Supplementary file

Comprehensive analysis of lysine acetylome reveals a site-specific pattern in rapamycin-induced autophagy

Zhuo Zhou[1#], Yu Chen[1#], Mengqi Jin[1], Jianqin He[2], Ayiding guli[1], ChunlanYan[3*], Shiping Ding[1*]

[1]The National Education Base for Basic Medical Sciences, School of Medicine, Zhejiang University, Hangzhou, Zhejiang Province, China

[2]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, the First Affiliated Hospital of Medical College, Zhejiang University, Hangzhou, Zhejiang Province, China

[3]Department of Biophysics, Key Laboratory of Medical Neurobiology, Ministry of Health of China, School of Medicine, Zhejiang University, Hangzhou, Zhejiang province, China

[#]Co-first author

*Corresponding author, Shiping Ding: Tel.: +86-571-88208521

E-mail address: dingshiping@zju.edu.cn (Shiping Ding)

clyan@zju.edu.cn (Chunlan Yan)

Postal address: The National Education Base for Basic Medical Sciences, School of Medicine, Zhejiang University, Hangzhou 310058, Zhejiang Province, PR China.

Supplemental Table S1-S9

Supplemental Figure S1-S2

Supplemental Material and Methods

MS/MS raw data analysis

Acetylated quantitative analysis

Clustering Method

The normalization of the acetylated ratios for protein abundances

Analysis of sequence model around Kac

IceLogo generation

Protein annotation and clustering of acetylated proteins

Protein complex enrichment and clustering analysis

Functional enrichment analysis

Public access to MS/MS data

Supplemental References

Supplemental Table

Table S-1. The inhibitive effect of rapamycin on the proliferation of HeLa cells

Table S-2. Summary of acetylated proteins and acetylated sites for three replicates. 2a: replicate 1; 2b: replicate 2; 2c: replicate 3.

Table S-3. Summary of acetylated peptides quantification for three replicates. 3a: replicate 1; 3b: replicate 2; 3c: replicate 3.

Table S-4. Summary of proteome data for three replicates. 4a: replicate 1; 4b: replicate 2; 4c: replicate 3.

Table S-5. Summary of protein peptides quantification for three replicates. 5a: replicate 1; 5b: replicate 2; 5c: replicate 3.

Table S-6. (a) The list of proteins affiliating to the autophagy-related interaction network shown in Table S2. (b) The overlapped sites between our data and Morselli's.

Table S-7. Gene Ontology analysis of the significantly-regulated proteins based on three categories: (a) biological process; (b) molecular function.

Table S-8. KEGG enrichment analysis of the significantly-regulated proteins.

Table S-9. Complex analysis of the acetylated proteins with CORUM database.

Supplemental Figure

FigureS-1 Acetylation-modulated functional networks

Interaction networks of acetylated proteins in different cellular processes from STRING analysis of the acetylome. Individual networks were generated for

each functional category with a confidence cutoff of 0.70. (A) ribosome; (B) spliceosome; (C) AcCoA-related metabolism.

Figure S-2. Representative MS/MS spectrum of acetylated proteins.

(A) The acetylation of ACLY at K468.

(B) The acetylation of CBP at K1583 and K1586.

(C) The acetylation of PFKL at K469.

(D) The acetylation of p300 at K373.

Supplemental Materials and methods

MS/MS raw data analysis

Xcaliber software version 3.0.63 was used for data collection. The resulting MS/MS (1) data was processed using MaxQuant (2) with integrated Andromeda search engine (v.1.5.2.8). Tandem mass spectra were searched against SwissProt Human database (20,203 entries, 8/2015) concatenated with reverse decoy database. Trypsin/P was specified as cleavage enzyme allowing up to 4 missing cleavages, 5 modifications per peptide. Mass error was set to 20 ppm for precursor ions and 0.02 Da for fragment ions. Carbamidomethylation on Cys was specified as fixed modification and oxidation on Met, acetylation on Lys, and acetylation on protein N-terminal were specified as variable modifications. False discovery rate (FDR) thresholds for protein, peptide and modification site were specified at 1%. Minimum peptide length was set at 7. The site localization probability was set as > 0.75. Peptide identification score for each peptide was specified at more

than 40. The normalized Ratio H/L from the MaxQuant output files ProteinGroups.txt and acetyl (K) Sites.txt were used for proteins and acetylation sites quantification respectively.

