

Supplementary materials

STEP: Profiling cellular-specific targets and pathways of bioactive small molecules in tissues via integrating single-cell transcriptomics and chemoproteomics

Jiayun Chen ^{a, #}, Zheng Chu ^{a, #}, Qian Zhang ^{b, #}, Chen Wang ^a, Piao Luo ^b, Ying Zhang ^a, Fei Xia ^a, Liwei Gu ^a, Yin Kwan Wong ^c, Qiaoli Shi ^a, Chengchao Xu ^{a*}, Huan Tang ^{a*}, Jigang Wang ^{a, b, c, d*}

^a State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Artemisinin Research Center, and Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

^b School of Traditional Chinese Medicine, Southern Medical University, Guangzhou 510515, China

^c Department of Nephrology, Shenzhen key Laboratory of Kidney Diseases, and Shenzhen Clinical Research Centre for Geriatrics, Shenzhen People's Hospital, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen 518020, China

^d State key Laboratory of Antiviral Drugs, School of Pharmacy, Henan University, Kaifeng 475004, China

[#]Jiayun Chen, Zheng Chu, and Qian Zhang contributed to this work equally.

*To whom correspondence should be addressed. Email: ccxu@icmm.ac.cn; htang@icmm.ac.cn; jgwang@icmm.ac.cn.

* Corresponding author:

Jigang Wang (*1st corresponding author)

E-mail addresses: jgwang@icmm.ac.cn (J. Wang);

Huan Tang (*2nd corresponding author)

E-mail addresses: htang@icmm.ac.cn (H. Tang);

Chengchao Xu (*3rd corresponding author)

E-mail addresses: ccxu@icmm.ac.cn (C. Xu)

1. Supplementary figures

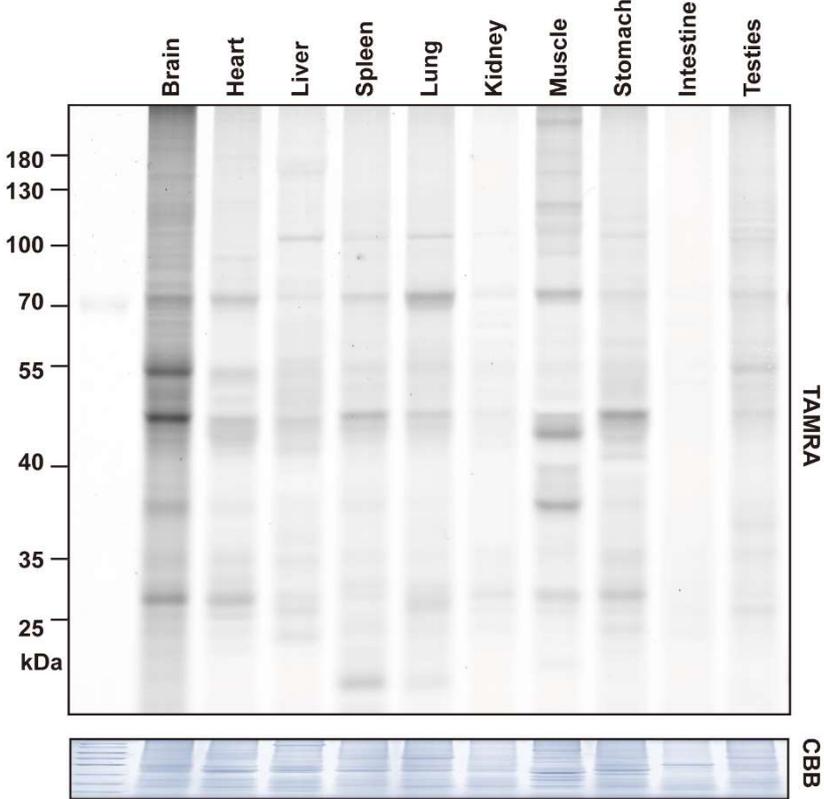


Figure S1. *In vitro* labeling intensity in murine brain and other 9 tissues lysis by aspirin probe (100 μ M). TAMRA, 5-carboxytetramethylrhodamine fluorescence; CBB, Coomassie brilliant blue.

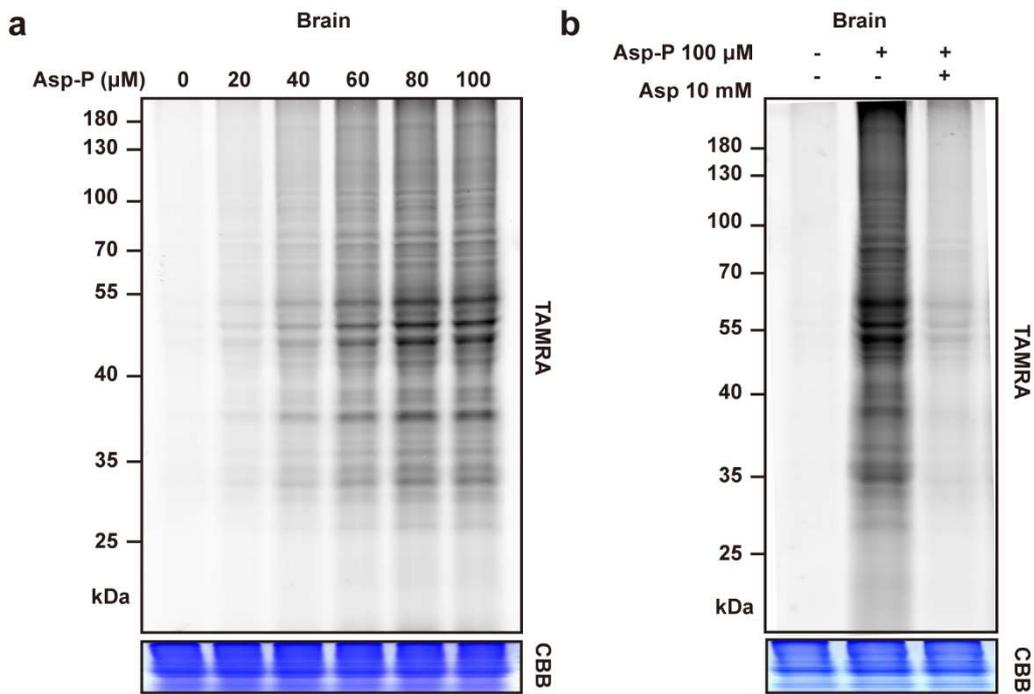


Figure S2. **a)** The fluorescent labeling in the brain tissue lysis by aspirin probe. **b)** The competition of *in vitro* protein labeling with Asp probe by Asp in mouse healthy brain lysate. TAMRA, 5-carboxytetramethylrhodamine fluorescence; CBB, Coomassie brilliant blue.

