Revealing dynamic protein acetylation across subcellular compartments

Authors: Josue Baeza*[1,2], Alexis J. Lawton*[1,2], Jing Fan[2,3], Michael J. Smallegan[1,2], Ian Lienert[4], Tejas Gandhi[4], Oliver M. Bernhardt[4], Lukas Reiter[4], John M. Denu[1,2,5]

Author Affiliations:

[1]Department of Biomolecular Chemistry, University of Wisconsin, Madison, WI

[2]Wisconsin Institute for Discovery, University of Wisconsin, Madison, WI

[3]Morgridge Institute for Research, University of Wisconsin, Madison, WI

[4]Biognosys AG, Wagistrasse 25, CH-8952 Schlieren, Switzerland

[5]Corresponding author. Email- john.denu@wisc.edu

*Co-first authorship

Running Title:

Dynamic protein acetylation across subcellular compartments

Abbreviations:

ANOVA – analysis of variance

ACN – acetonitrile

DDA – data-dependent acquisition

DIA – data-independent acquisition

DDT – dithiothreitol

GO – gene ontology

HCD – higher energy collisional dissociation

HPRP – high pH reverse phase

IAA – iodoacetamide

KAT – lysine acetyltransferase

KDAC – lysine deacetylase

NCE – normalized collision energy

PSM – peptide spectrum matches

QSSA – quantitative site set functional score analysis

Highlights:

- Developed an integrated workflow for quantifying lysine acetylation stoichiometry and proteomic abundances.
- Identified site-specific acetylation stoichiometry distribution across cellular compartments
- Revealed rapid changes in protein acetylation that define large groups of sites and major pathways

- Determined that highest acetylation stoichiometries are found in compartments with well-described acetyltransferase enzymes.

In Brief:

Lysine acetylation is a widespread post-translational modification, but the dynamics of rapid changes in acetylation stoichiometry across cellular compartments are not known. Here, we developed an improved mass-spectrometry method to quantify the distribution and dynamic nature of acetylation stoichiometry upon growth factor stimulation. Nuclear proteins displayed the highest acetylation levels, and serum starvation-refeeding led to rapid acetylation changes, revealing major cellular pathways such as splicing, translation, and protein homeostasis are dynamically acetylated.

Abstract:

Protein acetylation is a widespread post-translational modification implicated in many cellular processes. Recent advances in mass spectrometry have enabled the cataloging of thousands of sites throughout the cell, however identifying regulatory acetylation marks have proven to be a daunting task. Knowledge of the kinetics and stoichiometry of site-specific acetylation are important factors to uncover function. Here, an improved method of quantifying acetylation stoichiometry was developed and validated, providing a detailed landscape of dynamic acetylation stoichiometry within cellular compartments. The dynamic nature of site-specific acetylation in response to serum stimulation was revealed. In two distinct human cell lines, growth factor stimulation led to site-specific, temporal acetylation changes, revealing diverse kinetic profiles that clustered into several groups. Overlap of dynamic acetylation sites among two different human cell lines suggested similar regulatory control points across major cellular pathways that include splicing, translation, and protein homeostasis. Rapid increases in acetylation on protein translational machinery suggest a positive regulatory role under pro-growth conditions. Lastly, higher median stoichiometry was observed in cellular compartments where active acetyltransferases are well-described.

Introduction:

Protein lysine acetylation is now acknowledged as a widespread modification, rivaling phosphorylation in scope (1). Protein acetylation was first characterized on the N-terminal lysine residues of histone proteins that wrap DNA in the nucleus (2, 3) and has since been described throughout the cell including cytoplasm (4), mitochondria (5), endoplasmic reticulum (6), and peroxisomes (7). In the nucleus, histone acetylation is associated with active gene expression, acting in part to open chromatin and allowing access for transcriptional machinery. Acetylation of cytoplasmic proteins affects diverse cellular processes which include cell migration, cytoskeleton dynamics, metabolism, and aging. Mitochondrial protein acetylation has been linked to metabolic regulation, oxidative stress, OXPHOS, and mitochondrial gene expression (8).

Reversible acetylation is catalyzed by the enzymatic activity of lysine acetyltransferases (KATs) and deacetylases (KDACs), however other evidence also suggests that a significant proportion of acetyl-lysine sites result from nonenzymatic mechanisms (9–11). We previously measured second-order rate constants for nonenzymatic acetylation *in vitro* and found that the chemical conditions of the mitochondrial matrix was sufficient to drive considerable nonenzymatic acetylation, and was generally consistent with lysine sites with low *in vivo* acetylation stoichiometry (12). With the use of a general-base catalyst in the active site of KAT families that include GCN5 and MYST, these reactions are not affected by the protonation state of the ε-amino group, whereas the nonenzymatic reactions are directly dependent on the amount of unprotonated lysine (13, 14).

Mass spectrometry has enabled the identification of over 20,000 acetylation sites in human cells (15). This comprehensive catalog was facilitated by the development of antibody enrichment strategies for acetylated peptides (16, 17). While immunoenrichment of acetylated peptides helps with identification of acetyl-lysine sites on lowly abundant proteins, it can also introduce potential quantification bias through adding additional experimental steps and antibody selectivity biases.

Unlike a plethora of examples from protein kinase signaling pathways, lysine acetylation has not been associated with analogous cascades, where one acetylation event of an acetyltransferase leads to

acetylation of a second acetyltransferase to transmit a biological signal. Such cascades are used to amplify and rapidly propagate information down a signal transduction pathway. From available evidence, acetylation can modulate protein-protein and protein-DNA interactions, cellular localization, enzyme activity and stability (8). For example, bromodomain-containing proteins recognize and bind acetyl-lysine residues for recruitment of larger multi-subunit complexes and permit efficient activation of gene transcription (18, 19). In mitochondria, current evidence suggests that protein acetylation generally serves as an inhibitory modification of metabolic enzymes (8). In this regard, acetylation appears to function as a rheostat to modulate the degree of a biochemical process. Given these examples of regulation, quantifying the level of stoichiometry is critical for understanding the biological effect of acetylation.

Protein phosphorylation cascades are well-known mechanisms by which cells respond in seconds to minutes to external stimuli. However, the time scales at which protein acetylation occur is poorly understood. Determining both stoichiometry and dynamic responses during cellular stimulation are key features to understand the role of protein acetylation at a site- and protein-specific level. In this study, we provide an improved method using data-independent acquisition (DIA) to quantify acetylation stoichiometry at the proteome level, which was benchmarked using cellular proteomes with defined acetylation stoichiometry. To understand how acetylation and proteome dynamics are concomitantly modulated in cells, two human cell lines were synchronized by serum depletion/refeeding and monitored for changes in the proteome and site-specific acetylation, revealing rapid and dynamic changes in acetylation and protein expression profiles. Quantifying acetylation stoichiometry dynamics will be a critical tool for prioritizing the ever-increasing number of detected lysine acetylation sites for further investigation and towards a deeper understanding of this regulatory modification.

Experimental Procedures:

Experimental Design and Statistical Rationale

Samples: For NCE optimization, MCF7 cultured cells were acquired using eight NCE parameters as described. For the acetylation stoichiometry calibration curve, a cell culture stock of HEK293 cells was used to prepare the light and heavy acetyl-modified samples which are combined into eleven mixed samples at varying amounts as described. Steady-state and dynamic stoichiometry analysis (MCF7 and HCT116 cells) were measured in biological triplicate. Samples used for spectral library generation were also completed in biological triplicate.

Statistical tests: Various statistical tests including linear regression analysis and nonparametric analysis were performed throughout the study. Linear regression was performed on the stoichiometry calibration curve, while nonparametric tests were used to compare changes in the distribution of acetylation stoichiometry.

Cell Culture conditions

MCF7 and HCT116 cells were grown using DMEM supplemented with 10% FBS. For global acetylation stoichiometry and single amino acid analysis, MCF7 and HCT116 cells were harvested at ~80% confluency. Four hours prior to harvesting, cells were washed with PBS and replaced with fresh media. MCF7 and HCT116 cells were cultured for a total of 48 hours.

Sample preparation

<u>Protein chemical acetylation and digestion</u>

Equal amount of protein (200 µg) was resuspended into 25-30 µL of urea buffer (8 M urea (deionized), 500 mM ammonium bicarbonate pH = 8.0, 5 mM DTT). Incubation steps throughout the sample preparation are carried out using the Eppendorf ThermoMixer® C. Sample was incubated at 60 °C for 20 minutes while shaking at 1500 RPM. Cysteine alkylation was carried out with 50 mM iodoacetamide

and incubating for 20 minutes. Chemical acetylation of unmodified lysine residues was performed as previously described (12, 20, 34). Briefly, ~20 µmol of the "light" ^{12}C-acetic anhydride (Sigma) or "heavy" D_6-acetic anhydride (Cambridge Isotope Laboratories) was added to each sample and incubated at 60 °C for 20 minutes at 1500 RPM. The pH of each sample was raised to ~8 using ammonium hydroxide and visually checked with litmus paper. Two rounds of chemical acetylation were performed for each sample to ensure near-complete lysine acetylation. To hydrolyze any O-acetyl esters formed during the chemical acetylation, the pH of the sample raised to ~8.5 and each sample was incubated at 60 °C for 20 minutes at 1500 RPM. For protein digestion, the urea concentration of each sample was diluted to ~2 M by adding 100 mM ammonium bicarbonate pH = 8.0 followed by addition of trypsin (Promega) at a final ratio of 1:100. The sample was digested at 37 °C for 4 hours while shaking at 500 RPM. If a second digestion using gluC (Promega) occurred, the urea concentration was further diluted to ~1 M using 100 mM ammonium bicarbonate pH = 8.0 and digested with gluC (1:100) at 37 °C overnight while shaking at 500 RPM. Each sample was acidified by the addition of 15 µL of acetic acid.

Digesting protein sample to single amino acids

For complete digestion of proteins, which converts all unmodified lysine residues to free lysine, and all N-ε-acetylated lysine residue to acetyl-lysine, 20 µg of sample was diluted into 50 µL of digestion buffer (50 mM ammonium bicarbonate, pH 7.5, 5 mM DTT, in LC-MS grade water). A sample with 50 µL digestion buffer without protein was also included as a procedural blank. The samples were digested to single amino acids by treatment with three enzymes sequentially: First, samples were treated with 0.4 µg Pronase and incubated for 24 hr at 37°C. Then the Pronase activity was stopped by heating to 95°C for 5 min. After cooling down to ambient temperature, samples were then treated with 0.8 µg aminopeptidase and incubated at 37°C for 18 hr. Aminopeptidase activity was again stopped by heating samples to 95°C for 5 min and cooling down. Finally, samples were digested with 0.4 µg prolidase and incubated at 37°C for 3 hr. To extract the resulted single amino acids, 200 µL LC-MS grade acetonitrile (ACN) was added to

each sample. The mixture was vortexed for 5 sec, spun at maximal speed for 5 min, and the supernatant was saved for analysis by LCMS.