Acetylated quantitative analysis

Firstly, the quantified proteins in this study were divided into four quantitative categories according to the quantification ratio to generated four quantitative categories: Q1 (0< Ratio L/H < 0.667), Q2 (0.67 < Ratio L/H < 0.77), Q3 (1.3 < Ratio L/H <1.5) and Q4 (Ratio L/H >1.5).

The normalization of the acetylated ratios for protein abundances

The acetylated ratios were normalized for protein abundances. We selected the average of three ratios of three replicates as the reference.

Clustering Method

All the substrates categories obtained after enrichment were collated along with their *P* values, and then filtered for those categories which were at least enriched in one of the clusters with *P* value <0.05.

Analysis of sequence model around Kac

Soft motif-x (3) was used to analyze the model of sequences constituted with amino acids in specific positions of acetyl-13-mers (6 amino acids upstream and downstream of the acetylation site) in all protein sequences. And IPI human proteome sequences were used as background database parameter, other parameters with default.

IceLogo generation

For acetylation site analysis of all significantly-regulated sites, amino acid sequence windows of 6 amino acids downstream as well as 6 amino acids upstream of the modified lysine were extracted from the corresponding proteins, generating 13 amino acid (13AA) sequence windows. IceLogo software version 1.2 8 was used to overlay 13AA acetylation site sequence windows in order to generate a consensus sequence, and compensated against the expected random occurrence frequencies of amino acids across all human proteins (iceLogo). Alternatively, subsets of acetylation sites were compared directly to other subsets of modification sites, generating consensus sequences showing differential occurrence of amino acids between the subsets (SubLogo). Heat maps were generated in a similar fashion to IceLogos. For IceLogos, SubLogos and heat maps, all amino acids shown as enriched or depleted are significant with $P < 0.05$, as determined by IceLogo software version 1.2 (4).

Protein annotation and clustering of acetylated proteins

Gene Ontology (GO) annotation proteome was derived from the UniProt-GOA database (http://www.ebi.ac.uk/GOA/). Kac proteins were classified by Gene Ontology (5) analysis based on three categories: biological process, cellular component and molecular function. Kyoto Encyclopedia of Genes and Genomes (KEGG) database (6) was used to analyze protein pathway. Domain analysis was performed by using InterProScan on InterPro domain database via Web-based interfaces and services. WoLF PSORT was used for

subcellular localization prediction.

Protein complex enrichment and clustering analysis

CORUM (7) (the Comprehensive Resource of Mammalian protein complexes）database was used for protein complex analysis. Overrepresented complexes were identified using hypergeometric test for each quantile. Protein-protein interaction database STRING (8) was used for protein interaction network analysis. MCODE (9) plug-in tool in Cytoscape (version 3.1.0) (10) was used to analyze the interactions among all acetylated proteins and detected protein networks were exported.

Functional enrichment analysis

Fisher's exact test was used to test for enrichment or depletion (right-tailed test) of specific annotation terms among members of resulting protein clusters. Derived p-values were further adjusted to address multiple hypotheses testing by Benjamini and Hochberg's method. Any terms having adjusted p-values below 0.05 in any of the clusters were treated as significant.

Public access to MS/MS data

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (11) partner repository with the dataset identifier PXD005793.

