In-depth Analysis of the Sirtuin 5-regulated Mouse Brain Malonylome and Succinylome using Library-free Data-Independent Acquisitions

Joanna Bons[1], Jacob Rose[1], Ran Zhang[1], Jordan B. Burton[1], Christopher Carrico[1], Eric Verdin[1], Birgit Schilling[1]

[1] Buck Institute for Research on Aging, 8001 Redwood Boulevard, Novato, California 94945, USA

Corresponding author:

Birgit Schilling, PhD, Buck Institute for Research on Aging, 8001 Redwood Boulevard, Novato, California 94945, USA.

Email: bschilling@buckinstitute.org

Materials and Methods

Chemicals

LC-MS-grade acetonitrile and water were obtained from Burdick & Jackson (Muskegon, MI). Reagents for protein chemistry, including urea, iodoacetamide (IAA), dithiothreitol (DTT), triethylammonium bicarbonate (TEAB), and formic acid (FA) were purchased from Sigma-Aldrich (St. Louis, MO). Sequencing-grade trypsin was purchased from Promega (Madison, WI). HLB Oasis SPE cartridges were purchased from Waters (Milford, MA).

Mouse Brains

Animal study was performed according to protocols approved by IACUC (the Institutional Animal Care and Use Committee). SIRT5 knock-out (KO) mice were obtained from the Jackson Laboratory (Strain#012757). Mice were housed (12-h light/dark cycle, 22°C) and given unrestricted access to water. Brain tissues were collected from 18 months old females, after 6 hours of refeeding following 24 hours of fasting.

Protein Digestion and Desalting

Mouse brain tissues were collected from 2 different conditions with 4 biological replicates each: i) wild-type (WT, n=4), ii) $SIRT5^{(-/-)}$ (SIRT5-KO, n=4). Frozen brains were homogenized in lysis buffer containing 8 M urea, 50 mM Tris, pH 7.5, 7.5, 1x HALT protease inhibitor cocktail (Thermo Fisher Scientific, Waltham, MA), 150 mM NaCl, 5 μM trichostatin A, and 5 mM nicotinamide, and homogenized for 2 cycles with a Bead Beater TissueLyser II (Qiagen, Germantown, MD) at 25 Hz for 3 min each. Lysates were clarified by spinning at 16,500 x *g* for 15 min at 4°C, and the supernatant containing the soluble proteins was collected. Protein concentrations were determined using a Bicinchoninic Acid Protein (BCA) Assay (Thermo Fisher Scientific, Waltham, MA), and subsequently 3-mg and 5-mg protein from each sample were aliquoted for malonyl and succinyl peptide analysis, respectively. Proteins were reduced with 20 mM DTT for 30 min at 37 °C, and after cooling to room temperature, alkylated with 40 mM IAA for 30 min in the dark at room temperature. Samples were diluted 4-fold with 50 mM TEAB, pH 7.5, and proteins were digested overnight with a solution of sequencing-grade trypsin in 50 mM TEAB at a 1:50

(wt:wt) enzyme:protein ratio at 37°C. This reaction was quenched with 1% FA and the sample was clarified by centrifugation at 2,000 x *g* for 10 min at room temperature. Clarified peptide samples were desalted with Oasis 10-mg Sorbent Cartridges. 100 µg of each peptide elution were aliquoted for analysis of protein-level changes, after which all desalted samples were vacuum dried. The 100 µg whole lysate aliquots were re-suspended in 0.2% FA in water at a final concentration of 1 µg/µL and stored for MS analysis. The remaining 4.9 mg digests (for malonyl enrichment) and 2.9 mg digests (for succinyl enrichment) were re-suspended in 1.4 mL of immunoaffinity purification (IAP) buffer (Cell Signaling Technology, Danvers, MA) containing 50 mM 4-morpholinepropanesulfonic acid (MOPS)/sodium hydroxide, pH 7.2, 10 mM disodium phosphate, and 50 mM sodium chloride for PTM enrichment. Peptides were enriched for malonylation with anti-malonyl antibody conjugated to agarose beads from the Malonyl-Lysine Motif Kit (Kit #93872), and for succinylation with anti-succinyl antibody conjugated to agarose beads from the Succinyl-Lysine Motif Kit (Kit #13764; both from Cell Signaling Technology, Danvers, MA). This process was performed according to the manufacturer protocol; however, each sample was incubated in half the recommended volume of washed beads. Peptides were eluted from the antibody-bead conjugates with 0.1% trifluoroacetic acid in water and were desalted using C18 stagetips made in-house. Samples were vacuum dried and re-suspended in 0.2% FA in water. Finally, indexed retention time standard peptides (iRT; Biognosys, Schlieren, Switzerland) [1] were spiked in the samples according to manufacturer's instructions.