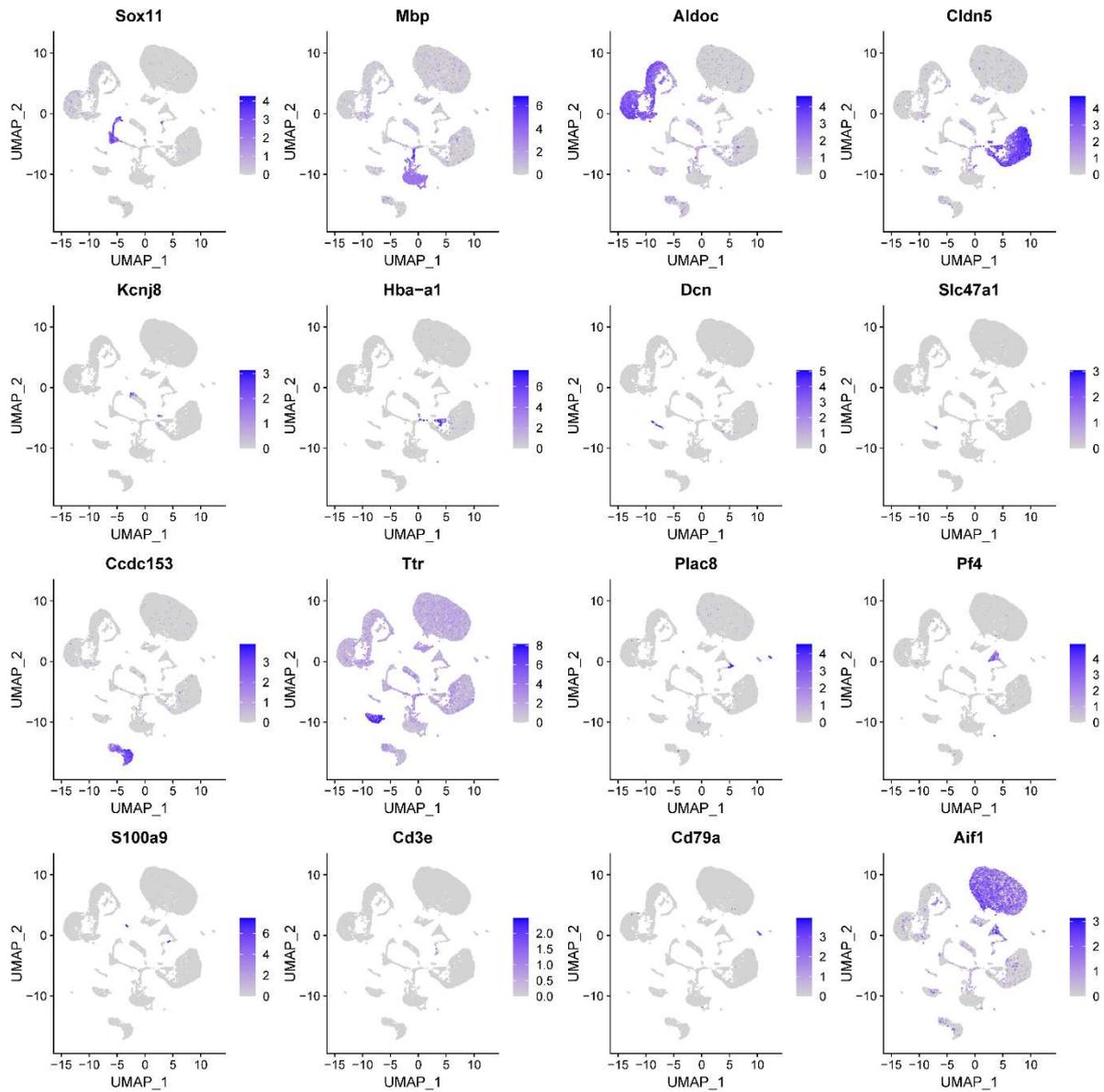


Figure S3. UMAP (Uniform Manifold Approximation and Projection) visualization showing the expression level of canonical markers for each cell type in the scRNA-seq dataset of healthy murine brain.

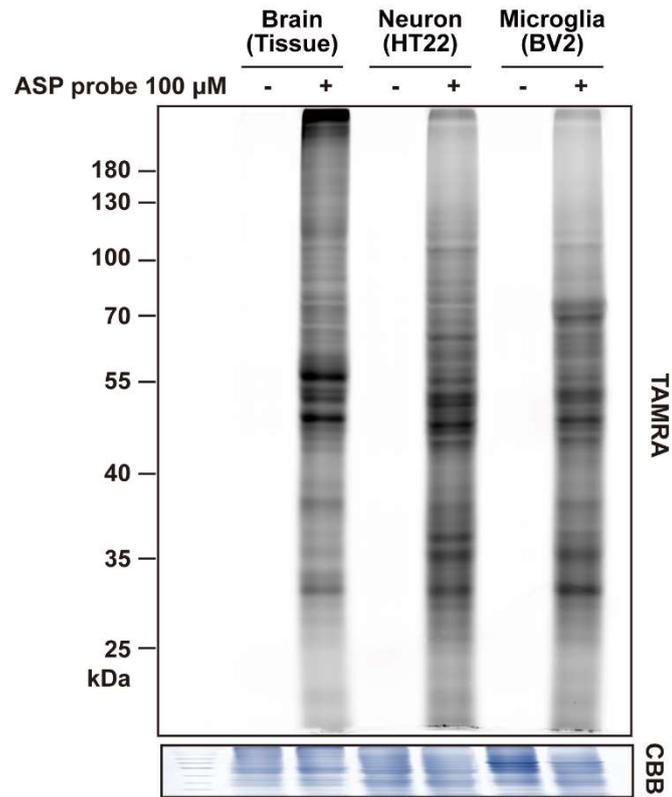


Figure S4. *In vitro* labeling in healthy mice brain tissue, HT-22 mouse neuron cell line, and BV2 mouse microglia cell line by aspirin probe. TAMRA, 5-carboxytetramethylrhodamine fluorescence; CBB, Coomassie brilliant blue.

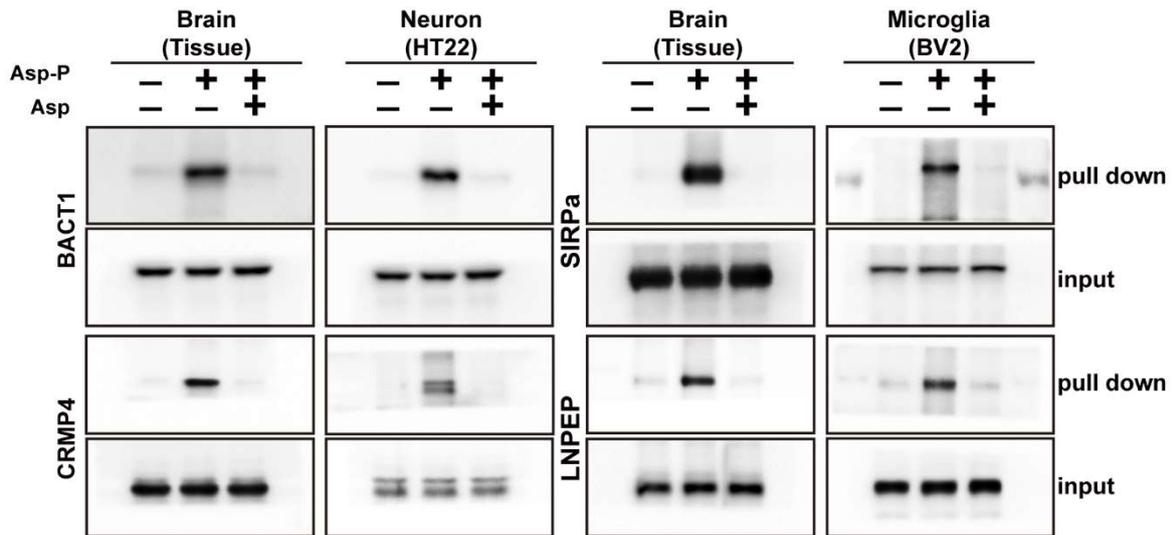


Figure S5. Pull-down western blotting experiments verifying aspirin directly targeting to BCAT1, and CRMP4 proteins in healthy brain tissue and neuron cell line, while also directly targeting to SIRPa and LNPEP proteins in healthy brain tissue and microglia cell line.