Supplemental References

1. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein

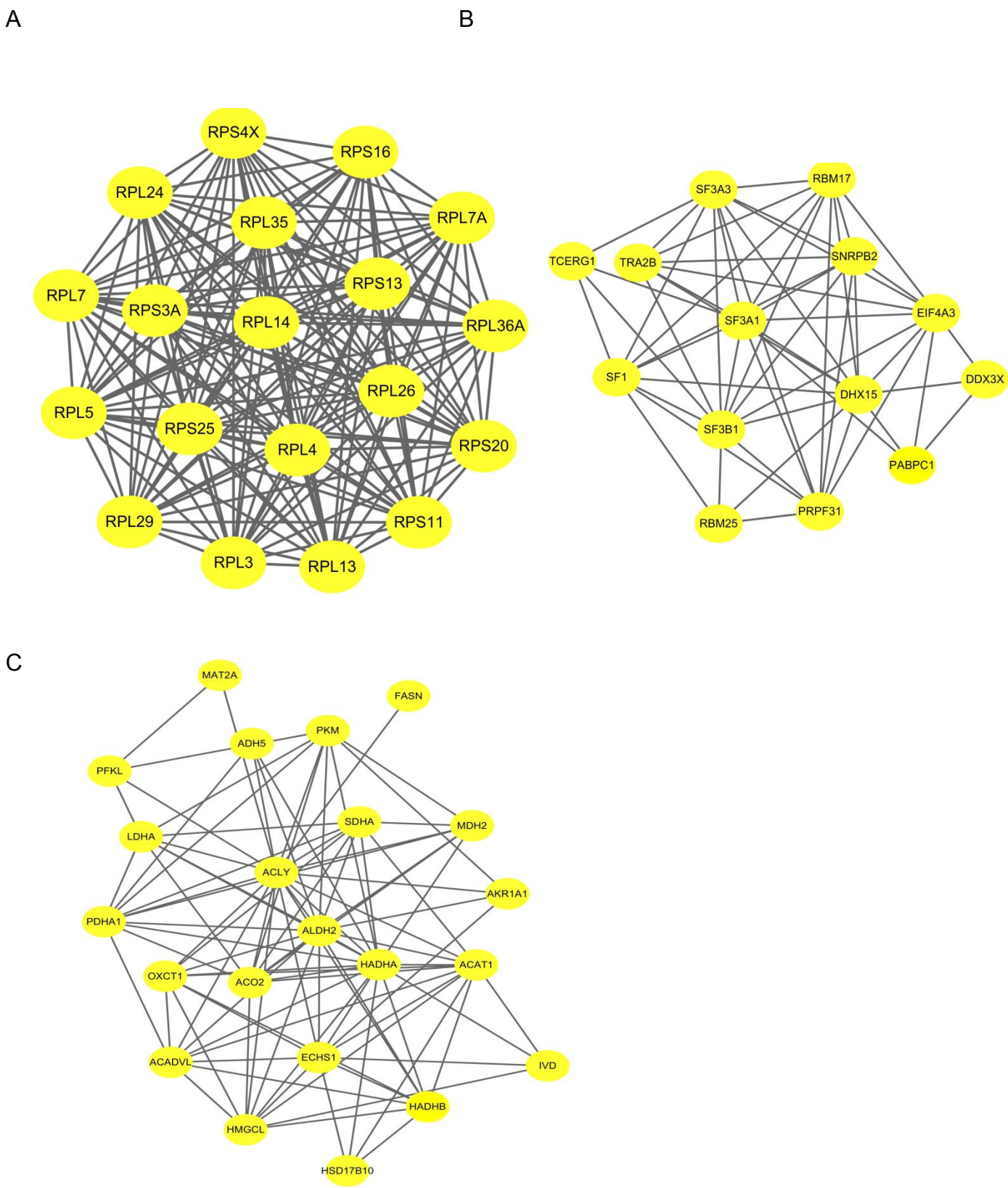
quantification. *Nat Biotechnol.* 2008;26(12): 1367-1372.

2. Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, Mann M. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc.* 2009;4(5):698-705.
3. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* 2005;23(11):1391-1398.
4. Hendriks IA, D'Souza RC, Yang B, Verlaan-de Vries M, Mann M, Vertegaal AC. Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol*. 2014;21(10): 927-936.
5. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1): 25-29.
6. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1): 27-30.
7. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW. CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Res.* 2010;38:D497-501.
8. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T,

Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 2009;37: D412-D416.