Offline High pH Reverse Phase (HPRP) Prefractionation

Chemically acetylated peptides were resuspended into ~2mL of HPRP buffer A (100 mM Ammonium Formate pH = 10) and injected onto a pre-equilibrated Phenomenex Gemini® NX-C18 column (5µm, 110Å, 150 x 2.0mm) with 2% buffer B (10% Buffer A, 90% acetonitrile). Peptides were separated with a Shimadzu LC-20AT HPLC system using a 2% - 40% buffer B linear gradient over 30 minutes at 0.6 mL/min flow rate, collecting 24 fractions throughout the length of the gradient. Fractions were dried down using a speedvac and pooled by concatenation into 6 final fractions as described previously (60).

Mass spectrometry

Liquid chromatography

Peptides were separated with a Dionex Ultimate 3000 RSLCnano HPLC using a Waters Atlantis dC18 (100 µm x 150 mm, 3µm) C18 column. The mobile phase consisted of 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B). Peptides were eluted with a linear gradient of 2 – 35% B at a flow rate of 800 nL/min over 90 minutes. Peptides were injected by nanoelectrospray ionization (Nanospray Flex™) into the Thermo Fisher Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass spectrometer.

Data-dependent acquisition mass spectrometry

For data-dependent acquisition (DDA), the MS survey scan was performed in positive ion mode with a resolution of 70,000, AGC of 3e6, maximum fill time of 100 ms, and scan range of 400 to 1200 m/z in profile mode. Data dependent MS/MS was performed in profile mode with a resolution of 35,000, AGC of 1e6, maximum fill time of 200 ms, isolation window of 2.0 *m/z*, normalized collision energy of 25, dynamic exclusion was set for 30 seconds, and a loop count of 20.

Data-independent acquisition mass spectrometry

For data-independent acquisition (DIA), the MS survey scan was performed in profile mode with a resolution of 70,000, AGC of 1e6, maximum fill time of 100 ms in the scan range between 400 and 1000 *m/z*. The survey scan was followed 30 DIA scans in profile mode with a resolution of 35,000, AGC 1e6, 20 *m/z* window, and NCE of 25 or 30. For both DDA and DIA methods, the source voltage was set at 2000 V and capillary temperature at 250 °C.

LCMS analysis of single amino acids

The abundances of free lysine, acetyl-lysine, and other amino acids from completely digested protein samples were analyzed using a Thermo Fisher Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass spectrometer coupled to a Dionex UltiMate 3000 UHPLC system. Samples are separated using a 5 µm polymer 150 2.1 mm SeQuant® ZIC®-pHILIC column, with the following gradient of solvent A (ACN) and solvent B (10 mM ammonium acetate in water, pH 5.5) at a flow rate of 0.3 mL/min: 0-2min, 10% solvent B; 2-14min, linearly increase solvent B to 90%; 14-17min, isocratic 90% solvent B; 17-20min, equilibration with 10% solvent B. Samples are introduced to the mass spectrometer by heated electrospray ionization. Settings for the ion source are: 10 aux gas flow rate, 35 sheath gas flow rate, 1 sweep gas flow rate, 3.5 kV spray voltage, 320°C capillary temperature, and 300°C heater temperature. Analysis is performed under positive ionization mode, with scan range of 88–500 m/z, resolution of 70 K, maximum injection time of 40 ms, and AGC of 1E6.

To quantify absolute levels of lysine and acetyl-lysine, an external calibration curve was run in the same sequence with the experimental samples. Lysine standard ranges between 10 to 200 µM, and acetyl-lysine standard ranges between 0.5 to 10 µM. Signal from procedural blank was subtracted from samples.

Data Processing

Generating ^{12}C-AcK and D_3-AcK Spectral Library

The spectral library consists of a catalog of high-quality MS/MS fragmentation spectra resulting from data-dependent acquisition (DDA) MS runs. For the MCF7 and HCT116 stoichiometry, we performed DDA runs on three MCF7 lysate samples which were chemically acetylated with ^{12}C-acetic anhydride, digested with trypsin and gluC, followed by HPRP prefractionation (see above). Prior to MS analysis, iRT peptides (Biognosys) were spiked into each sample following manufacturer's guidelines. Database search was performed using MaxQuant with Andromeda as the peptide search engine version 1.6.1.0 using lysine acetylation and methionine oxidation as variable modifications and cysteine carbamidomethylation as a fixed modification. Enzymes for digestion were set to trypsin, which cleaves after lysines and arginines, and gluC, which cleaves after glutamate. We increased the maximum missed cleavages to 4, because our labeling scheme, which modifies all unmodified lysines, prevents cleavage from trypsin. PSM and Protein FDR were both set to 1% calculated by the target-decoy approach, per default settings. Decoy entries were created in MaxQuant by reversing the original protein sequences. First search peptide mass tolerance was set to 20 ppm and main search peptide mass tolerance was set to 4.5 ppm, per default settings. Mass tolerance for fragment ions was set to 0.5 Da. The DDA runs were searched against the Swiss-Prot reviewed sequence database downloaded from UniProtKB on 12/12/2017 (20244 entries). The MaxQuant search results were imported into Spectronaut to build the 12C-AcK library. The 12C-AcK spectral library was then exported as a spreadsheet from Spectronaut and imported into a custom spectral library modifier, which completes the spectral library for all combinations of light and heavy acetylated peptides. The custom spectral library modifier can be accessed from the GitHub (DOI). With this *in silico* approach to inflate the 12C-AcK library, every acetylated peptide precursor will be represented by 2^n versions differing in the number and position of heavy/light acetylated lysine, where n is the number of acetylation sites in the peptide. The spectral library was completed with the corresponding precursor m/z values and fragment m/z values. The most intense fragment ions selected from the initial MS2 spectrum were cloned to the other

peptide precursor versions. All peptide precursor versions will have identical retention time and hence iRT was also cloned. The generated library contained 468585 entries corresponding to 2846 proteins. MaxQuant result files and inflated library can be accessed through ProteomExchange via the MassIVE repository (PXD014453).

DIA MS data analysis

Data from DIA-MS was analyzed using Spectronaut10 and Spectronaut Pulsar. Thermo raw files were converted to HTRMS files with the Spectronaut Raw to HTRMS converter using the default settings and input into Spectronaut. The Spectronaut default settings for quantitation were used with slight modification: Identification-Qvalue score was set to 0.1 and Workflow-Unify peptide peaks were selected. This will cause Spectronaut to use the same integration boundaries for all light/heavy versions of one acetylated peptide within one LC-MS run. This change in the workflow will instruct Spectronaut to select for a given acetylated peptide precursor the best signal (by q-value) of the 2^n versions in the spectral library (see above). With this workflow, Spectronaut will then transfer the integration boundaries of the best scoring peptide precursor to the other peptide precursors. Because all of the 2^n peptide precursor versions only differ by the number of heavy instead of light acetylated lysine the retention time is expected to be identical. The spectral libraries which were completed as described above for all the light/heavy peptide precursor versions were used with this workflow. A Spectronaut output file containing all the fragment ion peak areas along with the corresponding peptide and protein identification was exported and used to compute the lysine site stoichiometry. A list containing all the data categories used for downstream stoichiometry analysis is found in the supplemental information.

Stoichiometry data processing

Data processing was performed in R v3.5.0 (http://www.r-project.org/) using an in-house made R script, which is available in the supplementary information. The stoichiometry preprocessing pipeline consists of two major steps: quantifying fragment ion stoichiometry and natural abundance isotopic

correction. R-scripts utilized to analyze the results presented here can be accessed through GitHub (DOI:10.5281/zenodo.3360892).

Quantifying site-specific stoichiometry

DIA MS measures multiple peptide fragment ion abundances so this approach allows for quantitation of multiple lysines within a peptide. Acetylation stoichiometry of unique lysine sites are quantified by matching light and heavy fragment ion pairs and using the equation:

$$\frac{XIC_L}{XIC_L + XIC_H} \qquad \text{Equation 1}$$

where XIC_L is the peak area of the light fragment ion and XIC_H is the peak area of the heavy fragment ion.

Isotopic purity correction

The mass shift of the light and heavy AcK peptides is 3 Da. This causes the M+0 peak of the heavy AcK peptide to overlap with the M+3 peak of the light AcK peptide. Therefore, we are correcting for the isotopic distribution overlap between the peptide pairs. This is done using an in-house R script as well as the R package, BRAIN v1.16.0 (Baffling Recursive Algorithm for Isotopic DistributioN calculations), available from Bioconductor, the open-source, software project (http://www.Bioconductor.org/) (61). To correct for natural abundance of ^{13}C isotope, the M+0 and M+I, where I represents the isotopic mass shift +1 or +3, were used to calculate the correction coefficient.

$$\text{Correction coefficient} = \frac{1}{\left(\frac{M+0}{M+I}\right)} \qquad \text{Equation 2}$$

The correction coefficient is used to calculate the correction value:

$$\text{Correction value} = XIC_L * \text{correction coefficient} \qquad \text{Equation 3}$$

where the XIC_L is the peak area of the light fragment ion. Finally, the corrected heavy peak area ($^{Corr}XIC_H$) is calculated:

$$^{Corr}XIC_H = XIC_H - \text{Correction value} \qquad \text{Equation 4}$$

where the XIC_H is the peak area of the heavy fragment ion. The corrected stoichiometry is quantified using equation 1, substituting with $^{Corr}XIC_H$.

MSstats

Protein abundance summarization was performed using MSstats v3.12.0 with the output of Spectronaut as the input. The function "SpectronauttoMSstatsFormat" was used with the following arguments: intensity set to "PeakArea", filter_w_Qvalue set to TRUE, qvalue_cutoff set to 0.01, useUniquePeptide set to TRUE, fewMeasurements set to "remove", removeProtein_with1Feature set to FALSE, and summaryforMultipleRows set to "max". The dataProcesses function was then performed using the default arguments.

NCE Optimization

To quantify site-specific acetylation stoichiometry from peptides containing multiple lysines, the fragmentation spectra of precursor ions must contain a high b- and y-ion coverage. To this end, we compared and optimized the number of peptide spectral matches (PSMs) as well as b- and y-ion coverage of MCF7 peptides (chemically acetylated with ^{12}C-acetic anhydride followed by trypsin and gluC digestion) with a Q-Exactive MS using varying NCE settings (15, 20, 25, 30, 35, 40, 45, 50). For all NCE conditions, precursors between 400 - 1200 *m/z* were selected for fragmentation. MS1 resolution was set to 70,000, 3e6 target AGC, and 100 ms max IT in profile mode. MS2 resolution was set to 35,000, 1e6 target AGC, 200 ms max IT in profile mode with 15-sec dynamic exclusion. Database search was performed using MaxQuant version 1.5.4.1 followed by data analysis in R.

Stoichiometry curve

We determined the accuracy and precision of the stoichiometry method by generating an 11-point stoichiometry curve using a complex sample. For this, we used a HEK293 lysate that was grown using standard culture conditions and harvested by centrifugation. The packed cell volume was resuspended using urea buffer (6-8M urea, 100mM ammonium bicarbonate pH = 8.0) and lysed by sonication. Protein concentration was measured using Bradford reagent (Bio-Rad).

