

Supplementary Materials

Comprehensive analysis of the gut microbiome and post-translational modifications elucidates the route involved in microbiota-host interactions

Hai-Yang Wang^{1,2,3,#}, Lan-Xiang Liu^{4,#}, Xue-Yi Chen^{5,#}, Yang-Dong Zhang^{6,#}, Wen-Xia Li¹, Wen-Wen Li⁵, Lian Wang¹, Xiao-Long Mo^{1,6}, Hong Wei^{7,*}, Ping Ji^{2,3*}, Peng Xie^{1,2,6,*}

¹NHC Key Laboratory of Diagnosis and Treatment on Brain Functional Diseases, First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

²College of Stomatology and Affiliated Stomatological Hospital of Chongqing Medical University, Chongqing 401147, China

³Chongqing Key Laboratory for Oral Diseases and Biomedical Sciences, Chongqing 401147, China

⁴Department of Neurology, Yongchuan Hospital of Chongqing Medical University, Chongqing 402160, China

⁵Department of Pathology, Faculty of Basic Medicine, Chongqing Medical University, Chongqing 400016, China

⁶Department of Neurology, First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

⁷Yu-Yue Pathology Scientific Research Center, Chongqing 401329, China

#Authors contributed equally to this work

*Corresponding authors, E-mail: Weihong63528@163.com; jiping@hospital.cqmu.edu.cn; xiepeng@cqmu.edu.cn

Supplementary Materials and Methods

Sample collection and preparation

GF and SPF mice were anesthetized with 10% chloral hydrate (0.3 mL per 100 g weight), perfused with chilled phosphate buffered saline, then sacrificed by immediate decapitation. The hippocampus was immediately dissected on ice, then frozen in liquid nitrogen and stored at -80°C until use. The hippocampal tissue was grinded with liquid nitrogen into cell powder and then transferred to a 5-mL centrifuge tube. After that, lysis buffer (8 M urea, 1% protease inhibitor cocktail, 2 mM EDTA, 3 μM TSA, 50mM NAM and 10 mM DTT) was added to the cell powder, followed by sonication three times on ice using a high intensity ultrasonic processor (Scientz). The solution was centrifuged at 20,000 g for 10 min at 4°C to remove debris. The protein was precipitated for 2 h at -20°C . After centrifugation at 4°C for 10 min, the supernatant was discarded. The remaining precipitate was washed three times with cold acetone. The protein was redissolved in buffer (8 M urea, 100 mM TEAB, pH 8.0) and protein concentration was determined using a BCA kit. The protein solution was then reduced (10 mM DTT, 1 h, 37°C) and alkylated (25 mM IAM, 45 min, RT) in darkness. After that, 100 mM TEAB was used to dilute the protein sample to reduce the concentration of urea to under 2M. Trypsin was added (1:50 trypsin-to-protein mass ratio) for the first digestion at 37°C overnight and the second digestion (1:100 trypsin-to-protein mass ratio) was carried out over 4 h. Finally, the peptide was desalted by a Strata X C18 SPE column (Phenomenex) and vacuum-dried.

TMT labeling and HPLC fractionation

According to the TMT kit manufacturer's protocol, the tryptic peptides were firstly dissolved in 0.5 M TEAB. Each channel of peptide was labeled with their respective TMT reagent (based on manufacturer's protocol, Thermo Scientific), and incubated for 2 hours at room temperature. Samples were quenched by adding 5% hydroxylamine. The pooled samples were then desalted with Strata X C18 SPE column (Phenomenex) and dried by vacuum centrifugation. The sample was fractionated by high pH reverse-phase HPLC (Agilent 300Extend C18 column, 5 μm particles, 4.6 mm ID, 250 mm length). Then, the peptides were combined into 8 fractions and dried by vacuum centrifuging.