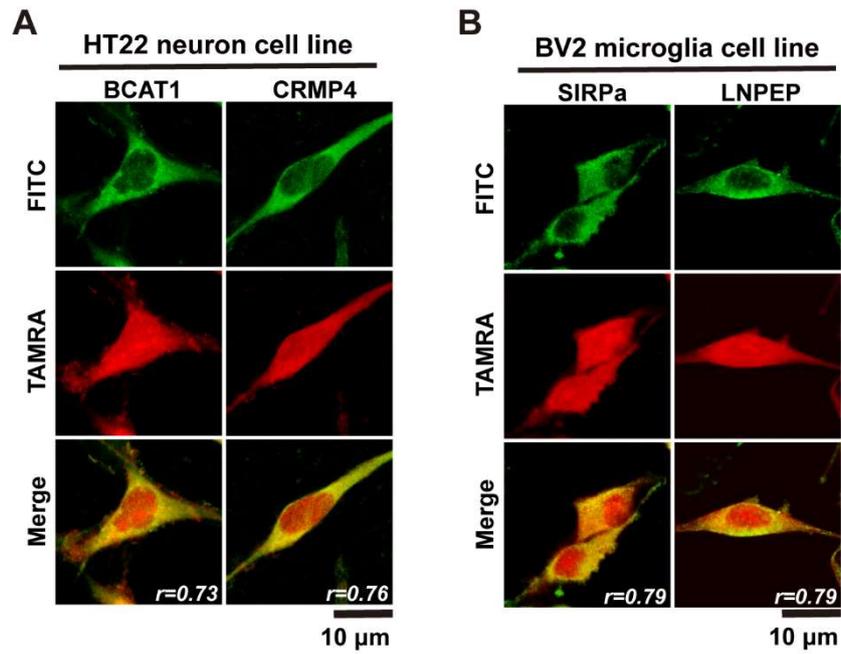


Figure S6. Fluorescent cellular imaging to track the subcellular location of the aspirin probe in HT22 neuron cell (A) and BV2 microglia cell (B). Fluorescence staining of target proteins (FITC, green) and aspirin probe (TAMRA, red), the scale bar (white) represents 10 μ m and the r value represents the Pearson's coefficient.

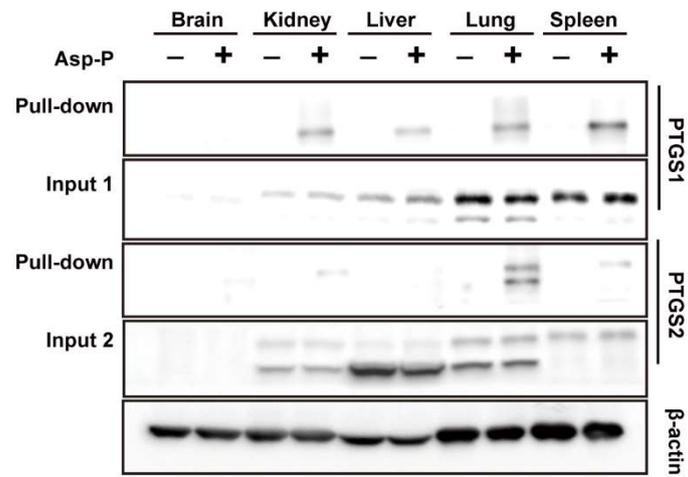


Figure S7. Pull-down western blotting experiments verifying aspirin directly targeting to PTGS1 and PTGS2 proteins in mouse brain, kidney, liver, lung, spleen tissues. Asp-P: aspirin probe.

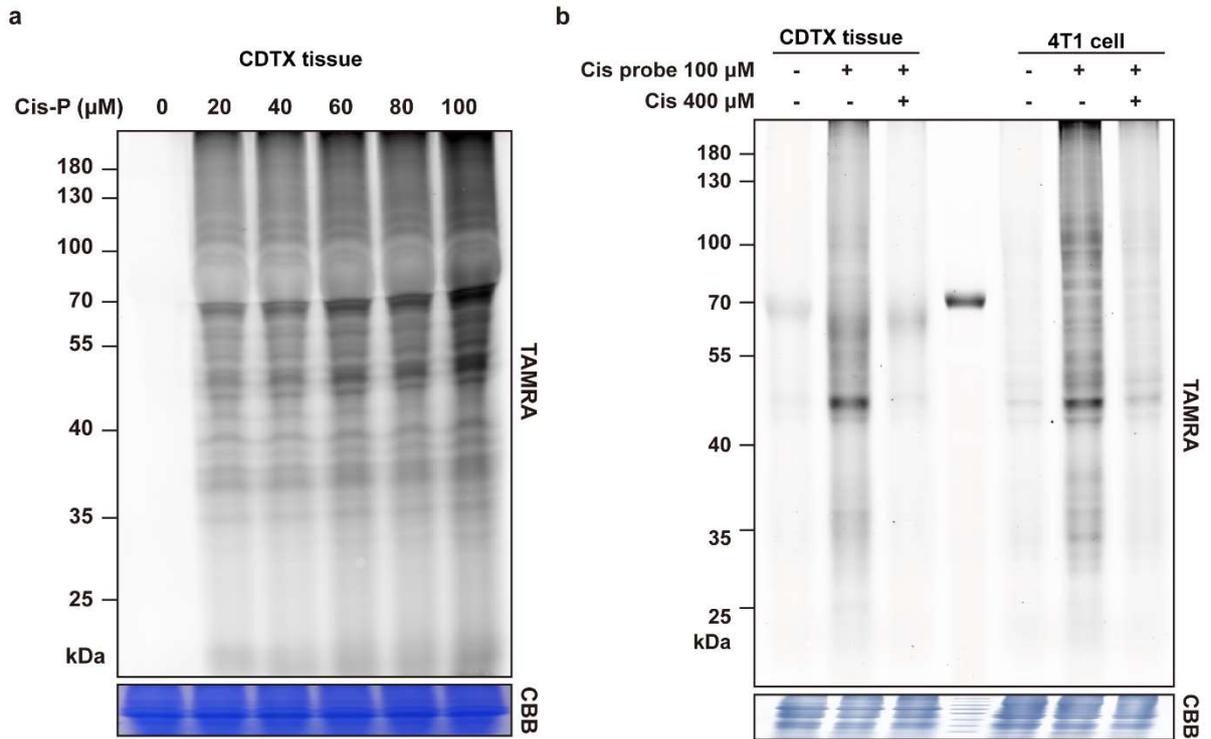


Figure S9. a) The fluorescent labeling in the CDTX tissue lysis by cisplatin probe. **b)** *In vitro* labelling in CDTX tissue lysis (left panel) and 4T1 tumor cell line lysis (right panel) by cisplatin probe. Flu, fluorescence; CBB, Coomassie brilliant blue.