Mass Spectrometric Analysis

LC-MS/MS analyses were performed on a Dionex UltiMate 3000 system coupled to an Orbitrap Eclipse Tribrid mass spectrometer (both from Thermo Fisher Scientific, San Jose, CA). The solvent system consisted of 2% ACN, 0.1% FA in H_2O (solvent A) and 98% ACN, 0.1% FA in H_2O (solvent B). Proteolytic peptides were loaded onto an Acclaim PepMap 100 C18 trap column (0.1 x 20 mm, 5 µm particle size; Thermo Fisher Scientific) for 5 min at 5 µL/min with 100% solvent A. For the protein lysates, an amount of 200 ng was loaded (protein level analysis), and for the enriched malonylated and succinylated peptides (PTM level analysis), 4 µL of each PTM-enriched sample were injected. Peptides

were eluted on an Acclaim PepMap 100 C18 analytical column (75 µm x 50 cm, 3 µm particle size; Thermo Fisher Scientific) at 300 nL/min using the following gradient of solvent B: 2% for 5 min, linear from 2% to 20% in 125 min, linear from 20% to 32% in 40 min, and up to 80% in 1 min, with a total gradient length of 210 min.

All samples – protein level analysis and PTM level analyses – were acquired in data-independent acquisition (DIA) mode. Full MS spectra were collected at 120,000 resolution (AGC target: 3e6 ions, maximum injection time: 60 ms, 350-1,650 m/z), and MS2 spectra at 30,000 resolution (AGC target: 3e6 ions, maximum injection time: Auto, NCE: 27, fixed first mass 200 m/z). The isolation scheme consisted in 26 variable windows covering the 350-1,650 m/z range with an overlap of 1 m/z (**Table S1**) [2]. A detailed set-by-step procedure for building the DIA method and subsequent data processing can be found in **Appendix 1**.

Data analysis with directDIA (Spectronaut)

DIA data was processed in Spectronaut (version 14.10.201222.47784) using directDIA for both the protein level as well as PTM enriched samples. Data was searched against the *Mus musculus* proteome with 58,430 protein entries (UniProtKB-TrEMBL), accessed on 01/31/2018. Trypsin/P was set as digestion enzyme and two missed cleavages were allowed. Cysteine carbamidomethylation was set as fixed modification, and methionine oxidation and protein N-terminus acetylation as variable modifications. A relative intensity filter was selected to remove fragment ions relative ion intensity < 5% from the spectral library. Data extraction parameters were set as dynamic. Identification was performed using 1% precursor and protein q-value. For the protein lysate samples, quantification was based on the extracted ion chromatograms (XICs) of 3 – 6 MS2 fragment ions, local normalization was applied, and iRT profiling was selected. For the malonyl-enriched samples, lysine malonylation was additionally set as variable modification, and six missed cleavages were allowed, while for the succinyl-enriched samples, lysine succinylation was additionally defined as variable modification, and four missed cleavages were allowed. PTM localization was selected with a probability cutoff of 0.75. Quantification was based on the XICs of 3 – 6 MS2 fragment ions, specifically b- and y-ions, without normalization as well as data filtering using q-value sparse. Grouping and quantitation of

PTM peptides were accomplished using the following criteria: minor grouping by modified sequence and minor group quantity by mean precursor quantity. A detailed step-by-step procedure for processing the DIA data – protein-level and PTM-level analyses – can be found in **Appendix 1**.

Differential expression analysis was performed using a paired t-test, and p-values were corrected for multiple testing using the Storey method [3, 4]. For whole lysate (protein level) analysis, protein groups are required with at least two unique peptides. For determining differential changes, a q-value < 0.01 and absolute Log_2(fold-change) > 0.58 was required at the protein level, and a q-value < 0.05 and absolute Log_2(fold-change) > 0.58 at the PTM level. An absolute Log_2(fold-change) cutoff of 0.58, that is a fold-change cutoff of 1.5, was applied to highlight more subtle changes, which are robust and highly confident changes when combined with the two unique peptides and q-value < 0.01 criteria.

