Cell PRESS
Open ACCESS

EXTENDED EXPERIMENTAL PROCEDURES

Rat Tissues and Peptide Extraction

The study was carried out according to approved national regulations in Denmark and with animal experimental license granted by the Animal Experiments Inspectorate, Ministry of Justice, Denmark. Five Sprague Dawley rats (Crl:SD, male, 350 g, Charles River, Germany) were anesthetized with isoflurane and perfused (1.5 min, 30 ml/min) with isotonic saline containing protease inhibitors (0.120 mM EDTA, 14 μM aprotinin, 0.3 nM valine-pyrrolidide and, Roche Complete Protease Inhibitor tablets (Roche), pH = 7.4) before being decapitated. Organs were quickly dissected and snap frozen in 2-methylbutane. Dissected organs were: brain, pancreas, brown adipose tissue, testis fat, testis, liver, kidney, spleen, lung, stomach, intestine, thymus, heart, muscle and skin. Organs were heat inactivated (Denator T1 Heat Stabilizor, Denator, Gothenburg, Sweden) followed by tissue homogenization and micro-tip sonication on ice. Proteins were acetone precipitated and resuspended in urea solution (6 M urea/2 M thiourea, 10 mM HEPES, pH 8.0) and the concentration was measured (Quick Start Bradford Dye Reagent X1, Bio-Rad). 50 mg protein per tissue was reduced (final concentration 1 mM dithiothreitol, 750 rpm, 30 min), and alkylated (final concentration 5.5 mM chloroacetamide, 750 rpm, 20 min in darkness) before protein digestion with 250 μg endoproteinase Lys-C (Wako) (750 rpm, 3 hr). The samples were diluted with 25 mM ammonium bi-carbonate to lower the urea concentration below 2M, and then further digested with 250 μg modified trypsin (Sequencing grade, Promega) (750 rpm, 8 hr). Trypsin digestion was quenched by lowering pH ≈2 with trifluoroacetic acid (TFA). Samples were centrifuged (20 min, 16,000 x g) and supernatants were desalted and concentrated on Sep-Pack C_{18} Cartridges (Waters).

Enrichment of Acetylated Peptides

Peptides were eluted from the cartridges and acetylated peptides were enriched from the supernatants by incubation with agarose conjugated acetyl lysine antibody (acetyl lysine antibody, ICP0388, Immunechem) at 4C. Bound peptides were eluted from the agarose beads in 0.1% TFA. The peptides were subsequently separated by SCX fractionation on in-house packed micro-columns (Rappsilber et al., 2007; Weinert et al., 2011) and eluted at pH4, pH6 and pH8.5 (Wiśniewski et al., 2009). Organic solvents were evaporated using a vacuum centrifuge, and the peptides were acidified in 0.1% TFA, 0.5% acetic acid and loaded onto in-house packed C_{18} STAGE tips.

LC-MS/MS

Peptide mixtures were eluted into a 96 well microtiter plates with 2x20ul 40% MeCN, 0.5% AcOH, followed by removal of organic solvents in a vacuum centrifuge and reconstitution of peptides in 2% MeCN, 0.5% AcOH, 0.1% TFA. The eluate was separated by online reversed-phase C_{18} nanoscale liquid chromatography on a 15 cm x 75 μm column in-house packed with ReproSil-Pur C_{18}-AQ 3 μm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). A nanoflow Easy –nLC system (Proxeon Biosystems, Odense, Denmark) was connected through a nano-electrospray ion source to the mass spectrometer. Peptides were separated by a linear gradient of increasing acetonitrile in 0.5% acetic acid for 180 min with a flow rate of 250 nl/min. The tandem mass spectrometry was performed on a LTQ Orbitrap Velos mass spectrometer (Thermo Electron, Bremen, Germany) using a top10 HCD fragmentation method with the replicate experiments analyzed on a Q Exactive mass spectrometer (Thermo Electron, Bremen, Germany). Settings used on the LTQ Orbitrap Velos included full-scan MS spectra acquired at a target value of 1e6 and a resolution of 30,000, and the HCD-MS/MS spectra recorded at a target value of 5e4 and with a resolution of 7,500 using normalized collision energy of 40%. Settings used on the Q Exactive were identical to the 'fast' method as previously described (Kelstrup et al., 2012) with an isolation width of 1.3 m/z. Briefly, a target value of 1e6 for full-scan MS with a resolution of 70,000 at m/z 200 was set. Fragment scans were analyzed with a fixed ion injection time of 60ms measured at a resolution of 17,500 at m/z 200.