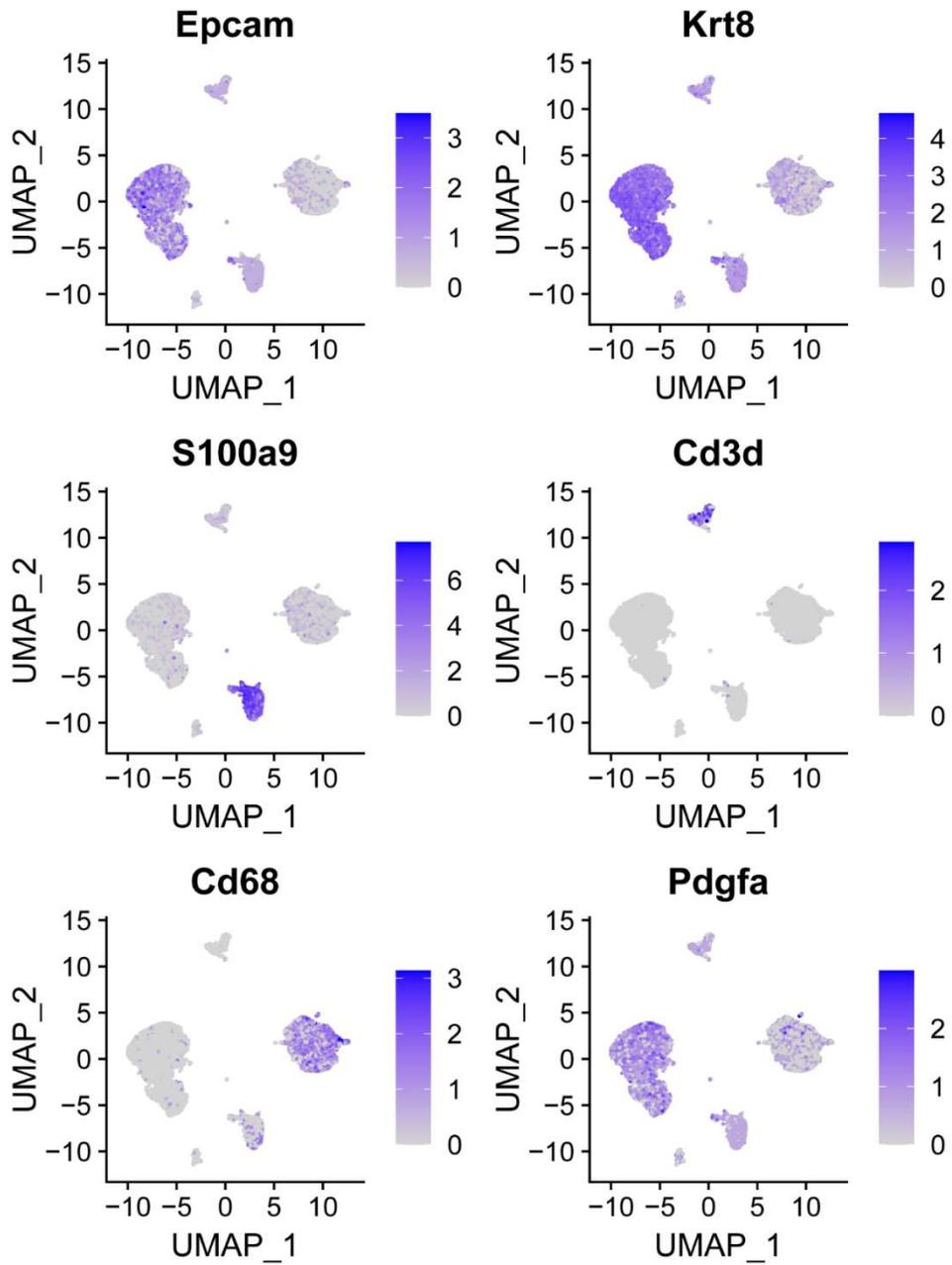


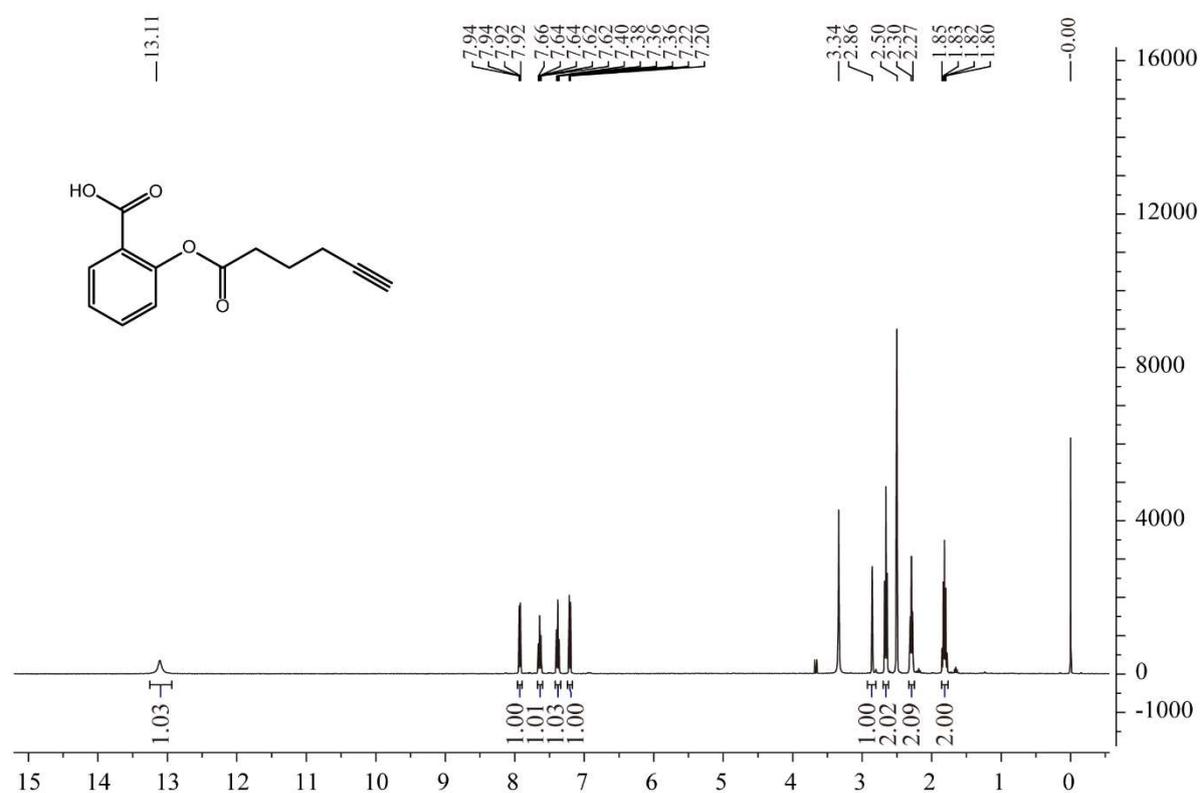
Figure S10. UMAP visualization showing the expression level of canonical markers for each cell type in the scRNA-seq dataset of cell-line-derived tumor xenograft (CDTX).