Data analysis with Skyline

To increase the confidence of the malonyl peptide analysis, malonylated peptides identified using the previously described analysis were further manually investigated in Skyline [5]. To do so, a spectral library was first built in Spectronaut from the malonyl peptide DIA acquisitions using the Biognosys (BGS) default settings, except that lysine malonylation was added as variable modification, and four missed cleavages were allowed, and contained 13,304 modified peptides and 3,112 protein groups (**Table S2**). The library was then imported into Skyline (Skyline-daily, version 21.2.1.377), and 468 malonylated peptides from 242 proteins and present in the library were finally targeted. Four missed cleavages were allowed, at least three product ions (mono- and doubly-charged y- and b-type ions) from ion-2 to last ion-1 were extracted, and all matching scans were used. Chromatographic peaks were manually checked to remove interfered transitions and correct peak boundaries. At least three transitions were kept per precursor ion. Peptide area was obtained by summing the corresponding transition peak areas.
Statistical analysis was performed in Skyline, and malonylated peptides with p < 0.05 and absolute fold-change > 1.5 were considered as significantly altered.

Clustering Analysis

Partial least squares-discriminant analysis (PLS-DA) of the proteomics data was performed using the package mixOmics [6] in R (version 4.0.2; RStudio, version 1.3.1093).

Enrichment Analysis

An over-representation analysis (ORA) was performed using Consensus Path DB-mouse (Release MM11, 14.10.2021) [7, 8] to evaluate which gene ontology (GO) terms were significantly enriched. Gene ontology terms identified from the ORA were subjected to the following filters: q-value < 0.01, and term level > 3 for GO. Dot plots were generated using the ggplot2 package [9] in R (version 4.0.5; RStudio, version 1.4.1106). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using the ClueGO package [10] (version 2.5.8), in Cytoscape [11] (version 3.8.2). Default settings were applied, except that two-sided hypergeometric test Bonferroni-adjusted p-value threshold was set to 0.01. Kappa score threshold was let at 0.4 for drawing pathway-connecting edges. Pathways with the same color indicate at least 50% similarity in genes/term.

iceLogo Heatmap

Heatmap of 15 amino acid-long sequences centered around the modified lysine residues were generated using the iceLogo tool [12] to determine the frequency of every amino acid residue around the modified lysine residue. The *Mus musculus* precompiled Swiss-Prot composition was chosen, and the start position was set to 0. The p-value was set at 0.05, and significantly up- and down-regulated residues are colored in shades of green and red, respectively.

Subcellular Localization

Subcellular localization was determined using Cytoscape [11] (version 3.8.2) and stringAPP [13] (version 1.7.0), by applying default settings, except that species was defined as *Mus musculus*. Only compartments with scores above 4.5 were considered.

References

[1] Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., . . . Rinner, O. (2012). Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics, 12*(8), 1111-1121.

[2] Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., . . . Reiter, L. (2017). Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol Cell Proteomics, 16*(12), 2296-2309.

[3] Burger, T. (2018). Gentle Introduction to the Statistical Foundations of False Discovery Rate in Quantitative Proteomics. *J Proteome Res, 17*(1), 12-22.

[4] Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(3), 479-498.

[5] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., . . . MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics, 26*(7), 966-968.

[6] Rohart, F., Gautier, B., Singh, A., & Le Cao, K. A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol, 13*(11), e1005752.

[7] Kamburov, A., Wierling, C., Lehrach, H., & Herwig, R. (2009). ConsensusPathDB--a database for integrating human functional interaction networks. *Nucleic Acids Res, 37*(Database issue), D623-628.

[8] Kamburov, A., Pentchev, K., Galicka, H., Wierling, C., Lehrach, H., & Herwig, R. (2011). ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res, 39*(Database issue), D712-717.

[9] Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*: Springer New York.

[10] Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., . . . Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics, 25*(8), 1091-1093.

[11] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res, 13*(11), 2498-2504.

[12] Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., & Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat Methods, 6*(11), 786-787.

[13] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res, 43*(Database issue), D447-452.