Identification and Quantification of Acetylated Peptides

Raw MS files were processed using the MaxQuant software (ver.1.0.14.7 and v.1.2.0.29), Max-Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Munich) by which the precursor MS signal intensities were determined and HCD MS/MS spectra were *de*isotoped and filtered such that only the ten most abundant fragments per 100-m/z range were retained. Acetylated proteins were identified using the Mascot search algorithm by searching all MS/MS spectra against a concatenated forward/reversed version of rat and mouse International Protein Index v.3.37 protein sequence database supplemented with protein sequences of common observed contaminants like human keratins and porcine trypsin. Human skeletal muscle samples were searched against a concatenated forward/reversed version of human International Protein Index v.3.68 protein sequence database. The HCD-MS/MS spectra were searched with fixed modification of Carbamidomethyl-Cysteine and we allowed for variable modifications of oxidation (M), acetylation (protein N-term), Gln- > pyro-Glu, and acetylation (K). Search parameters were set to an initial precursor ion tolerance of 7 ppm, MS/MS tolerance at 0.02 Da and requiring strict tryptic specificity with a maximum of two missed cleavages. Label-free peptide quantification and validation was performed in the MaxQuant software suite. Acetylated peptides were filtered based on Mascot score, PTM (Andromeda) score, precursor mass accuracy, peptide length, and summed protein score to achieve an estimated FDR < 0.01 based on the forward and reversed identifications. The minimum required peptide length was set to six amino acids. We required a minimum Mascot score of 10 and a minimum Andromeda score of 25 and a delta score to the next best match of at least 5.

Data Analysis

Data analysis was performed using Microsoft Office Excel and Perseus (Max-Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Munich) software. Hierarchical clustering was performed in Perseus using Euclidian distance and average linkage clustering.

Subcellular Compartment Pathway and GO Analysis

We grouped all lysine acetylated proteins based on their subcellular localization as defined by the GO term cellular component. For each tissue we calculated the fraction of identified acetylated proteins per compartment and calculated the deviations by dividing by medians for each compartment. Pathway and GO enrichment analyses were performed for proteins specific for tissue clusters against all other identified proteins residing in the same compartments using the innateDB web tool (Lynn et al., 2008). Enriched REACTOME pathways and GO-terms for biological processes were determined based on their corresponding *P*-values for overrepresentation in a given subcellular compartment. *P*-values were calculated using a hypergeometric test and corrected for multiple testing with a Benjamini-Hochberg. A cut-off of 0.01 was applied.

Sequence Motif Analysis

We performed sequence motif analysis using iceLogo with percentage difference as scoring system and a *P*-value cut-off of 0.01 using the *Rattus Norvegicus* protein sequence database as background. For heat maps: for all identified lysine acetylation sites as well as for lysine acetylation sites from each specific compartment we analyzed the amino acid sequences surrounding the sites by calculating their position-specific frequencies and dividing this by the amino acid frequencies of the *Rattus Norvegicus* protein sequence database. Significance for overrepresentation of identified motifs was evaluated by Fisher exact test calculations with a *P*-value cut-off of 0.05.