To quantify stoichiometry ranging between 1-99%, we varied the amount of starting material to be chemically acetylated with ^{12}C-acetic anhydride or D_6-acetic anhydride using a total of 200 µg of protein for each stoichiometry point. For example, to measure a sample as 10% acetylated, we labeled 20 µg of HEK293 lysate with ^{12}C-acetic anhydride and 180 µg of HEK293 lysate with D_6-acetic anhydride. The starting protein amounts were varied to generate stoichiometries of: 1, 5, 10, 20, 40, 50, 60, 80, 90, 95, and 99% acetylation. Upon chemical acetylation, the sample was pooled together, digested using trypsin and we performed an offline HPRP prefractionation as outlined above.

Bioinformatics

Subcellular localization assignment

To assign protein subcellular localization, we used the MitoCarta (62, 63) and UniProt (http://www.uniprot.org/) databases. For "Mitochondrial" assignment of proteins, we used the Mitocarta database. Additionally, we used "Subcellular location" or "GO - Cellular component" from the UniProt database to assign "Mitochondrial", "Nuclear", and "Cytoplasmic" pools. Other subcellular locations, such as endoplasmic reticulum, Golgi apparatus, cell membrane, etc., were assigned to the "Nuclear" fraction due to the likelihood that these cellular compartments, during differential centrifugation, would sediment in the "Nuclear" spin, which occurs at 1000 xg.

Quantitative Site set functional Score Analysis (QSSA)

The intersection of the KEGG pathway map (64) and proteins in the spectral library detected with < 1% FDR was used for the gene set background. Acetylation coverage for each (p) pathway was calculated as the number of acetyl sites identified (n_{ack}) over the total number of lysines in the pathway (n_k), counted using protein sequences from UniProt. The extent of acetylation was taken into account by summing the acetylation stoichiometry (s) across all conditions and all sites in each pathway. To allow for combining acetylation coverage and stoichiometry, the standard score of each quantity was taken. The overall pathway score was then calculated as the sum of the individual z-scores.

$$QSSA_p = z\left(\frac{n_{acK}}{n_K}\right) + z\left(\sum_{i \in p} s_i\right)$$

Functional analysis

Functional annotation of enriched gene ontology (GO) terms was assessed using DAVID v6.8(45, 46). For enrichment analysis, the background was set to all the proteins identified in the DDA spectral library totaling 2400 unique protein IDs. The list of proteins with significantly changing acetylation sites in both serum-stimulated MCF7 and HCT116 cells were analyzed using DAVID for the following GO terms: GOTERM_BP_DIRECT, GOTERM_CC_DIRECT, and GOTERM_MF_DIRECT. GO term fold-enrichment was plotted as a bar graph with the corresponding p-value in overlaying, white text. Terms were grouped according to DAVID Functional Annotation Clustering.

STRING Network analysis

The underlying interaction network was downloaded from the STRING database (version 11.0)(47). The thickness of edges in the STRING network display interaction confidence. Clusters of interactions were determined using k-means clustering with a set number of four clusters.

Results:

DIA acetylation stoichiometry method optimization

We previously reported a method to determine lysine acetylation stoichiometry across an entire proteome (20). This method employed an isotopic chemical acetylation approach to label all unmodified lysine residues within a sample and, upon proteolytic digestion coupled to LC-MS/MS, has been utilized to quantify proteome-wide acetylation stoichiometry in various biological conditions (20–23) as well as in *in vitro,* nonenzymatic acetylation kinetics (12). Here, we utilize an improved method to quantify acetylation stoichiometry using peptide prefractionation coupled with data-independent acquisition (DIA) mass spectrometry (**Figure 1A and Figure S1**). Briefly, a sample is chemically acetylated using isotopic acetic anhydride and digested with trypsin and gluC. The sequential digestion of the acetyl-proteome generates shorter peptides for MS analysis. Peptides are then prefractionated offline using high pH reversed-phase (HPRP) chromatography, analyzed using nano-LC-MS/MS in DIA mode and analyzed using Spectronaut (v10) (24–27). A project-specific spectral library is generated from a ^{12}C-acetic anhydride modified sample and acquired in data-dependent acquisition (DDA) mode. To account for both light and heavy acetyl-lysine fragment ions in the spectral library, a novel, standalone software was developed to be used with Spectronaut which generates the heavy labeled fragment ion *in silico* from the ^{12}C-AcK spectral library. This process assures that for a given peptide all heavy and light pairs are contained within the spectral library, rigorously requiring both isotopic pairs to be quantified before stoichiometry is calculated. Combining offline prefractionation and DIA analysis addresses unique limitations of the original study as previously discussed (23, 24, 27–30). HPRP prefractionation reduces interferences caused by coeluting peptides and has the added benefit of increasing the depth of the acetylome coverage (31). DIA analysis uses all quantified light and heavy fragment ions in MS2 which are used to calculate stoichiometry. Additionally, stoichiometry can be localized to the lysine site even when multiple lysines are present on a peptide when discriminating fragment ions are present (**Figure 1B, C**).

To assess the accuracy and precision of the improved workflow, a proteome-wide acetylation stoichiometry calibration curve was generated (**Table S1**). For this analysis, a proteome sample of HEK293 cells was chemically acetylated with either light (^{12}C-) or heavy (D_6-) acetic anhydride, which were then combined at varying ratios and subjected to the DIA workflow (**Figure 2A**). A cell-based proteome was used for method validation over standard peptides, which had several advantages: 1. The use of standard peptides at this scale would be cost prohibitive. 2. The combined acetyl-proteomes results in comprehensive acetylation stoichiometry, mimicking experimental samples. 3. The level of acetylation can be modulated to encompass a wide range of stoichiometries (1% - 99%) providing limits of sensitivity and quantification. A caveat of using a cell-based workflow is that endogenous acetylation can potentially confound the results, which can lead to overestimation of some lysine sites. However, because the vast majority of lysine sites do not display high levels of acetylation, this validation approach can be used to assess the accuracy and precision of the stoichiometry workflow.

Due to the 3 Da mass shift between the light and heavy acetyl peptide fragment ions, high acetylation stoichiometry will lead to an underestimation of stoichiometry due to increased intensity of the M+3 natural abundance isotopic peak (**Figure 2B**). To account for this, a natural abundance correction was applied to all heavy acetyl lysine fragment ions. This correction was performed by subtracting the M+3 isotopic peak of the light acetyl lysine fragment ion from the M+0 isotopic peak of the heavy acetyl lysine fragment ion (**Figure 2B**). This global correction improved the precision of the stoichiometry quantification, especially in the higher stoichiometry values (**Figure 2C and Figure S2**). An alternative method to assess the precision of the quantification is to measure the ratio of the light and heavy fragment ions. Quantification of the light/heavy ratios corresponding to stoichiometry profiles between 20 and 80% displayed the highest precision (**Figure 2D**). This is due to the abundance values of the light and heavy fragment ions near 1:1 ratio. In contrast, stoichiometries at the extreme ends of the curve (< 5% and > 95%) displayed the lowest precision since quantitation in these conditions requires the measurement of fragment ions greater than 20-fold difference (**Figure 2D**). In order to calculate stoichiometry, both the heavy and

light fragment ions must be observed, which can be challenging for the lowest stoichiometry if the light fragment ion is below the limit of detection. To avoid biasing the stoichiometry calculations by equally weighting high and low abundance fragment ions, the heavy and light fragment ions are summed together, which allows the higher intensity, higher confidence fragment ions to strongly influence the stoichiometry calculation. The stoichiometry curve analysis quantified stoichiometry for ~1400 acetyl lysine sites. The number of acetyl sites quantified in at least nine conditions was 616. Linear regression analysis using acetyl lysine sites quantified in at least nine conditions shows high reproducibility of this method (**Figure 2E**) with a median R^2 of 0.94, after correction for multiple regression analysis (**Figure 2F**). This global analysis with well-defined input stoichiometries highlights the quantitative nature of this method and is applicable to query acetylation stoichiometry of an entire proteome.

The improved stoichiometry workflow enables the quantitation of different acetyl-lysines from a single peptide, removing the ambiguity of site quantification. As an example, the histone H3 peptide (containing K18 & K23), $K_{Ac}QLATK_{Ac}AAR$, has fragment ions that are unique to each lysine site. K18 is quantified by the fragment ions b2-b3, while y4-y8 are specific for K23 (**Figure 1C**). Obtaining high quality and high coverage of b- and y-ions is essential for quantification of multiple lysines on the same peptide. Therefore, the normalized collision energy (NCE) was optimized for higher energy collisional dissociation (HCD) fragmentation (32). Peptide spectral matches (PSMs) were evaluated (**Figure 3A**) as well as the global b- and y-ion coverage (**Figure 3B**) across a wide range of NCEs (15-50 in 5-unit increments) using a chemically acetylated, trypsin and gluC digested proteome. A low number of PSMs with a c-terminal lysine are observed (blue bar). These peptide matches could arise from trypsin cleavage of unmodified lysines as well as proteins with a c-terminal lysine. However, comparing the frequency of lysine peptides to c-terminal glutamate (red) and arginine (green) peptides demonstrates that the chemical acetylation of the proteome progresses to near completion. To determine the global b- and y- fragment ion coverage, each fragment ion identified for a given PSM was counted and normalized to the peptide length. As y-ions

increase with higher NCE, the proportion of b-ions begin to decline at a similar rate (**Figure 3B**). The NCE 25 was used to balance the frequency of b-ions, y-ions, as well as the number of PSMs (**Figure 3A**).

Subcellular distribution of acetylation stoichiometry

There are few studies measuring (or estimating) acetylation stoichiometry in mammalian systems (21, 22, 33). Thus, a comprehensive analysis of the acetylation stoichiometry distribution across the cell remains uncertain. To address this, we first utilized the current quantitative stoichiometry approach using breast cancer cell line MCF7 and quantified a wide range of stoichiometry (< 1% up to 99%) with high correlation between acetyl lysine fragment ions (red) and peptides (blue) between three biological replicates (**Figure 4A and Table S2**). Quantifying acetylation stoichiometry in MCF7 cells shows the distribution of acetylation skewed towards low stoichiometry (**Figure 4A**). To determine if the distribution of stoichiometry varies across subcellular regions, we next grouped each protein into known subcellular localization based on UniProt localization and compared the acetylation stoichiometry distribution between cytoplasmic and nuclear localizations. Using this grouping and non-parametric analysis, the nuclear fraction contains more acetylation sites with a higher stoichiometry compared to the cytoplasmic fractions ($p = 0.00027$) (**Figure 4B**). To validate these findings, an orthogonal approach to quantify subcellular acetylation levels was utilized. Subcellular fractionation was performed on MCF7 cells by differential centrifugation and acid extraction to enrich for histone, nuclear non-histone, mitochondrial, and cytosolic proteins. Each fraction was treated with a combination of proteases to completely digest proteins of each subcellular compartment to individual amino acids. The relative abundance of acetyl-lysine and unmodified lysine can be measured using mass spectrometry (34) (**Figure 4C**). Acetyl-lysine was significantly more abundant on histone and nuclear proteins compared to the cytoplasm and mitochondrial fractions (**Figure 4D**) corroborating the peptide-level stoichiometry results (**Figure 4B**).