Affinity enrichment of modified peptides

To enrich Ksucc modified peptides, tryptic peptides dissolved in NETN buffer (100 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl, 0.5% NP-40, pH 8.0) were incubated with pre-washed succinylated antibody beads (Lot number PTM-402, PTM Bio) at 4°C overnight with gentle shaking. Then the beads were washed for four times with NETN buffer and twice with ddH₂O. The bound peptides were eluted from the beads with 0.1% TFA. Finally, the eluted fractions were combined and vacuum-dried. For LC-MS/MS analysis, the resulting peptides were desalted with C18 ZipTips (Millipore) according to the manufacturer's instructions.

LC-MS/MS Analysis

The tryptic peptides were dissolved in solvent A (0.1% formic acid, 2% acetonitrile/water), directly loaded onto a home-made reversed-phase analytical column (25 cm length, 75 μ m i.d.). Peptides were separated with a gradient from 6% to 22% solvent B (0.1% formic acid in 90% acetonitrile) over 24 min, 22% to 40% in 8 min and climbing to 80% in 5 min then holding at 80% for the last 3 min, all at a constant flowrate of 300 nL/min on an EASY-nLC 1000 UPLC system (Thermo Fisher Scientific).

The separated peptides were analyzed in Q ExactiveTM Plus (Thermo Fisher Scientific) with a nano-electrospray ion source. The electrospray voltage applied was 4.0 kV. The full MS scan resolution was set to 70,000 for a scan range of 350–1800 m/z. Up to 20 most abundant precursors were then selected for further MS/MS analyses with 15 s dynamic exclusion. The HCD fragmentation was performed at a normalized collision energy (NCE) of 28%. The fragments were detected in the Orbitrap at a resolution of 17,500. Fixed first mass was set as 100 m/z. Automatic gain control (AGC) target was set at 5E4, with an intensity threshold of 5E3 and a maximum injection time of 200 ms.

Database Search

The resulting MS/MS data were processed using MaxQuant search engine. Tandem mass spectra were searched against the Swissprot_Mouse.fasta concatenated with reverse decoy database. Trypsin/P was specified as cleavage enzyme allowing up to 4 missing cleavages. The mass tolerance for precursor ions was set as 20 ppm in First search and 5 ppm in Main search, and the mass tolerance for fragment ions was set as 0.02 Da. Carbamidomethyl on Cys was specified as fixed modification. Acetylation on protein N-terminal, oxidation on Met and succinylation on Lys were specified as variable modifications. TMT-6plex quantification was performed. FDR was adjusted to < 1% and minimum score for peptides was set > 40.

DNA extraction, library construction, and metagenomic sequencing

Total genomic DNA was extracted from 8 week-old SPF mice(n=6) fecal samples using the E.Z.N.A.[®] Soil DNA Kit (Omega Bio-tek, Norcross, GA, U.S.) according to manufacturer's instructions. Concentration and purity of extracted DNA was determined with TBS-380 and NanoDrop2000, respectively. DNA extract quality was checked on 1% agarose gel.

DNA extract was fragmented to an average size of about 400 bp using Covaris M220 (Gene Company Limited, China) for paired-end library construction. Paired-end library was constructed using NEXTFLEX[®] Rapid DNA-Seq (Bioo Scientific, Austin, TX, USA). Adapters containing the full complement of sequencing primer hybridization sites were ligated to the blunt-end of fragments. Paired-end sequencing was performed on Illumina Novaseq6000 (Illumina Inc., San Diego, CA, USA) at Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China) using NovaSeq Reagent Kits/HiSeq X Reagent Kits according to the manufacturer's instructions (www.illumina.com).

Sequence quality control and genome assembly

The data were analyzed on the free online platform of Majorbio Cloud Platform (www.majorbio.com). Briefly, the paired-end Illumina reads were trimmed of adaptors, and low-quality reads (length < 50 bp or with a quality value < 20 or having N bases) were removed by fastp (Chen et al., 2018) (<https://github.com/OpenGene/fastp>, version 0.20.0).

Reads were aligned to the mice genome by BWA (Li et al., 2009) (<http://bio-bwa.sourceforge.net>, version 0.7.9a) and any hit associated with the reads and their mated reads were removed.