Expression of Recombinant Proteins

Gateway Entry clones of human ALDOB and GPD1 were obtained from the Invitrogen Ultimate ORF collection (Invitrogen). The coding sequences were cloned into pDEST-15 (Invitrogen). The lysine to glutamic acid (K120Q for GPD1 and K147Q for ALDOB) point mutations were introduced by PCR-based mutagenesis. All constructs were verified by sequencing and transformed into Rosetta cells (Novagen). Expression of recombinant proteins were induced with 1mM IPTG (Sigma Aldrich) overnight at 4°. Cells were harvested and washed with PBS. Proteins were extracted with Bugbuster (Novagen) supplemented with protease inhibitors, (Roche Complete Protease Inhibitor tablets, Roche), 50mM Tris-HCl (pH 7,5), 150 mM NaCl, 1mM DTT, 1 μl/ml Benzonase and 200 μg/ml Lysozyme. The extract was subjected to sonication and snap freezing in liquid nitrogen. After clarification by centrifugation the extracts were incubated on Glutathione Sepharose Beads (GE Healthcare) overnight at 4°. After washing the proteins were eluted with 15 mM reduced glutathione (Sigma Aldrich) supplemented with 50mM Tris-HCl (pH 7,5), 150 mM NaCl and 1mM DTT. The purified proteins were analyzed by SDS-PAGE and the concentration determined (Quick Start Bradford Dye Reagent X1, Bio-Rad). The amino acid sequences of the purified proteins were furthermore checked by mass spectrometry.

Enzyme Activity Assay

The catalytic activity of GPD1 toward dihydroxyacetone phosphate was investigated by measuring the simultaneous oxidation of NADH. The GPD1 activity was measured in a 200 μl reaction volume containing GPD1, 1 mM D-Fructose-1,6-bisphosphate, 150 μM NADH, 2 U Aldolase (Sigma), 2 U Triosephosphate isomerase (Sigma) in 50 mM Tris-HCl. The enzyme activity is determined as initial velocity measurements by monitoring the NADH decrease as a function of time at 340 nm absorbance reads on a Biotek Synergy H4 reader. Reconstituted purified wild-type or mutant GPD1 (0.8 μg/ml) was diluted in assay buffer and added to the other reaction components resulting in 0.83 μg/ml wt and 0.83-212 μg/ml mutant GPD1 in the assay. Similarly, the catalytic activity of ALDOB toward fructose-1,6-bisphosphate was assayed by the downstream oxidation of NADH. Reconstituted purified wild-type or mutant ALDOB (13 μg/ml) was mixed in a 50 mM Tris-HCl buffer with 1 mM D-Fructose-1,6-bisphosphate, 150 μM NADH, 2 U Triosephosphate isomerase and 150 μg/ml GPD1 (wt). The cleavage of fructose-1,6-bisphosphate was measured as initial velocities using 3.3 μg/ml wt and 3.3 μg/ml - 212 μg /ml mutant ALDOB. Both enzymatic assays were conducted in triplicates at room temperature, and the results are shown as mean ± SEM. Assay detection limits are established based on blank samples without enzyme and converted to units of μg/min/mg using the enzyme concentration 212 μg/ml.

CPR PTM Resource

The CPR PTM Resource is a web-based data repository that integrates high-confidence *in-vivo* post-translational modification sites that we have identified by mass spectrometry-based proteomics in different tissue samples from various species. It is based on a MySQL database and developed with the Perl/CGI language. Modified proteins can be visualized based on their Uniprot identifiers. For each modified site in a protein we list the matching acetylation motifs we have identified and use the Reflect service to add additional information about the modified proteins.

Protein-protein interaction networks

Protein-protein interaction networks were built using a previously described up to date interaction network of quality controlled human protein interactions (InWeb, Lage et al., 2007 (Lage et al., 2007)). In brief, InWeb combines reported protein interactions