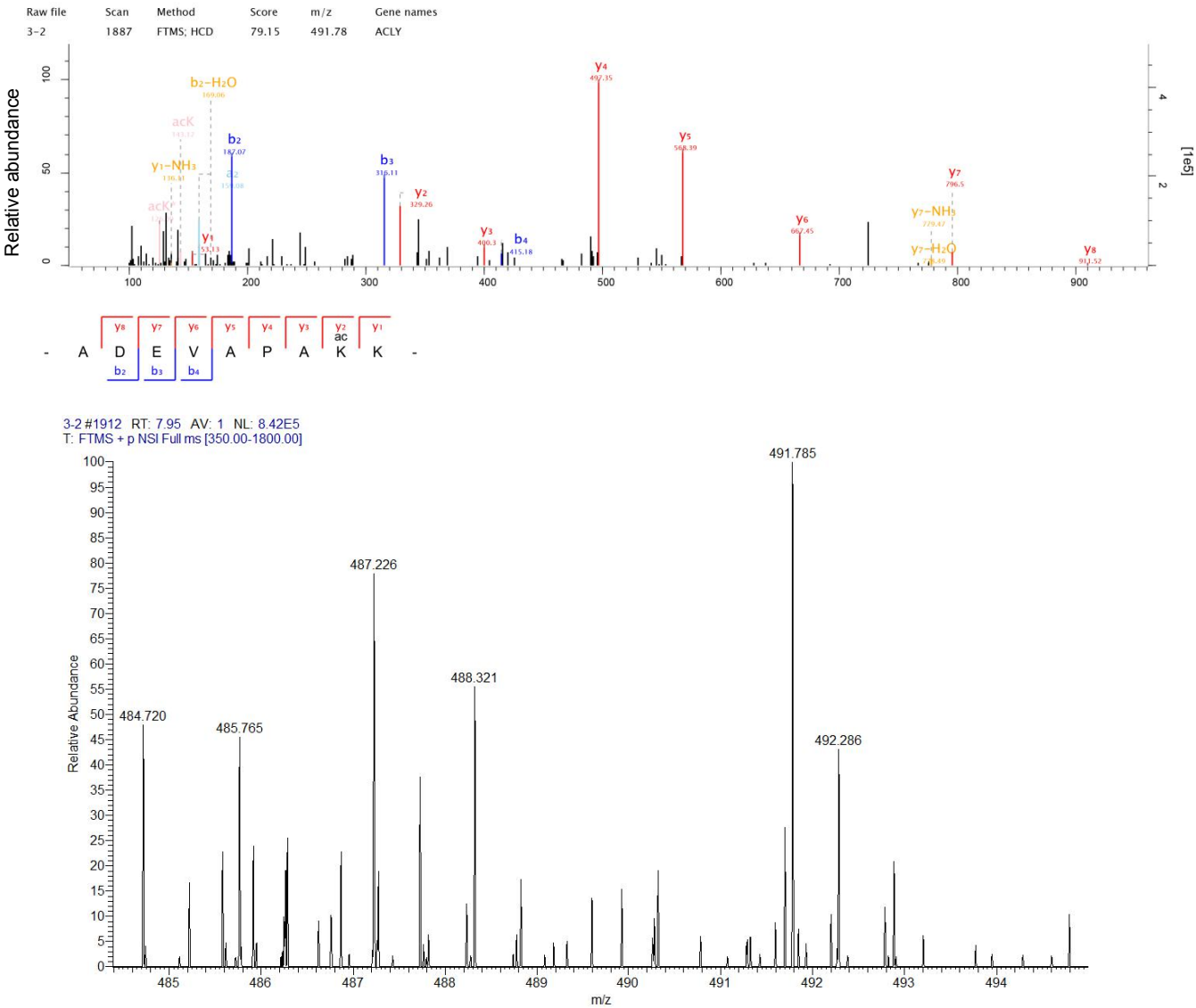
9. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4: 2.
10. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11): 2498-2504.
11. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H. 2016 update of the PRIDE database and related tools. *Nucleic Acids Res.* 2016;44(D1): D447-D456.