Metagenomics data were assembled using MEGAHIT (Li et al., 2015) (<https://github.com/voutcn/megahit>, version 1.1.2), which makes use of succinct de Bruijn graphs. Contigs with a length ≥ 300 bp were selected as the final assembling result, and then the contigs were used for further gene prediction and annotation.

Gene prediction, taxonomy, and functional annotation

Open reading frames (ORFs) from each assembled contig were predicted using Prodigal (Hyatt et al., 2010) / MetaGene (Noguchi et al., 2006) (<http://metagene.cb.k.u-tokyo.ac.jp/>). The predicted ORFs with a length ≥ 100 bp were retrieved and translated into amino acid sequences using the NCBI translation table (<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=tgencodes#SG1>).

A non-redundant gene catalog was constructed using CD-HIT (Fu et al., 2012) (<http://www.bioinformatics.org/cd-hit/>, version 4.6.1) with 90% sequence identity and 90% coverage. High-quality reads were aligned to the non-redundant gene catalogs to calculate gene abundance with 95% identity using SOAPaligner (Li et al., 2008) (<http://soap.genomics.org.cn/>, version 2.21).

Representative sequences of non-redundant gene catalog were aligned to NR database with an e-value cutoff of $1e-5$ using Diamond (Buchfink et al., 2015) (<http://www.diamondsearch.org/index.php>, version 0.8.35) for taxonomic annotations. Cluster of orthologous groups of proteins (COG) annotation for the representative sequences was performed using Diamond (Buchfink et al., 2015) (<http://www.diamondsearch.org/index.php>, version 0.8.35) against eggNOG database with an e-value cutoff of $1e-5$. The KEGG annotation was conducted using Diamond (Buchfink et al., 2015) (<http://www.diamondsearch.org/index.php>, version 0.8.35) against the Kyoto Encyclopedia of Genes and Genomes database (<http://www.genome.jp/kegg/>) with an e-value cutoff of $1e-5$.

REFERENCES

- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature methods*, 12(1), 59–60.
- Chen SF, Zhou YQ, Chen YR, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890.
- Fu LM, Niu BF, Zhu ZW, et al. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–3152.
- Hyatt D, Chen GL, Locascio PF, et al. 2010. Prodigal: prokaryotic gene recognition