2. Methods and materials

Experimental section

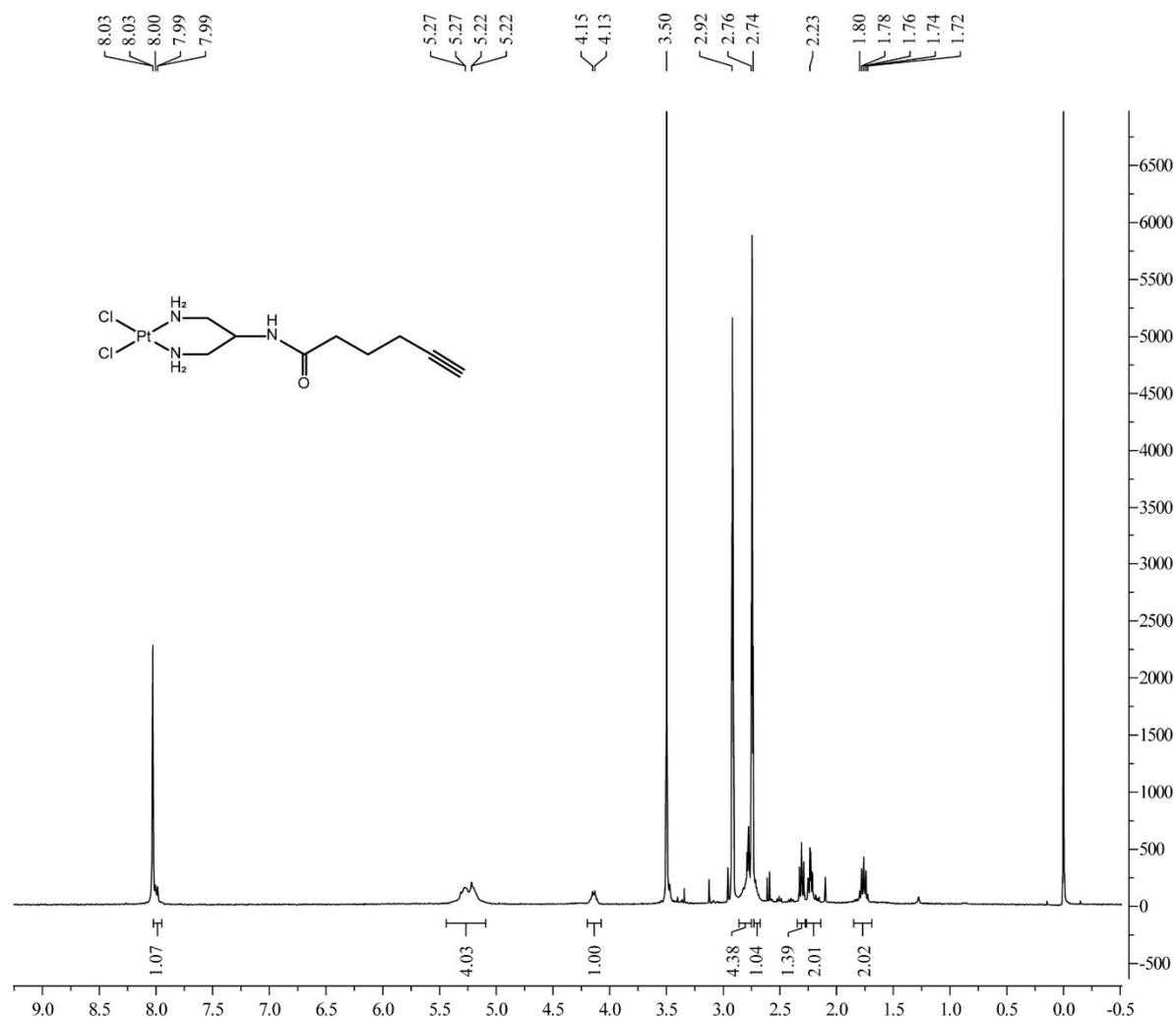
Chemical synthesis of probes

The probes used in our research were designed and synthesized as previously reported [1-3]. All the reagents and solvents were purchased from Sigma-Aldrich and AK Scientific used without further purification unless stated otherwise. Reactions were monitored by thin layer chromatography (TLC). Column chromatography was performed on silica gel 200~300 mesh. All ^1H NMR spectra were recorded. ^1H NMR Spectroscopy splitting patterns were designated as singlet (s), doublet (d), triplet (t), quartet (q). Splitting patterns that could not be interpreted or easily visualized were designated as multiplet (m).



(1) Aspirin Probe

^1H NMR of Aspirin Probe ^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ 13.11 (s, 1H), 7.93 (dd, $J = 7.8$, 1.7 Hz, 1H), 7.64 (td, $J = 7.8$, 1.7 Hz, 1H), 7.38 (td, $J = 7.6$, 1.1 Hz, 1H), 7.21 (dd, $J = 8.1$, 1.0 Hz, 1H), 2.85 (t, $J = 2.6$ Hz, 1H), 2.66 (t, $J = 7.4$ Hz, 2H), 2.29 (td, $J = 7.1$, 2.6 Hz, 2H), 1.83 (dd, $J = 14.5$, 7.2 Hz, 2H).



(3) Cisplatin probe

¹H NMR of Cisplatin Probe ¹H NMR (400 MHz, DMF-d₇): δ 8.02 – 7.95 (m, 1H), 5.44 – 5.09 (m, 4H), 4.14 (d, J = 7.7 Hz, 1H), 2.86 – 2.75 (m, 4H), 2.72 (m, 1H), 2.31 (t, J = 7.5 Hz, 2H), 2.23 (td, J = 7.1, 2.6 Hz, 2H), 1.76 (dt, J = 14.4, 7.2 Hz, 2H).

Materials

The following materials were obtained from commercial sources: dulbecco's modified eagle medium (DMEM, Gibco #11995-065), penicillin-streptomycin (Gibco #15140-122), fetal bovine serum (FBS, ExCell Bio #FSP500), protease inhibitor cocktail (Thermo Scientific #78429), NeutrAvidin Agarose beads (Thermo Scientific #29202), 5-TAMRA-PEG3-azide (Chomix), Biotin-PEG3-azide (Chomix), Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA, TCI #T2993), Tris (2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma #C4706), DL-Dithiothreitol (DTT, Sigma D916), iodoacetamide (IAA, Sigma #I6125), CuSO₄ (Sigma #61230), TMT 10plex Mass Tag reagents (Thermo Scientific #90110).

