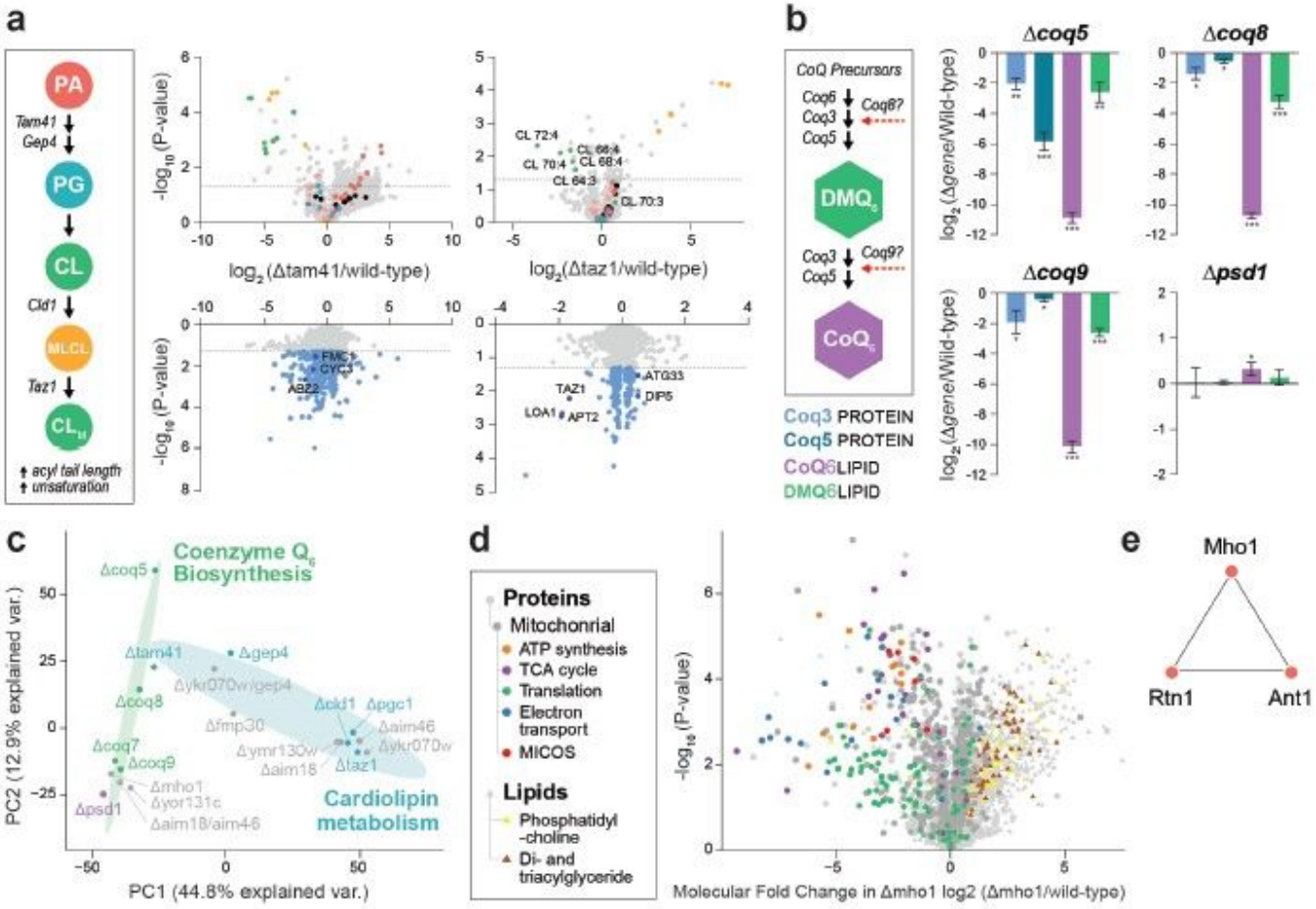


Figure 2

Biological study by MOST. (a) Schematic pathway of cardiolipin biosynthsis and volcano plots across related gene knockout strains. PA, phosphatidic acid. PG, phosphoglycerol. CL, cardiolipin. MLCL, monolysocardiolipin. (b) Schematic pathway of Co-enzyme Q6 biosynthsis and molecule abundance across related gene knockout strains. *, p-value < 0.05. **, p-value < 0.01. ***, p-value < 0.001. (c) Principal component analysis (PCA) of Δgene strains. (d) Volcano plot showing average fold-change in molecule abundances (mean log2[Δmho1/wild-type]) versus statistical significance, showing select functional groups (GO terms and lipid class) significantly (Benjamini-Hochberg adjusted p-value < 0.05) enriched in either upregulated or downregulated molecules. (e) Nearest neighbor covariance network for Mho1.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.jpg
- FigureS2.jpg
- FigureS3.jpg

- FigureS4.jpg