from MINT, BIND, IntAct, KEGG annotated protein-protein interactions (PPrel), KEGG Enzymes involved in neighboring steps (ECrel), Reactome and others. All human interactions are pooled and interactions in orthologous protein pairs passing a strict threshold for orthology are included. Each interaction is assigned a probabilistic score based on the neighborhood of the interaction, the scale of the experiment in which the interaction is reported and the number of different publications in which the interaction has been cited. To generate a network InWeb is scanned for interaction partners of each of the tissue-specific acetylation data sets (the input sets). For each input set, various parameter settings (density of interactions and different thresholds of confidence scores) are tested to identify the most significant network for the input set. The significance of each of the generated networks is determined by randomization testing. Specifically, for an input set of N input proteins yielding an interaction network (connected component) with G input proteins and T total proteins, a network score (NSinput) is determined. This network score is the fraction of input proteins of all proteins in the network (G/T). The significance of the network score is next determined by empirically estimating the probability of observing a similar or better network score in networks generated by using 10 000 random input sets of size Ninput. The random gene sets are chosen so the degree distribution of proteins in the random sets approximate the input set. As each query generates a varying number of networks (connected components) the probability estimates can be calculated from the total amount of networks produced by all 10 000 randomizations that have a network score > NSinput. All these algorithms have been thoroughly described elsewhere (Bergholdt et al., 2007; Lage et al., 2010). Furthermore, to probe for pathological processes associated with the tissue-specific protein-interaction networks, we analyzed each of the networks for genes known to be mutated in familial diseases, as described in Online Mendelian Inheritance in Man (OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), January 2011). The p values were adjusted for multiple hypothesis testing by Bonferroni correction.

SUPPLEMENTAL REFERENCES

Kelstrup, C.D., Young, C., Lavallee, R., Nielsen, M.L., and Olsen, J.V. (2012). Optimized Fast and Sensitive Acquisition Methods for Shotgun Proteomics on a Quadrupole Orbitrap Mass Spectrometer. J. Proteome Res. *11*, 3487–3497.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Wiśniewski, J.R., Zougman, A., and Mann, M. (2009). Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J. Proteome Res. *8*, 5674–5678.