To identify biological processes enriched in acetylation, a biological pathway analysis tool recently developed termed quantitative site set functional score analysis (QSSA) was used to analyze the acetylation

stoichiometry dataset (35). This tool was developed for PTM pathway enrichment analysis taking into account the number of modified sites as well as the fold-change across conditions. QSSA was adapted for acetylation stoichiometry datasets. The stoichiometry data from MCF7 cells was divided into quartiles (for stoichiometry ranges, see Materials and Methods). Each quartile was used as input for the QSSA. Gene Ontology processes that were enriched in this experiment include Metabolic Pathways, Ribosome, Spliceosome, and Protein Processing in Endoplasmic Reticulum (**Figure 4D**). Enrichment of metabolic pathways is a hallmark of acetylation studies (3, 35, 36). Proteins that form part of the ribosome are N-ε- (17, 37) and N-α-acetylated (38). Interestingly, decreases in N-α-acetylation of ribosomal proteins correlate with a decrease in 80S ribosome assembly and cell growth (38) demonstrating a functional link, however, it remains unknown whether N-ε-acetylation can regulate ribosomal function or what time scales are needed to observe changes in ribosomal acetylation.

Acetylation and proteome dynamics

While many studies focus on acetylation changes over longer periods of time, days in cell culture and months in animal models, understanding the dynamics of acetylation over shorter time scales (minutes and hours) can give critical insight into the mechanisms and functionality of acetylation. With time-course information, the relative timing, direction, and magnitude of the acetylation changes can help to distinguish change as primary or secondary responses. Levels of acetylation are dictated through additive mechanisms that involve acetyltransferase activity, nonenzymatic reactions and increased acetyl coenzyme A (acetyl-CoA) levels, while removal mechanisms involve deacetylase activity and protein degradation (8, 36, 39). Therefore, understanding the dynamics of not only protein acetylation, but also proteome dynamics will be important to understand the interplay between these two processes. The majority of studies quantifying acetylation dynamics utilize an antibody based workflow to enrich for the acetyl-peptides (16, 17). Using an enrichment strategy, it is necessary to account for changes in protein abundance in order to accurately report changes in acetylation by analyzing a sample of the proteome which was not subjected to the

immuno-enrichment procedure. The methods described here represent a label-free DIA workflow that does not require an enrichment step. Instead, all free lysine residues are chemically modified using acetic anhydride, a step which is analogous to the alkylation of cysteine residues with iodoacetamide. Therefore, precursor abundance data collected from the acetylation stoichiometry workflow can also be used to estimate protein abundance using label-free quantification techniques.

To initiate a robust stimulation of MCF7 cells, 24 hr serum-starved cells were activated with serum and harvested at 0, 1, 2, 4 hours (**Figure 5A**). Acetylation stoichiometry and protein abundance were determined as in Figure 1A (**Table S3**). This design has the benefits of synchronizing cells upon serum starvation followed by robust changes in signaling pathways that occur upon serum replenishment (40–42). To verify activation of major signaling pathways, we monitored the level of phosphorylation of ribosomal protein S6 (**Figure 5B, C**), as a proxy for mTOR signaling activation (40).

Acetylation stoichiometry was quantified using the described DIA-MS approach followed by a pattern recognition analysis using fuzzy c-means clustering (43). Clustering analysis identified four unique clusters where site-level acetylation dynamics revealed distinct profiles (**Figure 5D**). Over two-thirds of the acetylation sites identified in this clustering analysis were found in clusters 1 and 3 which display rapid changes upon growth factor stimulation. These clusters correspond to acetylation levels that rapidly increased and returned to pretreatment baseline levels (Cluster 1) as well as trends where acetylation rapidly decreased and remained low (Cluster 3), respectively. Protein abundance was determined by label-free quantification using MSstats in conjunction with the chemically acetylated proteome sample (44) followed by the clustering analysis using fuzzy c-means (**Figure 5E**) (43). QSSA analysis of the serum-stimulated acetylation stoichiometry dataset identified biological processes that are enriched in each of the acetylation clusters **(Figure 5F)**. Acetylation clusters 1 and 2, which exhibited rapid increases in acetylation stoichiometry within the first two hours were highly enriched for Metabolic Pathways. Cluster 4, which contained acetylation sites that more slowly increased over time, was highly enriched for the Ribosome.

To investigate possible links between protein and acetylation dynamics, the trends of acetylation were compared within the protein clusters (**Figure 5G**). For example, there is a strong overlap between acetylation stoichiometry cluster 3 within the protein clusters 2 and 4. This demonstrates that there is a subset of sites that have decreased acetylation stoichiometry with increasing protein abundance. This particular overlap of dynamics could be due to an increase in protein abundance without increased lysine acetylation, which would result in a trend showing a decrease in acetylation stoichiometry. Whereas, increases in lysine acetylation when protein abundance also increases must occur through active acetylation.

Coordinated acetylation dynamics in diverse human cell lines

The results from MCF7 cells suggest protein acetylation is dynamically controlled in a model of growth factor stimulation, and that major metabolic and cellular pathways are targets of these rapid acetylation changes. To further understand acetylation dynamics and determine if the serum-stimulated changes are conserved across cell types, the serum starve-replete cell culture model was applied to a colon cancer cell line, HCT116 (**Figure 5A**). Acetylation stoichiometry was quantified using the DIA-MS approach at 0 hours (no serum), 15 minutes, 1, 2, and 4 hours post-serum re-addition (**Table S4**). The total number of quantified acetylation sites in the HCT116 experiment was 3818, which is comparable to the 4310 acetylation sites quantified in the MCF7 experiment. Interestingly, there was a strong overlap of quantified acetylation sites across the two experiments, 3143 sites had quantifiable stoichiometry in both experiments resulting in a ~63% overlap (**Figure 6A**). To identify and understand which acetylation sites are dynamic in both experiments, we identified which acetylation sites are significantly changing over time using one-way analysis of variance (ANOVA). There were 647 and 1037 acetylation sites in MCF7 and HCT116 cells, respectively, significantly changing over time (p-value ≤ 0.05) with 228 sites significantly changing in both experiments (**Figure 6B**). Observing dynamic acetylation on the same lysine site in diverse cell lines suggests functional importance. Examples of overlapping acetylation sites are Mitochondrial fission 1 protein (FIS1) K89 and Sulfiredoxin (SRXN1) K116. Both of these acetylation

sites experienced rapid increases in acetylation in response to serum stimulation, a 30-40% increase at FIS1 K89 and a 25-60% increase at SRXN1 K116, which was maintained at a high level throughout the time course.

In addition to identifying the proteins and sites significantly changing in both experiments, determining the pathways similarly regulated between cell lines was an important goal. Therefore, we used DAVID Bioinformatics Functional Annotation Tool to identify GO term enrichment in the overlapping dynamic acetylation sites (45, 46). Grouping similar GO Terms together, we found that terms associated with cell junction, translation, protein folding, and splicing were enriched (**Figure 6C**). Furthermore, we took the most robustly changing sites in the overlap, ≥ 5% change up or down relative to the pre-stimulation time point and performed network analysis using STRING (**Figure 6D**) (47). K-means clustering of the resulting network demonstrated three main clusters of proteins that correlated with the pathway enrichment analysis. The clusters generally encompassed splicing and RNA-binding proteins, translational machinery, and proteins involved in maintaining protein homeostasis. Notably, within the overlapping sites we only observed a few acetyl sites on mitochondrial matrix localized proteins, of which many had only small (≤5%) changes over the time course.

Lastly, we used the same fuzzy c-means clustering to identify acetylation dynamic patterns in the serum-stimulated HCT116 cells (**Figure 6E**). Interestingly, similar acetylation trends in HCT116 cells were also observed in the MCF7 cells (**Figure 5D**). We identified several clusters with rapid changes in acetylation (clusters 1, 3, and 4) and one cluster with a delayed acetylation response (cluster 2). When we overlaid the pathway enrichment analysis onto the cluster analysis, we found that RNA metabolic processes, such as splicing and translation, had more sites that increased in acetylation over time (**Figure 5D – clusters 1 and 4, and Figure 6E – clusters 1, 2, and 3**). Cell-cell junction annotated proteins demonstrated more sites whose acetylation decreased over time (**Figure 5D – cluster 3 and Figure 6E – cluster 4**).

Discussion:

Integrated stoichiometry workflow

In this study, we have developed and utilized an improved DIA mass spectrometry method to quantify acetylation stoichiometry and investigate rapid acetylation dynamics in cells in response to serum stimulation. Understanding the speed and magnitude of protein acetylation changes on shorter timescales (minutes and hours) provide new insights into the cellular mechanisms of acetylation and help to prioritize specific sites for functional investigation. We previously developed a DDA analysis method to quantify acetylation stoichiometry in an entire proteome (20), however, identification and quantification accuracy decreased as the sample complexity increased, partly due to incorrect assignment of light and heavy precursors pairs for stoichiometry calculation. Therefore, a new workflow was needed which would identify and quantify acetylation with high accuracy and sensitivity. Here, we integrated the acetylation stoichiometry workflow with targeted DIA analysis, which allows identification and quantification of light and heavy acetyl-lysine fragment ions, enabled by a novel spectral library that contains all light and heavy acetyl-lysine feature pairs. Using this method, we demonstrate accurate and reproducible analysis of dynamic protein acetylation in different cell lines and highlight similar activated pathways upon serum stimulation.

Recently, Choudhary and colleagues developed an orthogonal method to quantify acetylation stoichiometry (30, 33, 48), which incorporates a heavy SILAC labeled proteome subjected to partial chemical acetylation, combined with an experimental sample by serial dilution, trypsin digested, followed by acetyl-lysine enrichment and MS analysis. Stoichiometry quantification is assessed by comparing the SILAC ratio across the dilution series. Validation of this approach was performed using AQUA peptides, which agreed with their stoichiometry calculation (30), when stoichiometries were less than 10%. Additionally, the authors demonstrate higher quantification accuracy for peptides with very low stoichiometry (< 1%) and higher error rates with higher stoichiometry (> 1%). Therefore, the two strategies

developed for quantifying acetylation stoichiometry by Choudhary *et.al.* and this study, represent orthogonal methods for accurately quantifying low and high acetylation stoichiometry, respectively.