Cell culture

HT-22 (mouse neuronal cell), BV2 (mouse microglial cell), 4T1 (mouse breast cancer cell) were maintained in DMEM, supplemented with 1% (v/v) Penicillin-Streptomycin, and 10% (v/v)

FBS. Cells were maintained at 37 °C with 5% CO₂ humidified atmosphere.

Animal experiments

All animal experimental procedures were approved by the China Animal Care and Use Committee and the Care and Use of Laboratory Animals of China Academy of Chinese Medical Sciences (MDKN-2022-015). The brain, kidney and other organ tissues were collected for further usage from C57 BL/6 wild-type mice (male, 20 ± 2g, 6-8 weeks old) after anesthetization. The cell-line-derived tumor xenograft (CDTX) tissues were collected from BALB/c wild-type mice (female, 20 ± 2g, 6-8 weeks old), after injected with 10⁶ 4T1 cells to form breast carcinoma. The tumor was passaged for two generations *in situ* (21 d as one generation), and the tumors tissues were then collected after anesthetization.

Chemoproteomics experiment and analysis

Sample preparation for in-gel fluorescence characterization

The collected tissues and cells were homogenized with Dounce Homogenizer in 0.1% (v/v) Triton/ PBS containing 1% (v/v) protease inhibitor cocktail, and then the samples were lysed on ice with sonication for 5 min at 2 s: 2 s duty cycle and 25% power. After centrifugation for 30 min at 4°C, supernatants were collected and the concentration of proteins was determined using a BCA assay. A total 100 µL of protein at 2 mg/mL was then incubated with the alkylated probes (final concentration of 100 µM) for 2 h at room temperature with or without pretreated with the competitor (final concentration of 400 µM) for 2 h. Probe-labeled samples were conjugated to rhodamine-azide by copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) as the literature reported^[2, 4]. After the click chemistry step, the proteins were precipitated by prechilled acetone (-20°C), then centrifuged (20,000 g, 5 min, 4°C) to pellet the precipitated proteins. The pellets were dissolved in 1×SDS-PAGE loading buffer containing DL-Dithiothreitol (DTT) via sonication and heating (5 min, 95°C). The samples were finally subjected to SDS-PAGE analysis and in-gel fluorescence scanning using Azure biosystem.

Sample preparation for cellular imaging

HT22 and BV2 cells were grown in 4-chamber glass bottom dishes and incubated with 0.4 mL DMEM containing 1 mM ASP probe. After 12 h, the cells were washed and then fixed with 4% paraformaldehyde at room temperature for 15 min, and further permeabilized in 0.2% Triton X-100 for another 15 min. As former mentioned, the cells were conjugated to rhodamine-azide by CuAAC. After being washed 3 times with PBS, the cells were then blocked with 5% bovine albumin (BSA) for 2 h, followed by incubating with primary antibodies (1:200) at 4 °C overnight. Subsequently, cells were gently washed with 0.1% Tween 20 in Tris buffered saline (TBST) for 3 times and then incubated with secondary fluorescent antibody (FITC goat anti-rabbit IgG, Abclonal) for 1 h. Finally, samples were incubated with Hoechst (1:5000) at room temperature for 30 min. All images were acquired with Leica TCS SP8 SR confocal fluorescence microscopy and the colocalization Pearson's coefficient were obtained with LAS X software.

Sample preparation for proteome analysis

The preparation of proteomic samples was performed according to the previous literature with slight modification^[1, 2, 4]. Briefly, the protein from tissues and cells samples were lysed and labeled similarly as mentioned above. Afterward, probe labeled proteins were conjugated to biotin-azide by CuAAC as previously described^[2, 4]. The mixtures were then incubated for 1 h at room temperature, and excess click chemistry reagents were removed with prechilled acetone (-20°C) through centrifugation (20,000 g, 5 min, 4°C). Subsequently, the airdried protein pellets were resuspended in 1.2% SDS/ PBS and the final concentration of SDS was diluted to 0.2% with PBS, then subjected to streptavidin enrichment of probe-labeled proteins as previously described^[5]. The enriched proteins on beads were followed by dithiothreitol reduction (10 mM, 30 min, 37°C) and iodoacetamide alkylation (25 mM, 30 min, 37°C), and then digested to peptides using trypsin (4 µg in 15 µL of ammonium bicarbonate, sequencing grade from Promega) for 17 h at 37°C. The digested peptides in the supernatant were collected and labeled with TMT 10plex Mass Tag reagents as previously described^[6]. Finally, the pooled samples were dried using a SpeedVac and stored at -80 °C until LC-MS/MS analysis.

LC-MS/MS analysis

The tryptic peptides were analyzed using an UltiMate 3000 nano-LC system coupled with an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific, USA). Dried peptide sample fractions were reconstituted in 10 µL of 0.1% formic acid and 1% acetonitrile aqueous solution, and transferred to a 96-well plate autosampler. 2 µL of each reconstituted fraction was loaded into an AcclaimPepMap100 C18 Nano-Trap Columns (100 Å, 3 µm, 75 mm × 2 cm) and washed with 0.5% formic acid and 2% acetonitrile at a flow rate of 5 µL/min for 5 min. The peptides were separated on an Acclaim PepMap100 C18 HPLC reversed-phase analytical column (130 Å, 2 µm, 75 µm × 250 mm). At a flow rate of 0.3 µL/min, the peptides were separated by a gradient formed by 0.1% formic acid aqueous solution (mobile phase A) and 0.1% formic acid in 80% acetonitrile (mobile phase B): 2%-4% mobile phase B (4 min), 4%-8% mobile phase B (4 min), 8%-35% mobile phase B (50 min), 35%-90% mobile phase B (4 min), 90% mobile phase B (5 min), 90%-2% mobile phase B (1 min), and 2% mobile phase B (7 min). The mass spectrometer was operated using Xcalibur software (version 4.2.4 SP1) in the data-dependent acquisition mode with an NSI (nanospray ion source) spray voltage of 2.5 kV. All MS spectra detection was performed in positive-ion mode with charge states of 2-6 and a full scan of MS spectrum from m/z 300-1500 at 60,000 resolutions. The top 20 most abundant precursors were subjected to HCD with a collision energy of 38% in the ion trap with a fixed first mass of 110 m/z and an isolation window of 0.7 m/z.

MS data process

Mass spectrometry (MS) raw files were processed by Proteome Discovery software integrated with Sequest HT search engine (version 2.4, Thermo Scientific). All the MS spectra were searched against the UniProtKB mus musculus FASTA database (version from January 2020) contaminated with reversed copies of all sequences. Search parameters were used as follows for all the pipelines: MS1 tolerance of 10 ppm; orbitrap-detected MS/MS mass tolerance of 0.02 Da; enzyme digestion specificity of trypsin; missed cleavages of up to 2; Minimum peptide length of seven amino acids; carbamidomethylation (C) as fixed modification; oxidation of methionine and acetyl (protein N-term) as variable modifications. Data were filtered to a 1%

false discovery rate (FDR) on peptide-spectrum matches estimated using the decoy hit distribution or the Percolator algorithm (version 3.02.1) based on the calculated q-values when specified. The TMT quantification for proteins was performed by Proteome Discovery software according to the manufacturer's specifications and the literature^[7].