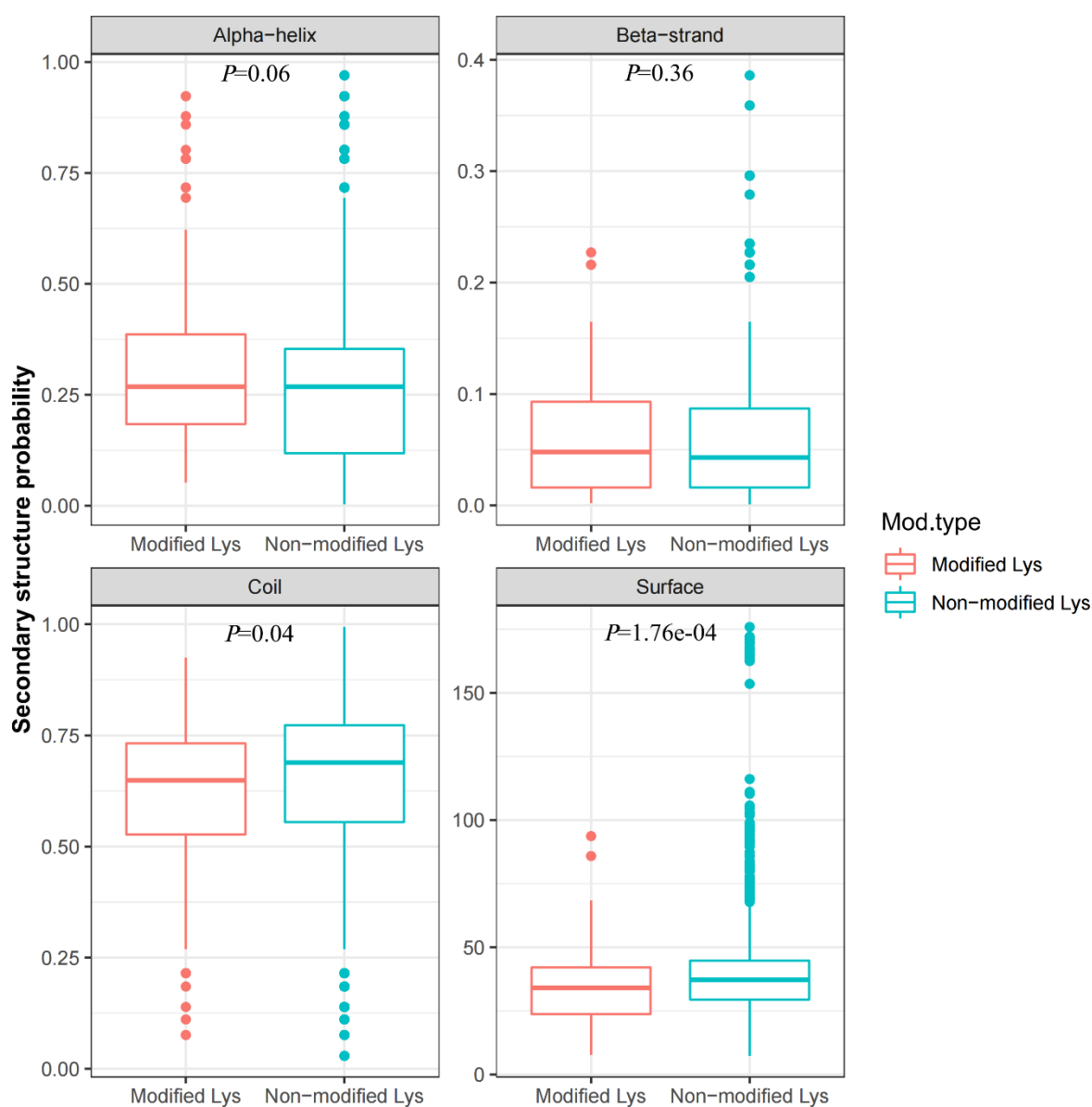
and translation initiation site identification. *BMC bioinformatics*, 11, 119.

Li DH, Liu CM, Luo RB, et al. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10), 1674–1676.