Acetylation dynamics and mechanistic implications

In this study, rapid acetylation dynamics were investigated using serum stimulation to initiate a major transcriptional, translational, and metabolic response (49). Analysis in two different human cell lines revealed that acetylation changes occur on time-scales (minutes) that rival those observed via phosphorylation-dependent signaling (50, 51). Rapid changes in acetylation included groups of sites that increased and groups of sites that decreased, many with similar kinetic profiles. Notably, many sites showed no significant change across the four-hour time-course. The majority of acetylation sites that decreased rapidly were from proteins that increased in abundance, as would be predicted from newly translated, unacetylated proteins under serum stimulation (52). Most importantly, lysine sites that exhibited rapid increases in acetylation are candidates for acetyltransferase control and functional regulation. The results presented here, enabled by a robust MS method, provide a critical resource for investigators studying the regulation of these pathways.

Acetylation stoichiometry distribution in the cell

One of the most interesting observations from this study is that the highest stoichiometry and dynamics, particularly increases in acetylation, occur on proteins that exist in subcellular compartments where acetyltransferases are known to reside. In continually serum-fed MCF7 cells, we observed a significantly larger proportion of proteins with higher acetylation stoichiometry within the nuclear compartment (**Figure 4C**), which was corroborated by a separate method (**Figure 4D**). Additionally, nuclear-localized proteins displayed the most dramatic changes upon serum stimulation. Nuclear localized proteins, such as transcriptional and post-transcriptional processing factors, exhibited rapid increases in acetylation in both MCF7 and HCT116 cells. For example, we observed increases in acetylation

stoichiometry ranging from 8-30% over the first hour on several splicing factors such as U4/U6.U5 tri-snRNP associated protein 1 (SART1) at K147, RNA binding-protein 8A (RBM8A) at K114, and heterogeneous nuclear ribonucleoprotein H3 (HNRNPH3) at K97. Given that these proteins display increased acetylation under pro-growth conditions and are localized to the nucleus where acetyltransferases reside, we predict that these sites are enzymatically regulated and functionally important for increasing basal splicing rates or for preferentially affecting certain types of splicing reactions. Our data provide a novel roadmap for investigating nuclear acetylation of non-histone proteins and suggest that acetyltransferases play a critical function well beyond histone acetylation.

Outside the nucleus, we observe dynamic acetylation across other cellular pathways known to be regulated during serum stimulation. We observed high enrichment for proteins involved in translation, including increases in acetylation on ribosomal proteins such as RPL3 at K103 and K155 and RPL7 at K77 (**Figures 6D and 6E**). Pro-growth increases in protein synthesis is a major cellular response to serum stimulation (52). A known mark of actively translating ribosomes is Ribosomal protein S6 phosphorylation. The acetylation of S6 mimicked the pattern of the MCF7 cluster showing increased but slower acetylation kinetics (**Figures 5C and 5D**) (53). Interestingly, this same cluster of acetylation sites was enriched for the Ribosome. Taken together with the contrasting observation that most newly translated proteins are unacetylated, increased acetylation of ribosomal proteins is positively associated with translation and would be predicted to be catalyzed by acetyltransferases (54).

Another cellular pathway known to be affected by serum stimulation is mitochondrial fission and fusion, in which mitochondria transition between being highly fragmented in the quiescent state to more tubular structures as cells approach the G1-S transition (55). On one protein involved in this process, mitochondrial fission protein 1 (FIS1) we observed large, rapid increases in acetylation at K89. FIS1 is localized to the cytoplasmic face of the mitochondrial outer membrane, therefore this site is likely to be enzymatically regulated by cytoplasmic acetyltransferases. Acetylation of FIS1 K89 might inhibit mitochondrial fission, potentially through disrupting interactions with other fission machinery proteins such

as Dynamin-1-like protein (DNM1L) (56). Lastly, we observed dynamic protein acetylation on proteins related to protein homeostasis, particularly ER-localized proteins. It was recently shown that the lysine acetyltransferases, NAT8/NAT8B, are localized in the lumen of the endoplasmic reticulum and function to acetylate properly folded proteins as proteins traverse through the secretory pathway (57). Acetylation by NAT8/NAT8B is proposed to signal correctly folded proteins and function in quality control. Collectively, the strong trend in these results suggest that dynamic protein acetylation occurs in subcellular compartments with known, localized acetyltransferases. The existence of bona fide protein acetyltransferases in the mitochondrial matrix is the subject of debate (9–11). While acetylation is prevalent in mitochondria(5, 35, 36, 58, 59), the considerably lower mean stoichiometry and no apparent kinetic trends in response to growth factor stimulation suggest that widespread enzyme-catalyze acetylation is not a major mechanism for matrix proteins. The method and results described in this study provide a valuable resource and tool to investigate regulatory acetylation events and as a rich acetylation data set for investigators to now directly test the role of acetylation in a pathway of interest.

Author contributions: J.B., A.L., and J.M.D. designed the research study. J.B. and A.L. performed acetylation stoichiometry mass spectrometry experiments and analyzed the data. J.F. performed single amino acid mass spectrometry experiments and analyzed data. I.L., T.G., O.M.B., and L.R. developed software to analyze DIA MS data as well as for custom modification of the library. J.B., A.L. and M.J.S. performed bioinformatic analysis. J.B. and A.L. designed the figures and drafted the manuscript. J.B., A.L., J.F., M.J.S., L.R., and J.M.D edited manuscript. J.M.D. is the corresponding author.

Acknowledgments: We would like to thank Greg Barrett-Wilt and Greg Sabat at the University of Wisconsin-Madison Biotechnology center for use of the Mascot database server. This work was supported, in whole or in part, by National Institutes of Health (NIH) Grant GM065386 (J.M.D.), NIH National

Research Service Award T32 GM007215 (J.B. and A.L.) and the National Science Foundation Graduate Research Fellowship Program (NSF-GRFP) DGE-1256259 (J.B.).

Additional Information

The authors I.L., T.G., O.M.B., and L.R. are employees of Biognosys AG (Zurich, Switzerland). Spectronaut is a trademark of Biognosys AG.

Data Availability

The mass spectrometry raw files, spectral libraries, and the result files from MaxQuant and Spectronaut used in this study have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with the dataset identifier PXD014453.

The R code used to calculate acetylation stoichiometry from the Spectronaut result files and perform all secondary analyses have been published to GitHub through Zenodo and can be accessed using the following DOI: 10.5281/zenodo.3360892.

References

1. Kouzarides, T. (2000) Acetylation: a regulatory modification to rival phosphorylation? *EMBO J.* 19, 1176–1179

2. Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964) ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc. Natl. Acad. Sci. U. S. A.* 51, 786–794

3. Verdin, E., and Ott, M. (2015) 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat. Rev. Mol. Cell Biol.* 16, 258–264

4. Close, P., Creppe, C., Gillard, M., Ladang, A., Chapelle, J.-P., Nguyen, L., and Chariot, A. (2010) The emerging role of lysine acetylation of non-nuclear proteins. *Cell. Mol. Life Sci.* 67, 1255–1264

5. Anderson, K. A., and Hirschey, M. D. (2012) Mitochondrial protein acetylation regulates metabolism. *Essays Biochem.* 52, 23–35

6. Pehar, M., and Puglielli, L. (2013) Lysine acetylation in the lumen of the ER: a novel and essential function under the control of the UPR. *Biochim. Biophys. Acta* 1833, 686–697

7. Pougovkina, O., te Brinke, H., Ofman, R., van Cruchten, A. G., Kulik, W., Wanders, R. J. A., Houten, S. M., and de Boer, V. C. J. (2014) Mitochondrial protein acetylation is driven by acetyl-CoA from fatty acid oxidation. *Hum. Mol. Genet.* 23, 3513–3522

8. Baeza, J., Smallegan, M. J., and Denu, J. M. (2016) Mechanisms and Dynamics of Protein Acetylation in Mitochondria. *Trends Biochem. Sci.* 41, 231–244

9. Olia, A. S., Barker, K., McCullough, C. E., Tang, H.-Y., Speicher, D. W., Qiu, J., LaBaer, J., and Marmorstein, R. (2015) Nonenzymatic Protein Acetylation Detected by NAPPA Protein Arrays.

ACS Chem. Biol. 10, 2034–2047

10. Wagner, G. R., and Payne, R. M. (2013) Widespread and enzyme-independent Nε-acetylation and Nε-succinylation of proteins in the chemical conditions of the mitochondrial matrix. *J. Biol. Chem.* 288, 29036–29045

11. James, A. M., Hoogewijs, K., Logan, A., Hall, A. R., Ding, S., Fearnley, I. M., and Murphy, M. P. (2017) Non-enzymatic N-acetylation of Lysine Residues by AcetylCoA Often Occurs via a Proximal S-acetylated Thiol Intermediate Sensitive to Glyoxalase II. *Cell Rep.* 18, 2105–2112

12. Baeza, J., Smallegan, M. J., and Denu, J. M. (2015) Site-specific reactivity of nonenzymatic lysine acetylation. *ACS Chem. Biol.* 10, 122–128

13. Tanner, K. G., Langer, M. R., and Denu, J. M. (2000) Kinetic mechanism of human histone acetyltransferase P/CAF. *Biochemistry* 39, 11961–11969

14. Berndsen, C. E., Albaugh, B. N., Tan, S., and Denu, J. M. (2007) Catalytic mechanism of a MYST family histone acetyltransferase. *Biochemistry* 46, 623–629

15. Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 43, D512–20

16. 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., and Zhao, Y. (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell* 23, 607–618

17. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840

18. Filippakopoulos, P., and Knapp, S. (2014) Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discov.* 13, 337–356

19. Josling, G. A., Selvarajah, S. A., Petter, M., and Duffy, M. F. (2012) The role of bromodomain proteins in regulating gene expression. *Genes* 3, 320–343

20. Baeza, J., Dowell, J. A., Smallegan, M. J., Fan, J., Amador-Noguez, D., Khan, Z., and Denu, J. M. (2014) Stoichiometry of site-specific lysine acetylation in an entire proteome. *J. Biol. Chem.* 289, 21326–21338

21. Nakayasu, E. S., Wu, S., Sydor, M. A., Shukla, A. K., Weitz, K. K., Moore, R. J., Hixson, K. K., Kim, J.-S., Petyuk, V. A., Monroe, M. E., Pasa-Tolic, L., Qian, W.-J., Smith, R. D., Adkins, J. N., and Ansong, C. (2014) A method to determine lysine acetylation stoichiometries. *Int. J. Proteomics* 2014, 730725

22. Zhou, T., Chung, Y.-H., Chen, J., and Chen, Y. (2016) Site-Specific Identification of Lysine Acetylation Stoichiometries in Mammalian Cells. *J. Proteome Res.* 15, 1103–1113

23. Meyer, J. G., D'Souza, A. K., Sorensen, D. J., Rardin, M. J., Wolfe, A. J., Gibson, B. W., and Schilling, B. (2016) Quantification of Lysine Acetylation and Succinylation Stoichiometry in Proteins Using Mass Spectrometric Data-Independent Acquisitions (SWATH). *J. Am. Soc. Mass Spectrom.* 27, 1758–1771

24. Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., and Reiter, L. (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* 14, 1400–1410

25. Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M. J., and Rinner, O. (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* 12, 1111–1121