Supplement figure 1

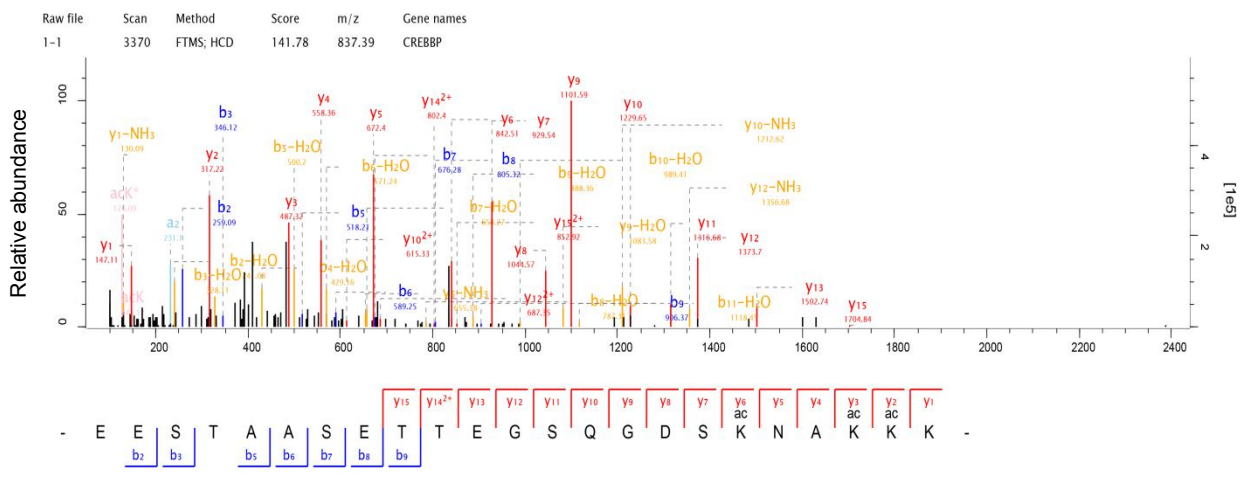


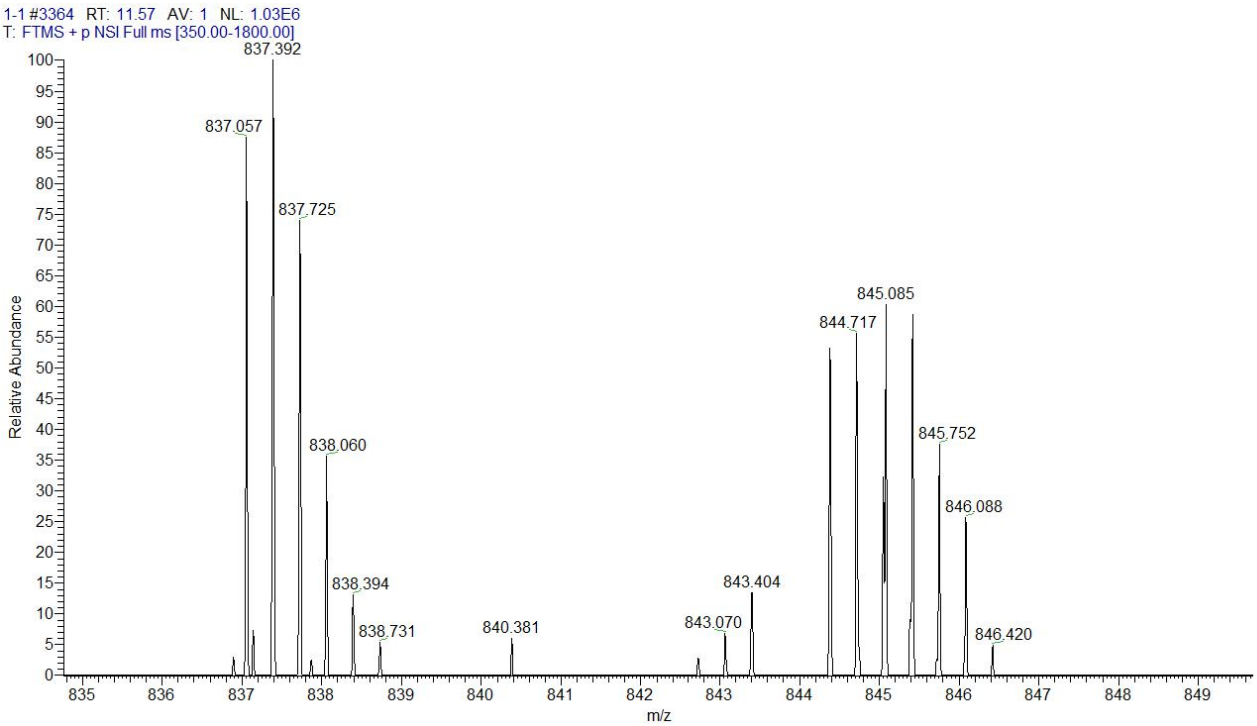
Supplement figure 2

A The acetylation of ACLY at K468



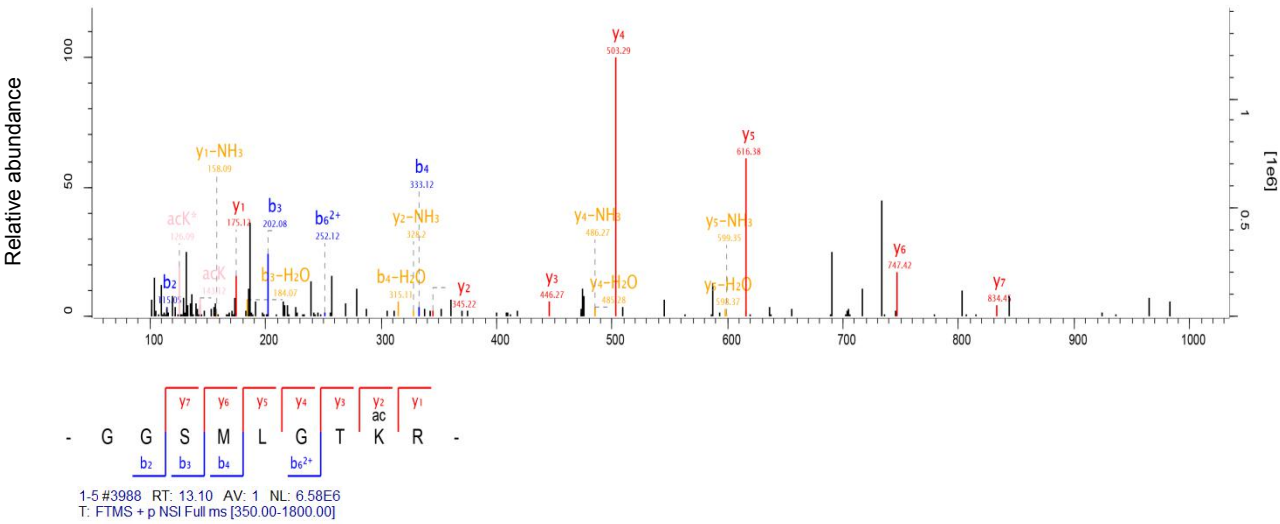