Bioinformatic analysis

Chemoproteomics datasets generated from LC-MS/MS were subjected to statistical analysis and visualization using statistical software R (version 4.1.1). For statistical analysis of proteins detected in different samples, the bar plot and pie chart were generated by ggplot2 R package (version 3.3.5), and the venn plot was generated using VennDiagram R package (version 1.6.2). For targeted proteins analysis, we filtered the protein that more than half of samples in such group contain the NA values, otherwise we applied knnImputation function in DMwR2 R package (version 0.0.2) to fulfill the missing value. Next, the analysis was performed using the "limma" R package (version 3.48.3)^[8]. P values generated from the empirical Bayes test model and were adjusted using Benjamini–Hochberg (BH). The proteins with absolute fold change ≥ 1.2 and adjusted P value (FDR) < 0.05 were considered to be significant targeted proteins. Volcano plot was used to highlight the significant targeted proteins which based on $\log_2(\text{FC})$ and $-\log_{10}(\text{FDR})$ of proteins by ggplot2 R package (version 3.3.5), and the heatmap plot was generated using pheatmap R package (version 1.0.12).

Single cell RNA sequencing and analysis

Generation of single-cell suspensions

As for the brain tissue, the brain samples from 3 healthy mice were cut into 3~4 pieces and enzymatically digested with the Adult Brain Dissociation Kit (Miltenyi Biotec) for 30 min on gentle MACS Dissociator according to the manufacturer's protocol. As for the CDTX tissue, the CDTX sample from 3 mice were cut into 5 mm particles and enzymatically digested with the Tumor Dissociation Kit (Miltenyi Biotec) for about 40 minutes on gentle MACS Dissociator. The dissociated cells were next passed through a 70 mm and 40 mm cell strainer (BD Biosciences) in the PBS (Sigma-Aldrich), until uniform cell suspensions were obtained. Subsequently, the suspended cells were passed through cell strainers and centrifuged at 300 g at 4 °C for 10 minutes. RBCs were removed using Red Blood Cell Lysis Solution (Miltenyi Biotec). After washing twice with $1 \times$ PBS, the cell pellets were resuspended in PBS sorting buffer to prepare single-cell suspension.

Single-cell RNA sequencing (scRNA-seq)

All scRNA-Seq libraries were prepared using the Chromium Next GEM Single-cell 3' Kit v3.1 from 10x Genomics, following the manufacturer's instructions. In brief, single cells were diluted into a final concentration of 800–1200 cells/ μL as determined by TC20 cell counter (Bio-Rad). About 10^4 cells were captured in droplets to generate nanoliter-scale gel beads in emulsion (GEMs). GEMs were then reverse transcribed in applied biosystems (Thermo Fisher Scientific) programmed at 53 °C for 45 minutes and 85 °C for 5 minutes and were held at 4 °C. After reverse transcription and cell barcoding, emulsions were broken and cDNA was isolated and purified with Cleanup Mix containing DynaBeads and SPRIselect reagent (Thermo Fisher

Scientific), followed by PCR amplification. For scRNA-Seq library construction, amplified cDNA was fragmented and end repaired, double-sided size selected, and PCR amplified with sample indexing primers, successively. Libraries prepared according to the manufacturer's user guide were then purified and profiled for quality assessment. Single-cell RNA was sequenced by an Illumina Novaseq 6000 sequencer (Illumina) with paired-end 150 bp (PE150) reads.

Cell type annotation

In the brain and CDTX scRNA-seq datasets, the Seurat's FindAllMarkers function was conducted to find expressed markers of each cluster. Each cluster was identified and annotated according to the expression level of canonical cell type markers, as previous work reported^[9-11].

scRNA-seq and bulk RNA-seq dataset re-analysis

As for the aristolochic acid nephropathy (AAN) dataset, we retrieved and integrated these existing datasets including the AAI targeted protein lists, bulk RNA-seq (3 healthy and 3 AAN samples), and scRNA-seq datasets (including 3 healthy and 3 AAN samples) in our previous studies^[1, 12], so as to profile the cellular protein targets and mediated biological pathways engaged by AAI-probe.

Identification of the cellular protein targets

After we get the candidates in the protein targets list identified by chemoproteomics, we use the AverageExpression function in Seurat, to access the averaged feature expression matrix of these genes (the candidate protein targets) across all cell types in the scRNA-seq dataset. After that, we scale the averaged feature expression matrix by "row" to comparable the average feature expression value of all cell types. For a protein target, if one cell type has the largest scaled value than the other cell types, we then defined this protein target as belonging to this cell type. After we iterated over all candidate targets, we obtained the cell distribution information of these protein targets.

Profiling the cellular targeted biological function

Gene ontology (GO) analysis was performed using the "clusterProfiler" R package (version 3.18.1), according to the cellular protein target profiles. P values were generated from the Hypergeometric test model and were adjusted using BH. The biological process (BP) category was selected to represent the functional profiles, and visualized on the basis of the count of proteins enriched and the adjusted P-value (Q value) of <0.01. We use the compareCluster function to generate a comparison gene clusters functional profiles in the scRNA-seq dataset, and use cnetplot function to visualize the concordant or different functional pathways as well as the protein targets. Specially, for pathways of oxidoreductase activity (acting on the CH-CH group of donors) and oxidoreductase activity (the CH-OH group of donors), we performed Gene Set Enrichment Analysis (GSEA) based on the whole DEGs profile of the AAN vs. healthy groups in bulk RNA-seq. As for the renal cell types, we also evaluated the enrichment scores of these two pathways based on the DEGs profile of the AAN vs. healthy groups in PT cells and DLH cell, respectively.