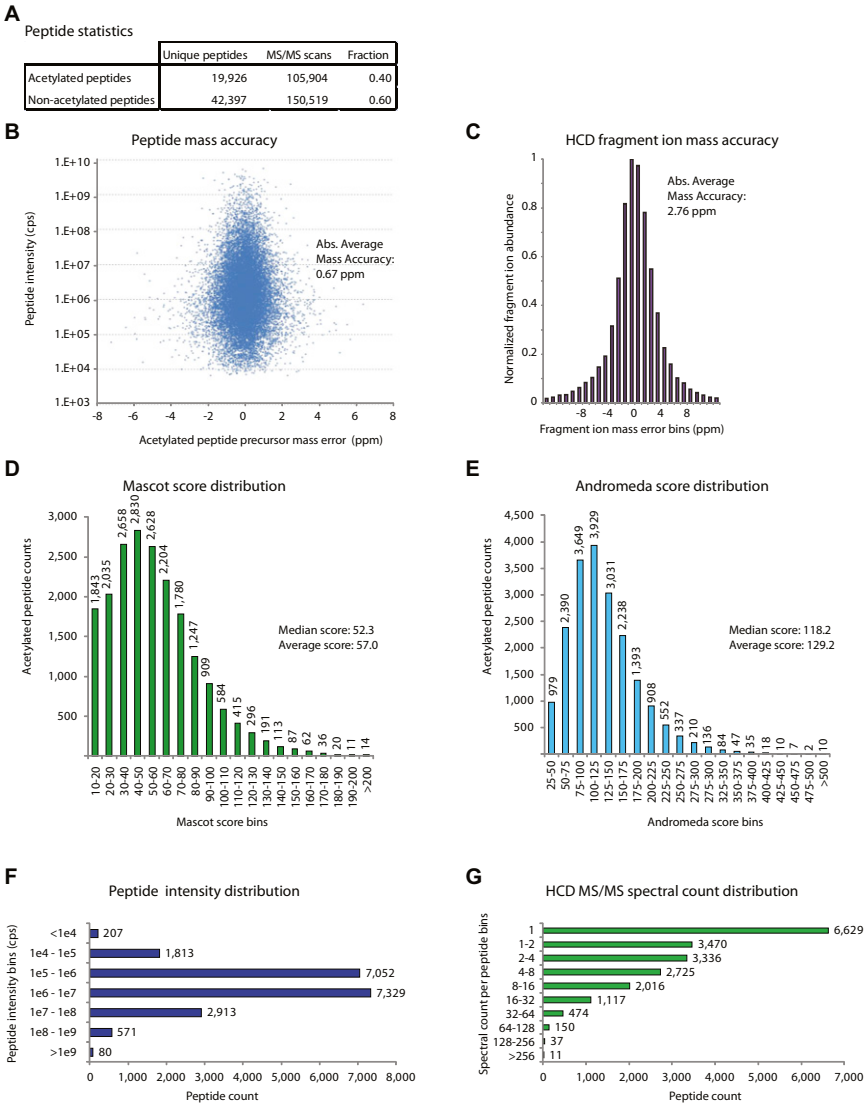


Figure S1. Evaluation of MS and MS/MS Data Quality, Related to Figure 1

(A) Peptide statistics for acetylated and non-acetylated peptides.

(B) Peptide intensity for all acetylated peptides is plotted as function of calibrated peptide precursor mass errors measured for all identified peptides in parts-per-million (ppm) and the absolute average mass accuracy is calculated.

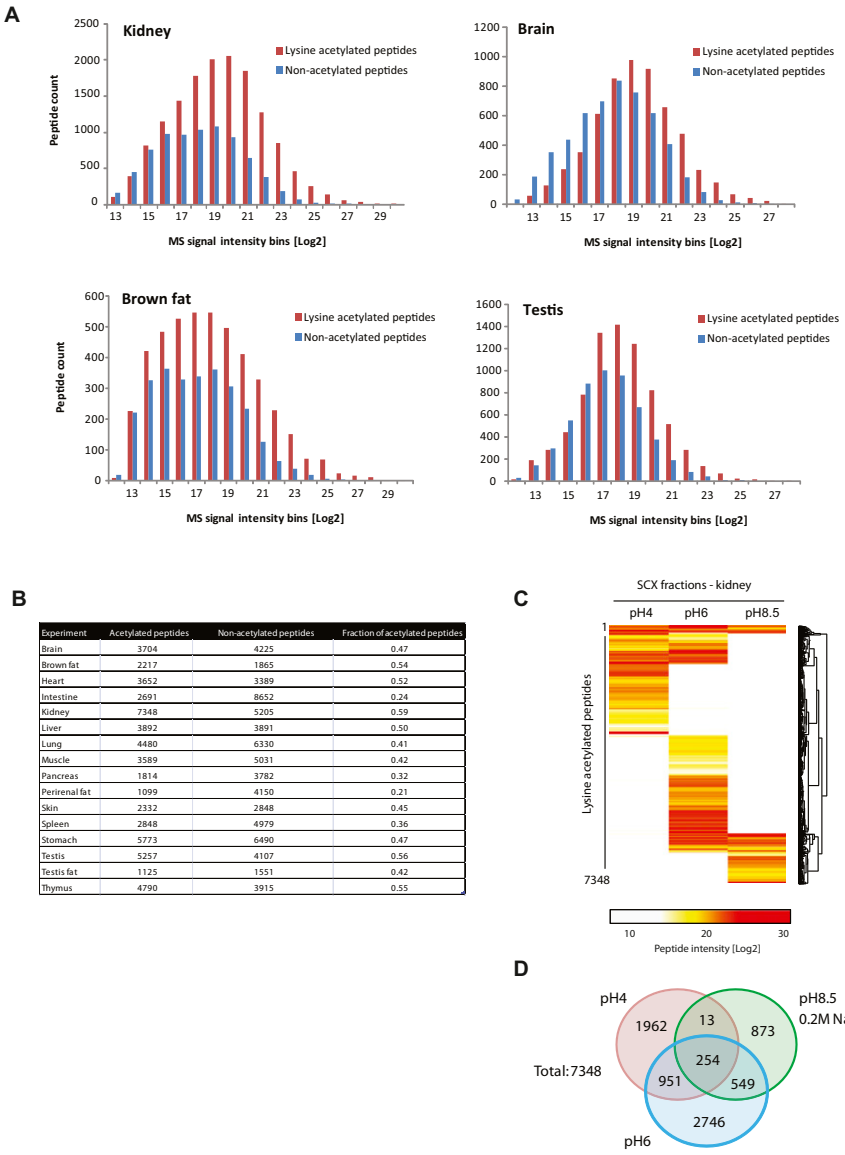
(C) Fragment ion mass accuracy. Histogram of HCD fragment ion errors binned in 1 ppm units.