26. Bruderer, R., Bernhardt, O., Gandhi, T., and Reiter, L. (2016) High precision iRT retention time prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. *Proteomics*,

27. Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., and Reiter, L. (2017) Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell. Proteomics* 16, 2296–2309

28. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11, O111.016717

29. Kelstrup, C. D., Bekker-Jensen, D. B., Arrey, T. N., Hogrebe, A., Harder, A., and Olsen, J. V. (2018) Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J. Proteome Res.* 17, 727–738

30. Hansen, B. K., Gupta, R., Baldus, L., Lyon, D., Narita, T., Lammers, M., Choudhary, C., and Weinert, B. T. (2019) Analysis of human acetylation stoichiometry defines mechanistic constraints on protein regulation. *Nat. Commun.* 10, 1055

31. Selevsek, N., Chang, C.-Y., Gillet, L. C., Navarro, P., Bernhardt, O. M., Reiter, L., Cheng, L.-Y., Vitek, O., and Aebersold, R. (2015) Reproducible and consistent quantification of the Saccharomyces cerevisiae proteome by SWATH-mass spectrometry. *Mol. Cell. Proteomics* 14, 739–749

32. Greer, T., Lietz, C. B., Xiang, F., and Li, L. (2015) Novel isotopic N,N-dimethyl leucine (iDiLeu) reagents enable absolute quantification of peptides and proteins using a standard curve approach. *J. Am. Soc. Mass Spectrom.* 26, 107–119

33. Weinert, B. T., Moustafa, T., Iesmantavicius, V., Zechner, R., and Choudhary, C. (2015) Analysis of acetylation stoichiometry suggests that SIRT3 repairs nonenzymatic acetylation lesions. *EMBO J.* 34, 2620–2632

34. Fan, J., Baeza, J., and Denu, J. M. (2016) Investigating Histone Acetylation Stoichiometry and Turnover Rate. *Methods Enzymol.* 574, 125–148

35. Dittenhafer-Reed, K. E., Richards, A. L., Fan, J., Smallegan, M. J., Fotuhi Siahpirani, A., Kemmerer, Z. A., Prolla, T. A., Roy, S., Coon, J. J., and Denu, J. M. (2015) SIRT3 mediates multi-tissue coupling for metabolic fuel switching. *Cell Metab.* 21, 637–646

36. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., and Mann, M. (2014) The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat. Rev. Mol. Cell Biol.* 15, 536–550

37. Liew, C. C., and Gornall, A. G. (1973) Acetylation of ribosomal proteins. I. Characterization and properties of rat liver ribosomal proteins. *J. Biol. Chem.* 248, 977–983

38. Kamita, M., Kimura, Y., Ino, Y., Kamp, R. M., Polevoda, B., Sherman, F., and Hirano, H. (2011) N(α)-Acetylation of yeast ribosomal proteins and its effect on protein synthesis. *J. Proteomics* 74, 431–441

39. Shi, L., and Tu, B. P. (2015) Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. *Curr. Opin. Cell Biol.* 33, 125–131

40. Magnuson, B., Ekim, B., and Fingar, D. C. (2012) Regulation and function of ribosomal protein S6 kinase (S6K) within mTOR signalling networks. *Biochem. J* 441, 1–21

41. Rosner, M., Schipany, K., and Hengstschläger, M. (2013) Merging high-quality biochemical fractionation with a refined flow cytometry approach to monitor nucleocytoplasmic protein expression throughout the unperturbed mammalian cell cycle. *Nat. Protoc.* 8, 602–626

42. Xu, Y. S., Liang, J. J., Wang, Y., Zhao, X.-Z. J., Xu, L., Xu, Y.-Y., Zou, Q. C., Zhang, J. M., Tu, C.-E., Cui, Y.-G., Sun, W.-H., Huang, C., Yang, J.-H., and Chin, Y. E. (2016) STAT3 Undergoes Acetylation-dependent Mitochondrial Translocation to Regulate Pyruvate Metabolism. *Sci. Rep.* 6, 39517

43. Schwämmle, V., and Jensen, O. N. (2010) A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinformatics* 26, 2841–2848

44. Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014) MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 30, 2524–2526

45. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13

46. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57

47. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–52

48. Weinert, B. T., Iesmantavicius, V., Moustafa, T., Schölz, C., Wagner, S. A., Magnes, C., Zechner, R., and Choudhary, C. (2014) Acetylation dynamics and stoichiometry in Saccharomyces cerevisiae.

Mol. Syst. Biol. 11, 833

49. Kirkconnell, K. S., Paulsen, M. T., Magnuson, B., Bedi, K., and Ljungman, M. (2016) Capturing the dynamic nascent transcriptome during acute cellular responses: The serum response. *Biol. Open* 5, 837–847

50. Kholodenko, B. N., Demin, O. V., Moehren, G., and Hoek, J. B. (1999) Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.* 274, 30169–30181

51. Kholodenko, B. N. (2006) Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* 7, 165–176

52. Montine, K. S., and Henshaw, E. C. (1989) Serum growth factors cause rapid stimulation of protein synthesis and dephosphorylation of eIF-2 in serum deprived Ehrlich cells. *Biochim. Biophys. Acta* 1014, 282–288

53. Roux, P. P., Shahbazian, D., Vu, H., Holz, M. K., Cohen, M. S., Taunton, J., Sonenberg, N., and Blenis, J. (2007) RAS/ERK signaling promotes site-specific ribosomal protein S6 phosphorylation via RSK and stimulates cap-dependent translation. *J. Biol. Chem.* 282, 14056–14064

54. Han, Z., Chou, C.-W., Yang, X., Bartlett, M. G., and Zheng, Y. G. (2017) Profiling Cellular Substrates of Lysine Acetyltransferases GCN5 and p300 with Orthogonal Labeling and Click Chemistry. *ACS Chem. Biol.* 12, 1547–1555

55. Mitra, K., Wunder, C., Roysam, B., Lin, G., and Lippincott-Schwartz, J. (2009) A hyperfused mitochondrial state achieved at G1-S regulates cyclin E buildup and entry into S phase. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11960–11965

56. Yu, T., Fox, R. J., Burwell, L. S., and Yoon, Y. (2005) Regulation of mitochondrial fission and apoptosis by the mitochondrial outer membrane protein hFis1. *J. Cell Sci.* 118, 4141–4151

57. Peng, Y., and Puglielli, L. (2016) Nϵ-lysine acetylation in the lumen of the endoplasmic reticulum: A way to regulate autophagy and maintain protein homeostasis in the secretory pathway. *Autophagy* 12, 1051–1052

58. 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., and Coon, J. J. (01/2013) Calorie Restriction and SIRT3 Trigger Global Reprogramming of the Mitochondrial Protein Acetylome. *Mol. Cell* 49, 186–199

59. 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., and 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.* 110, 6601–6606

60. Dowell, J. A., Frost, D. C., Zhang, J., and Li, L. (2008) Comparison of two-dimensional fractionation techniques for shotgun proteomics. *Anal. Chem.* 80, 6715–6723

61. Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., Pagès, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L., and Morgan, M. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121

62. 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., and Mootha, V. K. (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134, 112–123

63. Calvo, S. E., Clauser, K. R., and Mootha, V. K. (2016) MitoCarta2.0: an updated inventory of

mammalian mitochondrial proteins. *Nucleic Acids Res.* 44, D1251–7

64. Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G. D. (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 5, e13984

Figures

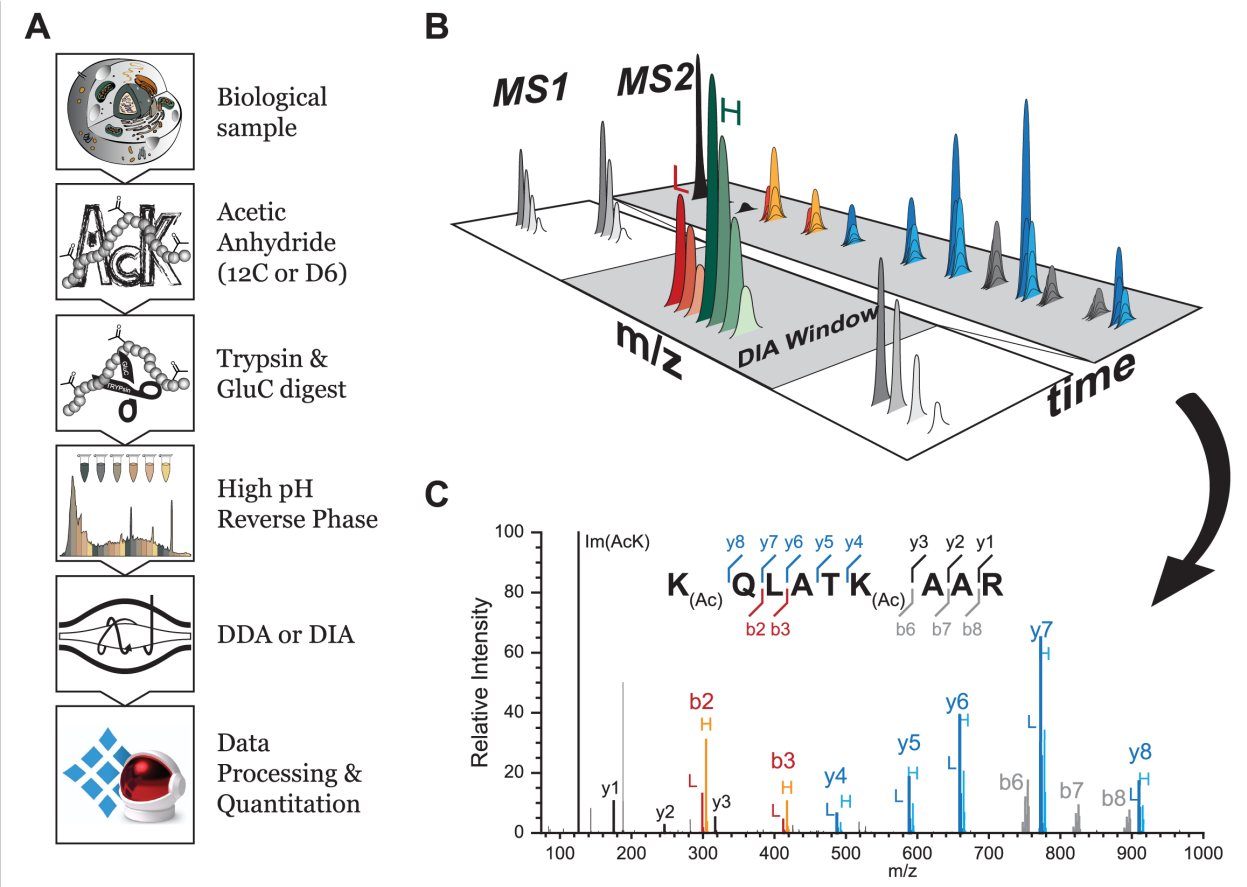

Figure 1: DIA acetylation stoichiometry workflow

(A) Workflow for measuring acetylation stoichiometry from whole cell lysate. (B) Diagram representing DIA quantitation of precursor and fragment ions across the elution profile. Peak areas for precursor and fragment ions are quantified. (C) Diagram illustrating MS2 spectra for the Histone H3 peptide containing lysine K18Ac and K23Ac. The fragments b2-b3 are specific for K18 and y4-y8 are specific for K23. The fragments b6-b8 are ambiguous as they contain K18Ac and K23Ac, while the fragments y1-y3 contain no acetyl lysine information.