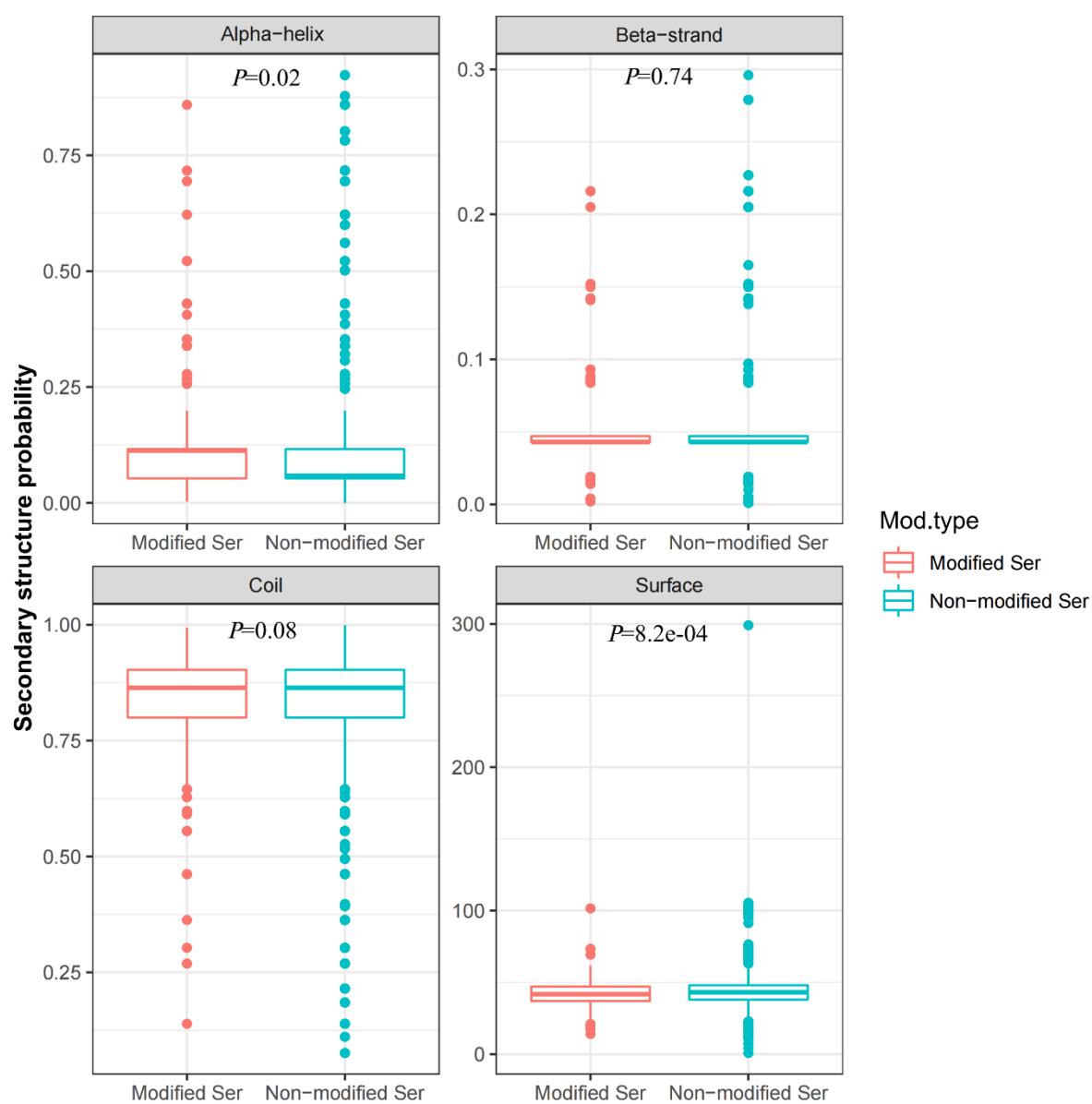
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.

Li RQ, Li YR, Kristiansen K, et al. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics*, 24(5), 713-714.