(D) Histogram illustrating the Mascot score distribution of all acetylated peptides. Mascot scores are binned in 10-score units and each bin is displayed as a bar indicating the peptide count.

(E) Histogram of Andromeda P-score distribution of all acetylated peptides. Andromeda scores are binned in 25-score units.

(F) Distribution of peptide precursor intensities. Histogram of the number of peptides per bin of normalized label-free XIC-based intensities for all identified peptides.

(G) Spectral count distribution. Histogram of HCD-MS/MS spectra count per identified peptide.



Experiment	Acetylated peptides	Non-acetylated peptides	Fraction of acetylated peptides
Brain	3704	4225	0.47
Brown fat	2217	1865	0.54
Heart	3652	3389	0.52
Intestine	2691	8652	0.24
Kidney	7348	5205	0.59
Liver	3892	3891	0.50
Lung	4480	6330	0.41
Muscle	3589	5031	0.42
Pancreas	1814	3782	0.32
Perirenal fat	1099	4150	0.21
Skin	2332	2848	0.45
Spleen	2848	4979	0.36
Stomach	5773	6490	0.47
Testis	5257	4107	0.56
Testis fat	1125	1551	0.42
Thymus	4790	3915	0.55

Figure S2. Lysine-Acetylated Peptide Enrichment and SCX Fractionation Efficiency, Related to Figure 1

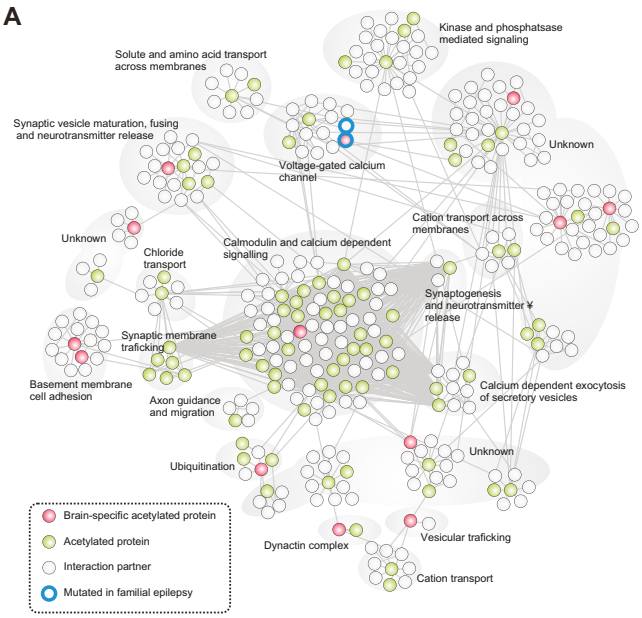
(A) Peptide signal intensity distributions of binned log2-transformed peptide XICs based on all lysine acetylated peptide spectrum matches (red bars) and all non-acetylated peptide spectrum matches (blue bars) are shown for kidney, brain, brown fat and testis.

(B) Table showing lysine acetylated peptide enrichment efficiencies across tissues based on unique lysine acetylated peptides.

(C) Heat-map depicting the unique lysine acetylated peptide signal intensities across the three kidney SCX fractions from pH elutions 4, 6 and 8.5.