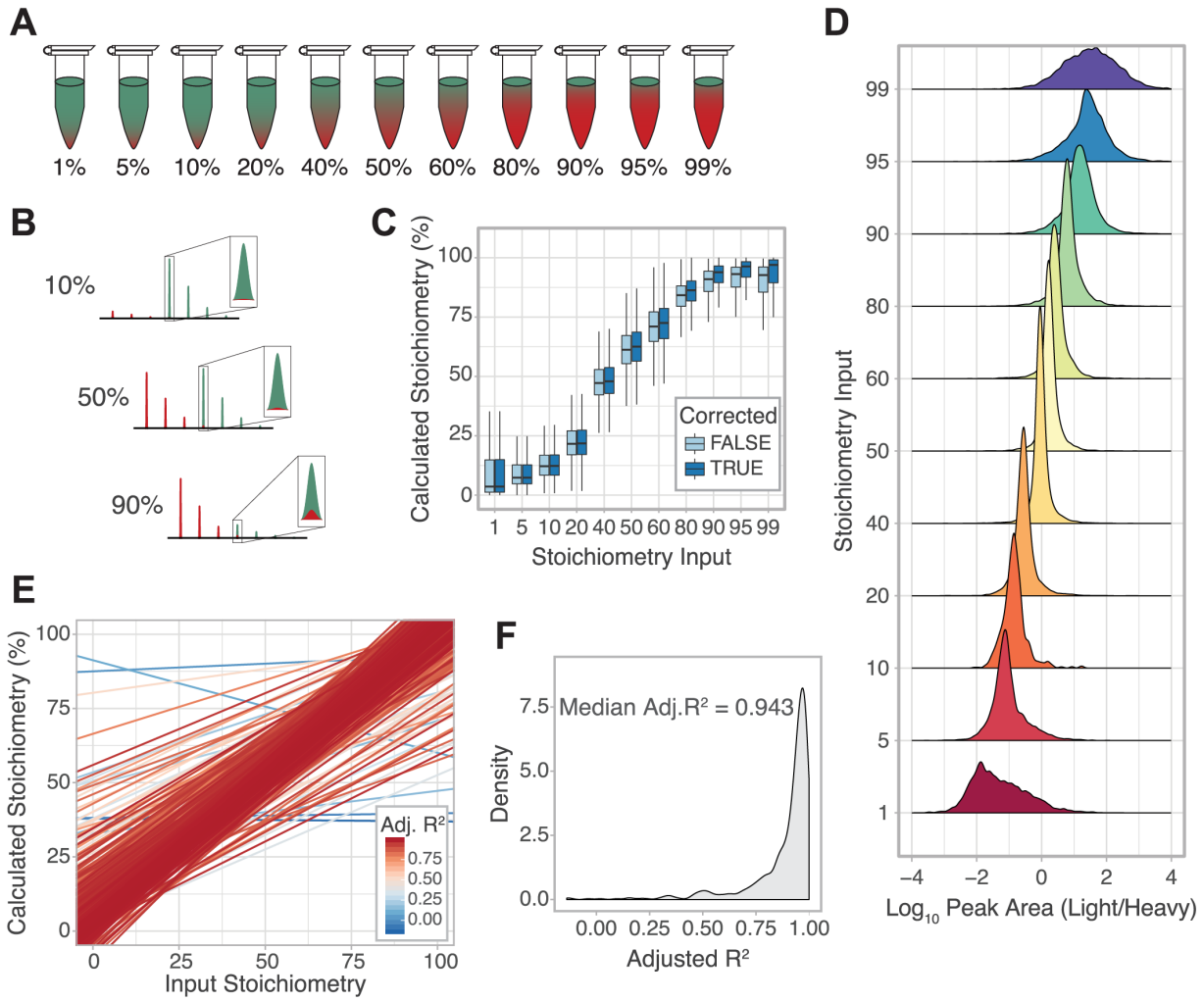

Figure 2: Benchmarking the DIA stoichiometry workflow

(A) Diagram of standard curve samples generated by combining fully heavy labeled (green) peptides with fully light labeled (red) peptides at the specified ratios. (B) Correction of natural abundance isotopic envelope between light and heavy acetyl peptides. Red peaks represent light acetyl containing fragment ions and the green peaks represent heavy acetyl fragment ions. Isotopic overlap is much more pronounced at high stoichiometry. (C) Boxplot of the proteome-wide stoichiometry distribution before and after natural abundance isotope correction. (D) Density plot of the Log_{10}-transformed (light/heavy) peak area for each input stoichiometry condition. (E) Linear regression analysis of all acetyl lysine sites (> 9 data points, n =

616) with the color of each line corresponding to the R^2 value. (F) Distribution of Adjusted R^2 values for the acetyl lysine linear regression analyses.

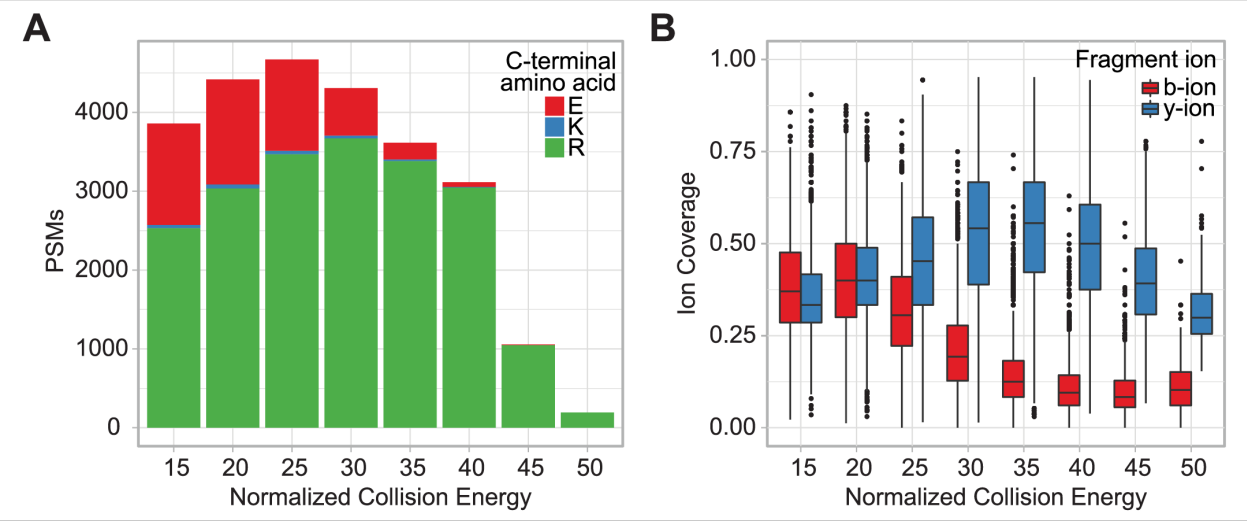

Figure 3: Optimization of instrument parameters for increased fragment ion coverage

(A) Peptide spectral matches (PSMs) at varying normalized collision energy (NCE) for trypsin (c-terminal arginine and lysine) and gluC (c-terminal glutamate) digested peptides. The data consists of an MCF7 whole cell lysate labeled with ^{12}C-acetic anhydride, digested with trypsin and gluC and analyzed in DDA mode using varying NCE settings. (B) Global fragment ion coverage for b- and y-ions of the data in A.

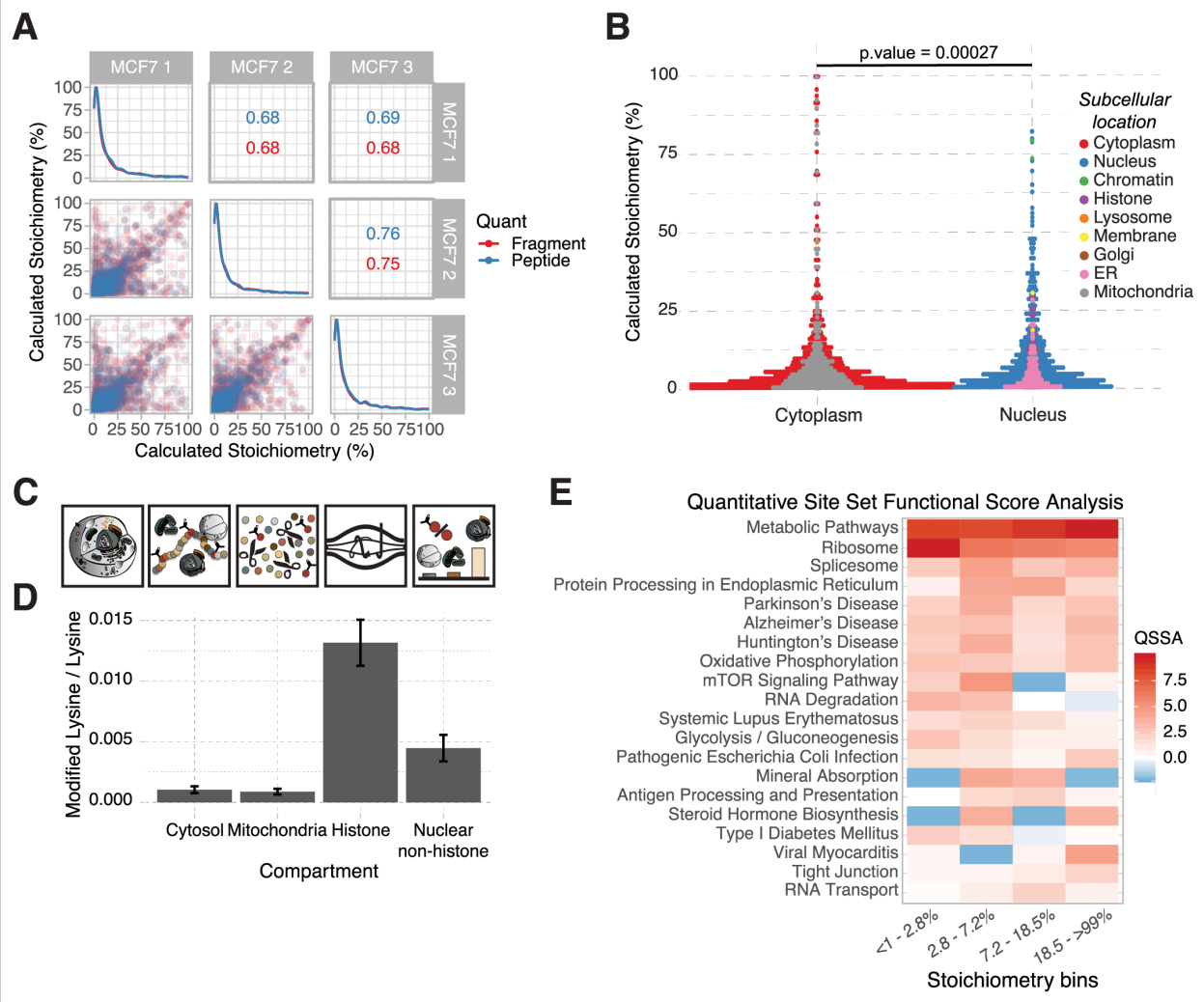

Figure 4: Nuclear-localized proteins have the highest lysine PTM abundances

(A) Scatterplot matrix of acetylation stoichiometry across the three biological replicates of MCF7 cells. Spearman correlation is displayed for each pairwise comparison of fragment ions (red) acetyl-lysine site stoichiometry (blue). (B) QSSA heatmap showing enriched KEGG pathways. Acetylation stoichiometry was binned into quartiles and used as input for the QSSA. (C) Acetylation stoichiometry distribution across cellular compartments. Compartments were grouped *in silico* as cytoplasmic or nuclear. Statistical analysis was performed using the posthoc Kruskal-Nemenyi test. Colored circles represent individual subcellular location assignment based on UniProt location and mitocarta database (62, 63). (D) Workflow for quantifying single amino acid abundances of modified and unmodified lysines of enriched subcellular