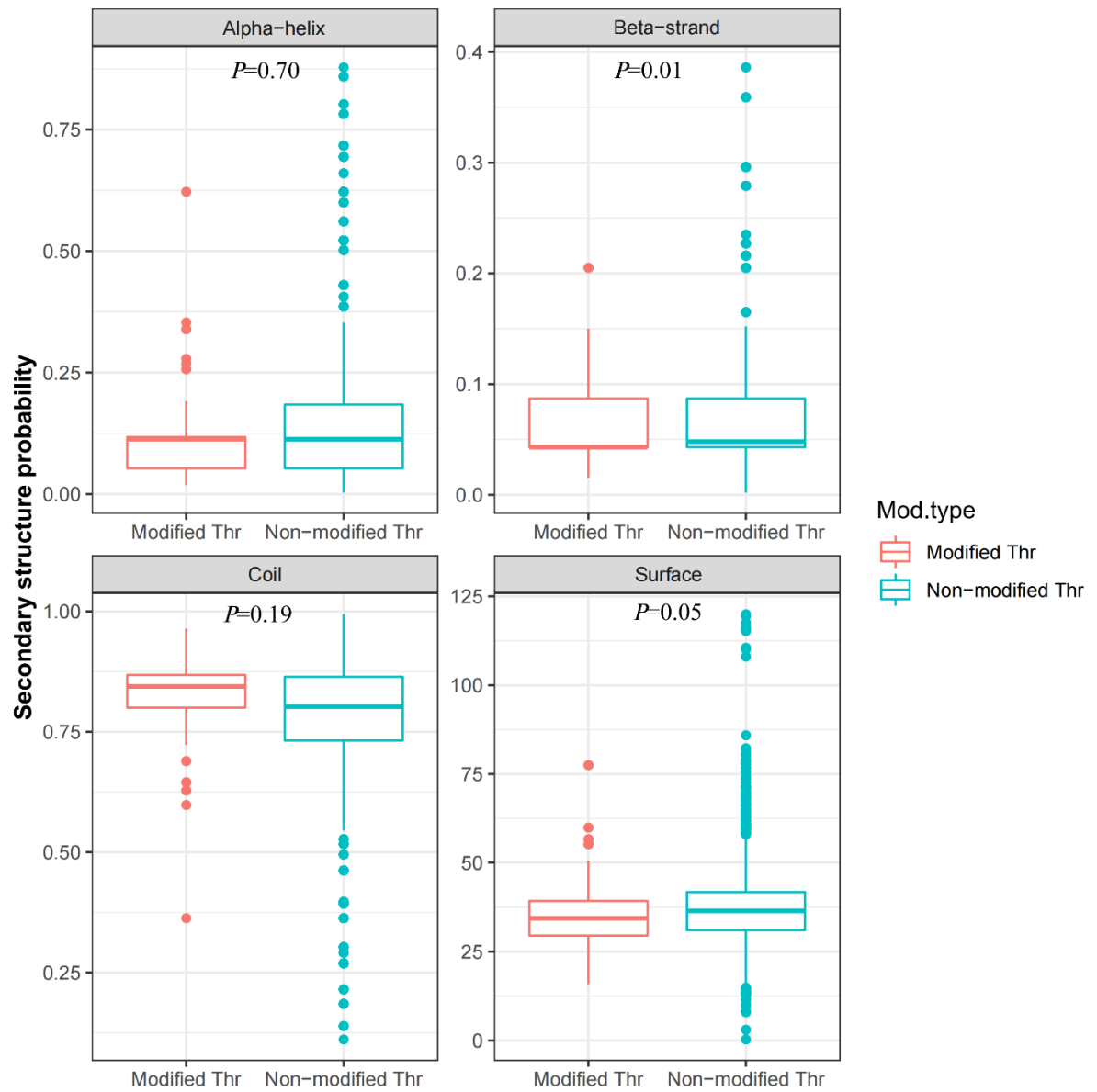
Noguchi H, Park J, Takagi T. 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res*, 34(19), 5623-5630.



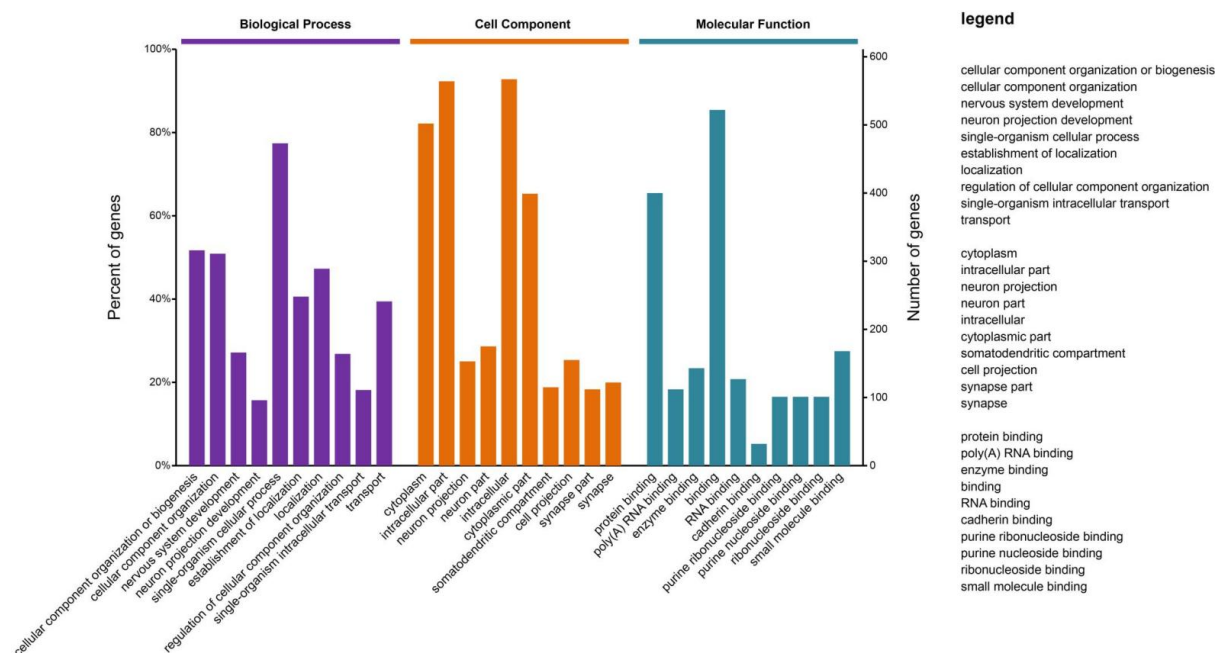
Supplementary Figure S1 Secondary structure distribution and surface accessibility prediction of acetylation sites