(D) Venn-diagram displaying the overlap of acetylated peptides between the three SCX pH-elution fractions displayed in panel C.

A



B

Brain, Adj. Pval = 0.047
Brown Fat, Adj. Pval = 1
Heart , Adj. Pval = 0.041
Intestine, Adj. Pval = 0.57
Kidney, Adj. Pval = 0.095
Liver, Adj. Pval = 0.026
Lung, Adj. Pval = 0.131
Muscle, Adj. Pval = 0.465
Pancreas, Adj. Pval = 0.033
Skin, Adj. Pval = 0.029
Spleen, Adj. Pval = 0.881
Stomach, Adj. Pval = 0.214
Testes, Adj. Pval = 0.442
Testes Fat, Adj. Pval = 0.359
Thymus, Adj. Pval = 0.182

Figure S3. Protein Interaction Network Based on Proteins Specifically Acetylated in Brain, Related to Figure 2

(A) Brain-specific proteins (red) and their direct interaction partners as well as indirect interaction partners through common partners are shown. Interaction partners that are identified to be acetylated in our data set are depicted in green, those that we have not identified to be acetylated are depicted in gray. The network consists of protein clusters with functions relevant to brain such as synaptic vesicle maturation, calcium mediated signaling, axon guidance and migration, synaptogenesis and neurotransmitter release, and vesicular trafficking, among others. Moreover this network is enriched for proteins coded by genes in which Mendelian mutations cause juvenile myoclonic epilepsy (adj. $p = 8.3 \times 10^{-3}$), suggesting that the 16 tissue acetylomes elucidated in this work can be used as a starting point for deciphering tissue-specific biology relevant for human pathology.

(B) The table summarizes the significances (adjusted for multiple hypothesis testing) of the interaction networks delineated for each tissue, which are available for download from http://cpr1.sund.ku.dk/cgi-bin/PTM.pl. We analyzed the diseases associated with each of the five significant tissue-specific networks, and found that two of the networks were significantly enriched for proteins coded by genes known to be mutated in familial diseases associated with the relevant tissue, as described in Online Mendelian Inheritance in Man (OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), January 2011, http://omim.org/). The network based on proteins acetylated in brain is enriched (adj. $p = 8.3 \times 10^{-3}$) for proteins coded by genes involved in juvenile myoclonic epilepsy (OMIM identifier 606904); the genes in question are EFHC1 and CACNB4 both involved in calcium homeostasis and the heart network is enriched (adj. $p = 1.45 \times 10^{-4}$) for proteins coded by genes involved in cardiomyopathy (OMIM identifier 192600); the genes in question are CAV3, MYBPC3 and TNNI3.

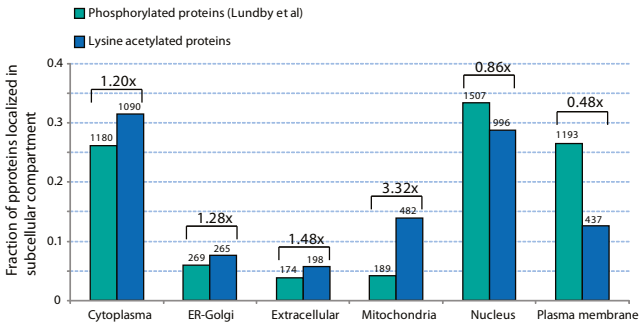


Figure S4. Global Subcellular Distribution of the Tissue Acetylome versus Phosphoproteome, Related to Figure 5

The fraction of lysine acetylated proteins (green) and the fraction of phosphorylated proteins (blue) residing in the main subcellular compartments was calculated based on global rat tissue data sets of acetylome and phosphoproteome (Lundby et al., 2012). The number of proteins underlying the fraction represented by each bin is displayed above the histogram bars as are the fold-change in fraction localization to a compartment for lysine acetylated proteins relative to phosphorylated proteins.