fractions. Each fraction is digested with a cocktail of proteases to generate single amino acids. Amino acids are analyzed by MS to quantify modified lysine / unmodified lysine as a measure of global subcellular stoichiometry. (E) Acetyl-lysine:lysine ratio across cytoplasmic, mitochondrial, histone, and non-histone proteins.

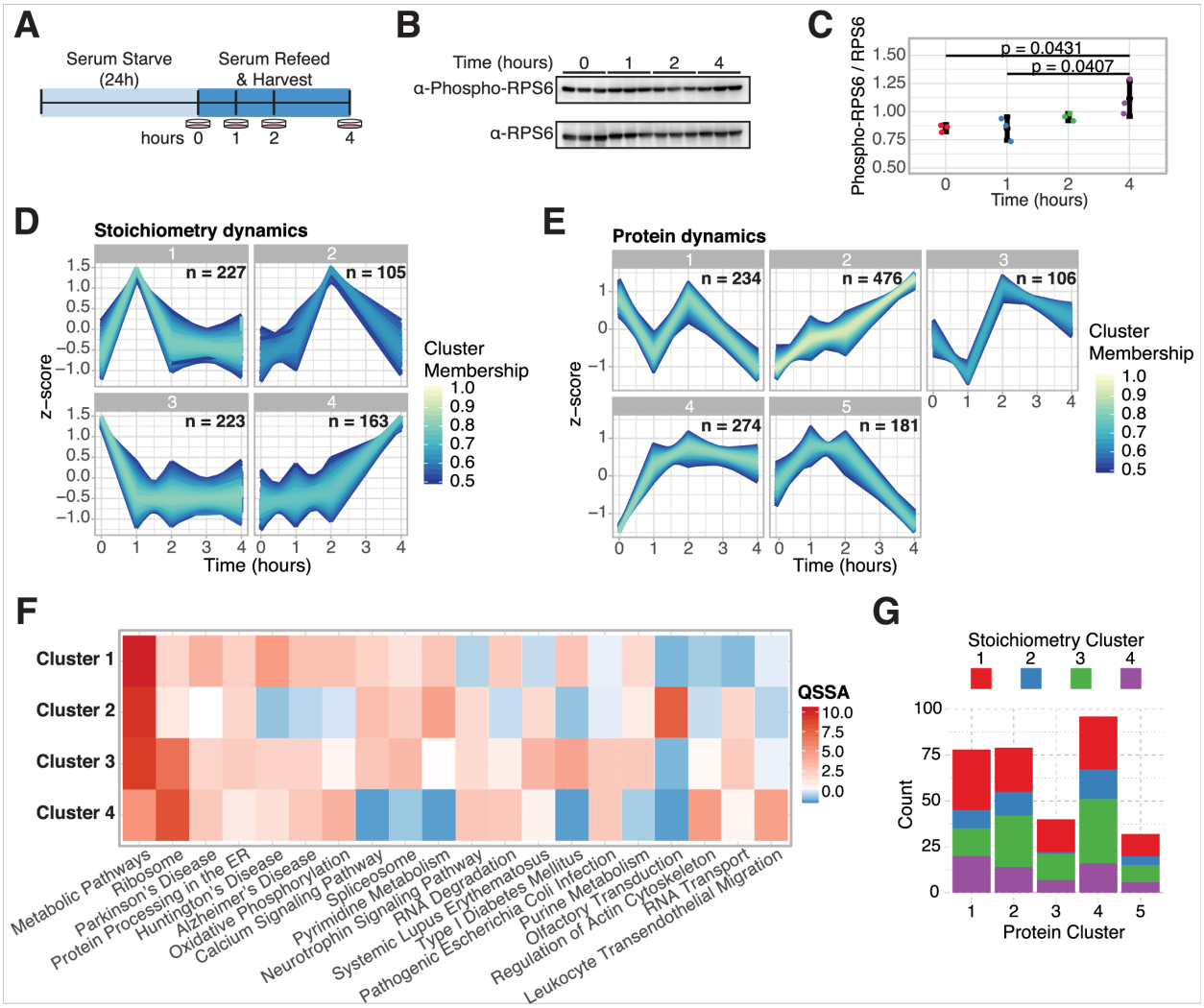

Figure 5: Acetylation and protein dynamics

(A) Diagram of experimental approach to quantify acetylation stoichiometry dynamics. (B) Western blot of phosphorylated Ribosomal Protein S6 (top) and Ribosomal Protein S6 (bottom). (C) Quantitation of the western blot. Statistical analysis was performed using an ANOVA. (D) Time-course clusters of acetylation stoichiometry dynamics by fuzzy c-means clustering. (E) Time-course clusters of protein abundance dynamics by fuzzy c-means clustering. (F) QSSA heatmap showing enriched KEGG pathways. Acetylation stoichiometry was separated by clusters and used as input for the QSSA. (G) Barplot quantifying the frequency of acetylation dynamic clusters within each protein dynamics cluster.

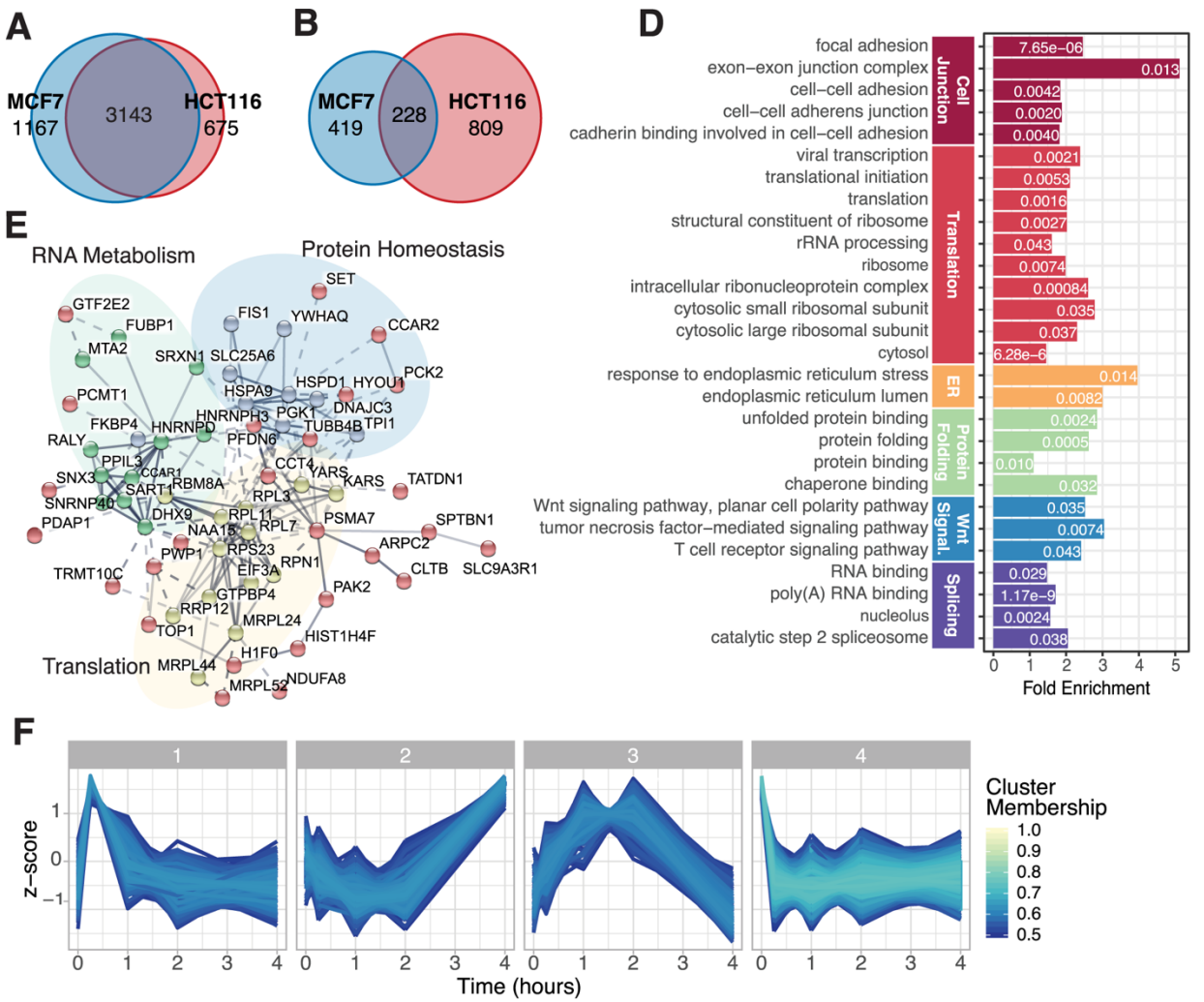

Figure 6: Coordinated acetylation dynamics in diverse cell lines, MCF7 and HCT116

(A) Venn diagram of quantified acetylation sites in serum-stimulated MCF7 and HCT116 cells. (B) Venn diagram of significantly changing acetylation sites (significance calculated using one-way ANOVA analyses) between MCF7 and HCT116 cells. (C) DAVID GO Term enrichment analysis of the proteins significantly changing in both MCF7 and HCT116 cells. Bar graphs represent fold enrichment with p-values in white text, and terms are grouped by similarity. (E) STRING network analysis of robustly (≥ 5%) and significantly changing sites in both experiments. Color of nodes represent cluster from k-means clustering in STRING and edges represent the confidence of the interaction. (F) Time-course clusters of acetylation stoichiometry dynamics by fuzzy c-means clustering.