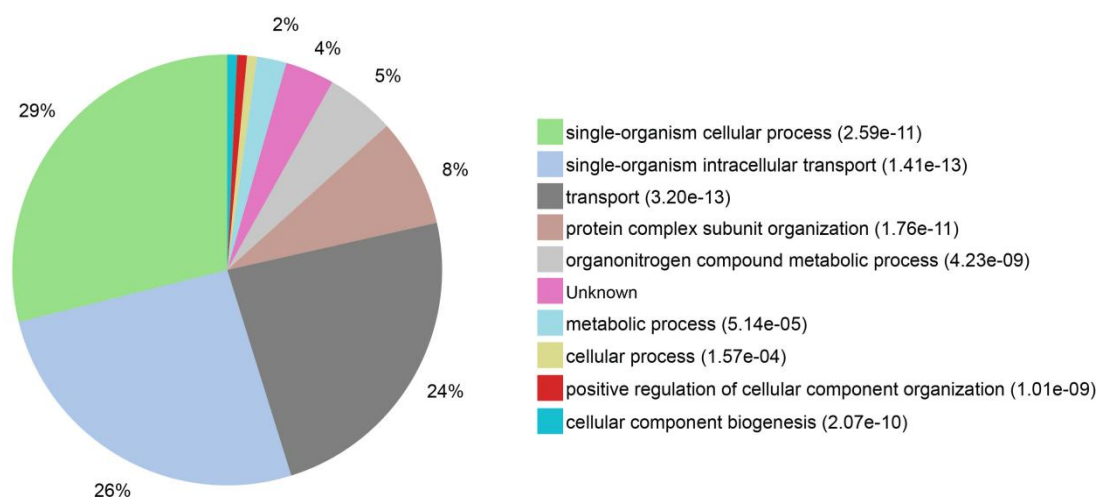
Supplementary Figure S2 Secondary structure distribution and surface accessibility prediction of serine phosphorylation sites



Supplementary Figure S3 Secondary structure distribution and surface accessibility prediction of threonine phosphorylation sites



Supplementary Figure S4 GO annotation analysis of differential succinylated proteins



Supplementary Figure S5 Details of cellular component enrichment based on GO annotation analysis of differential succinylated proteins

Supplementary Tables

Supplementary Table S1 Differential hippocampal succinylation sites between GF and SPF mice

Supplementary Table S2 Differential hippocampal phosphorylation sites between GF and SPF mice

Supplementary Table S3 Differential hippocampal acetylation sites between GF and SPF mice

Supplementary Table S4 Motif analysis of all succinylation sites

Supplementary Table S5 KEGG enrichment analysis of differential succinylated proteins

Supplementary Table S6 KEGG enrichment analysis of differential acetylated proteins

Supplementary Table S7 KEGG enrichment analysis of differential phosphorylated proteins

Supplementary Table S8 Distinct and shared/overlapping signaling pathways in different PTM proteins

Supplementary Table S9 Species distribution of NR annotations in SPF mice

Supplementary Tables S1–S9 are listed in a separate Excel file due to their large size.