B The acetylation of CBP at K1583 and K1586

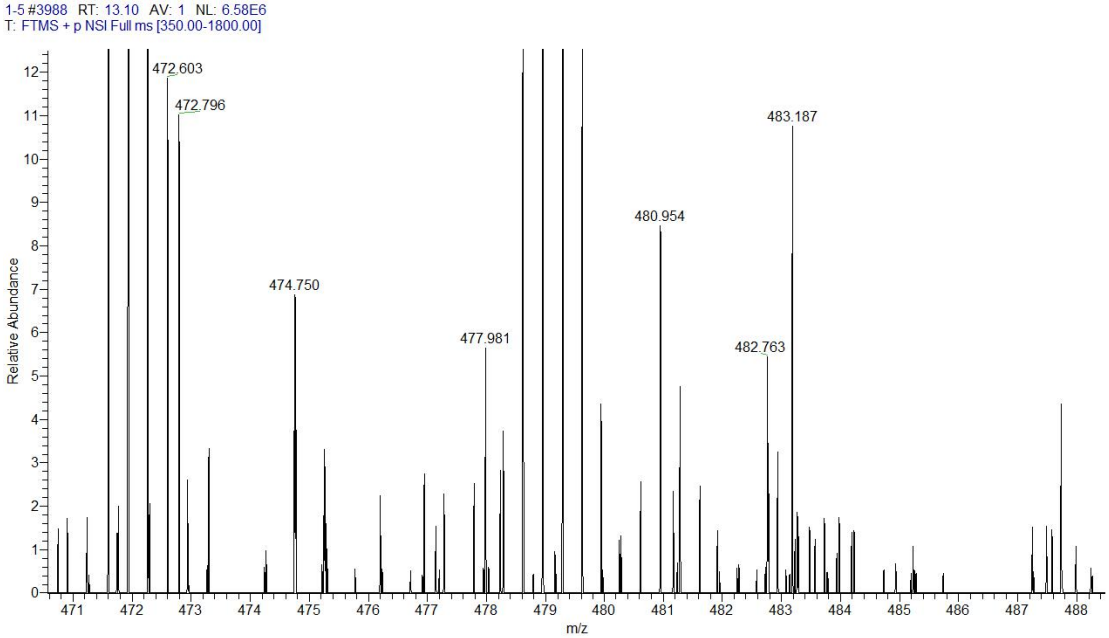




C The acetylation of PFKL at K469

Raw file	Scan	Method	Score	m/z	Gene names
1-5	3995	FTMS; HCD	65.93	474.75	PFKL





D The acetylation of p300 at K373