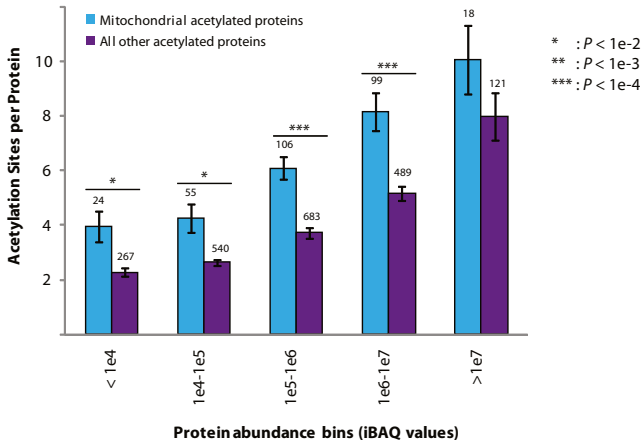


Figure S5. Relationship between Protein Abundance and Acetylation Site Counts, Related to Figure 5

Mitochondrial proteins and all non-mitochondrial proteins are binned according to their iBAQ abundance values and the average number of acetylation sites per protein in each bin is displayed with error bars (standard error of the mean). The total number proteins in each bin are indicated above the histogram bars, and t test based statistical significance is shown.

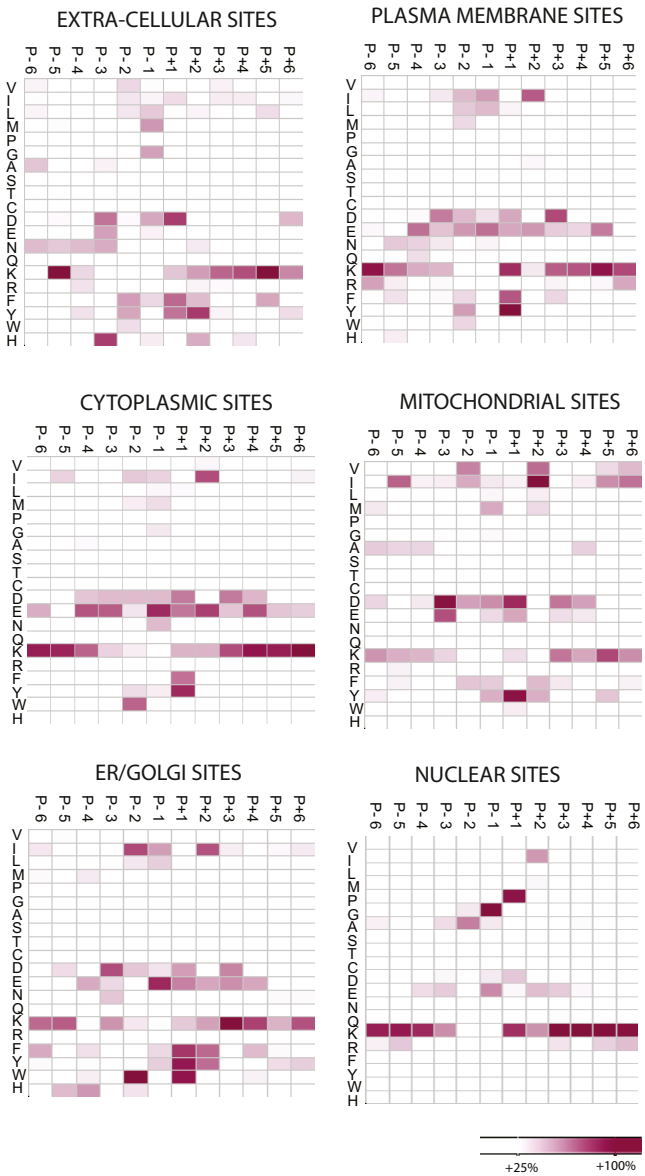


Figure S6. Cellular Compartment-Specific Sequence Logo Heatmaps, Related to Figure 6

Heat maps indicate overrepresentation of amino acids in all positions from −6 to +6 from the acetylated lysine residue based on all identified acetylation sites.