References

- [1] Q. Zhang, P. Luo, J. Chen, C. Yang, F. Xia, J. Zhang, H. Tang, D. Liu, L. Gu, Q. Shi, X. He, T. Yang, J. Wang, *Int J Biol Sci* 2022, 18, 2003.
- [2] J. Wang, C. J. Zhang, J. Zhang, Y. He, Y. M. Lee, S. Chen, T. K. Lim, S. Ng, H. M. Shen, Q. Lin, *Sci Rep* 2015, 5, 7896.
- [3] J. D. White, L. E. Guzman, L. N. Zakharov, M. M. Haley, V. J. DeRose, *Angew Chem Int Ed Engl* 2015, 54, 1032.
- [4] J. Wang, C. J. Zhang, W. N. Chia, C. C. Loh, Z. Li, Y. M. Lee, Y. He, L. X. Yuan, T. K. Lim, M. Liu, C. X. Liew, Y. Q. Lee, J. Zhang, N. Lu, C. T. Lim, Z. C. Hua, B. Liu, H. M. Shen, K. S. Tan, Q. Lin, *Nat Commun* 2015, 6, 10111.
- [5] P. Luo, D. Liu, Q. Zhang, F. Yang, Y. K. Wong, F. Xia, J. Zhang, J. Chen, Y. Tian, C. Yang, L. Dai, H. M. Shen, J. Wang, *Acta Pharm Sin B* 2022, 12, 2300.
- [6] Y. Shiio, R. Aebersold, *Nat Protoc* 2006, 1, 139-145.
- [7] P. Gao, Y. Q. Liu, W. Xiao, F. Xia, J. Y. Chen, L. W. Gu, F. Yang, L. H. Zheng, J. Z. Zhang, Q. Zhang, Z. J. Li, Y. Q. Meng, Y. P. Zhu, H. Tang, Q. L. Shi, Q. Y. Guo, Y. Zhang, C. C. Xu, L. Y. Dai, J. G. Wang, *Mil Med Res* 2022, 9, 30.
- [8] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, *Nucleic Acids Res* 2015, 43, e47.
- [9] F. J. Garcia, N. Sun, H. Lee, B. Godlewski, H. Mathys, K. Galani, B. Zhou, X. Jiang, A. P. Ng, J. Mantero, L. H. Tsai, D. A. Bennett, M. Sahin, M. Kellis, M. Heiman, *Nature* 2022, 603, 893.
- [10] M. A. Wheeler, I. C. Clark, E. C. Tjon, Z. Li, S. Zandee, C. P. Couturier, B. R. Watson, G. Scalisi, S. Alkwai, V. Rothhammer, A. Rotem, J. A. Heyman, S. Thaploo, L. M. Sanmarco, J. Ragoussis, D. A. Weitz, K. Petrecca, J. R. Moffitt, B. Becher, J. P. Antel, A. Prat, F. J. Quintana, *Nature* 2020, 578, 593.
- [11] K. Xu, R. Wang, H. Xie, L. Hu, C. Wang, J. Xu, C. Zhu, Y. Liu, F. Gao, X. Li, C. Wang, J. Huang, W. Zhou, G. Zhou, Y. Shu, X. Guan, *Oncogenesis* 2021, 10, 66.
- [12] J. Chen, P. Luo, C. Wang, C. Yang, Y. Bai, X. He, Q. Zhang, J. Zhang, J. Yang, S. Wang, J. Wang, *JCI Insight* 2022, 7.

3. Uncropped WB data

Fig. 2g BCAT1

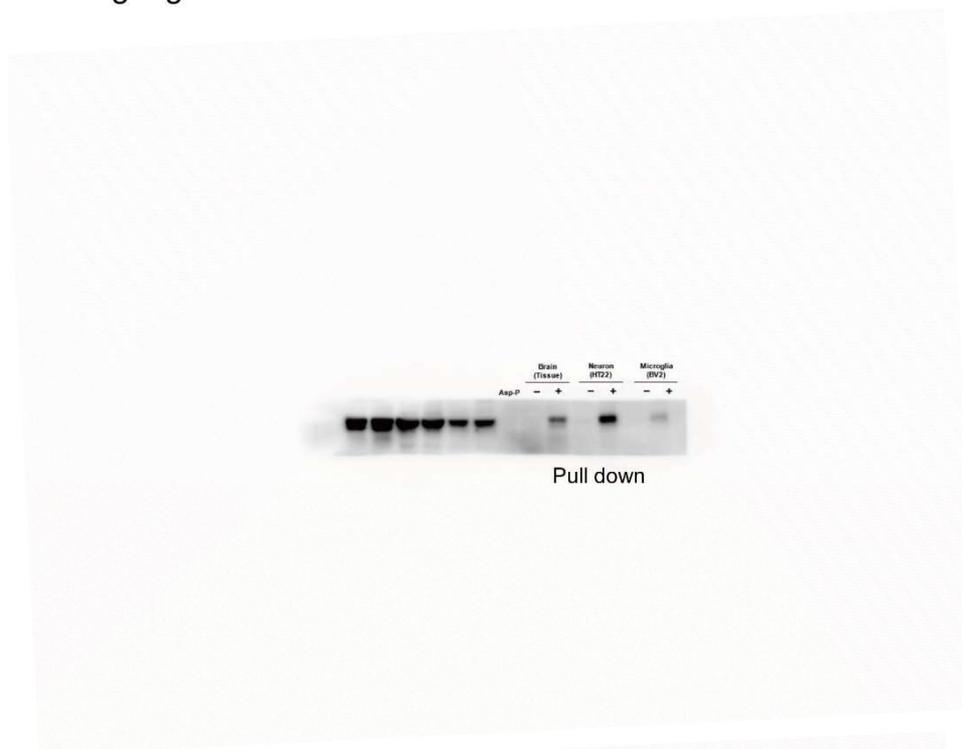


Fig. 2g CRMP4

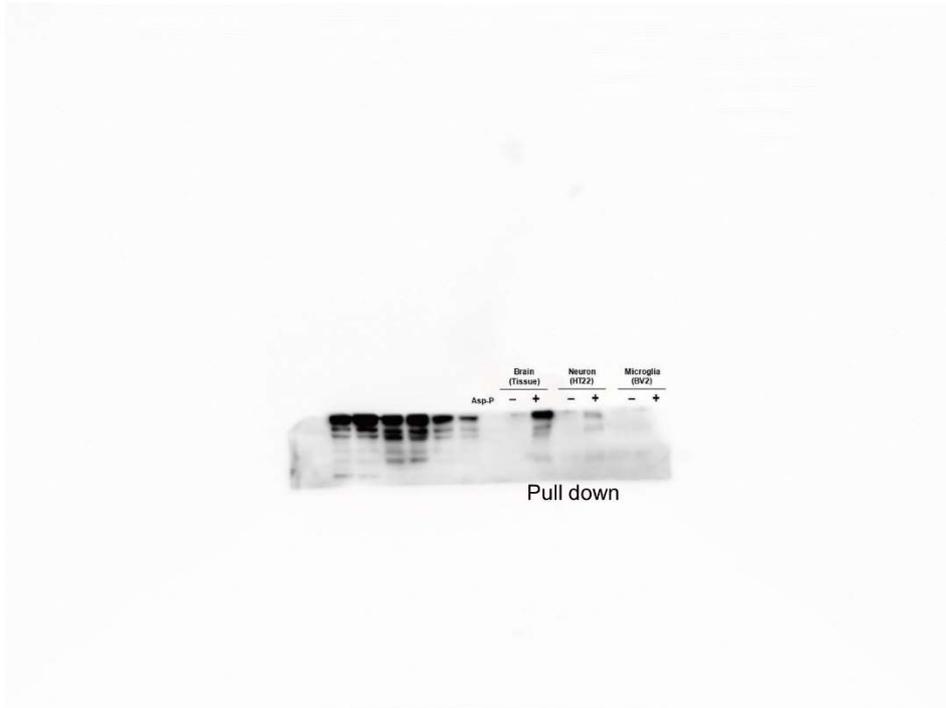


Fig. 2g LNPEP

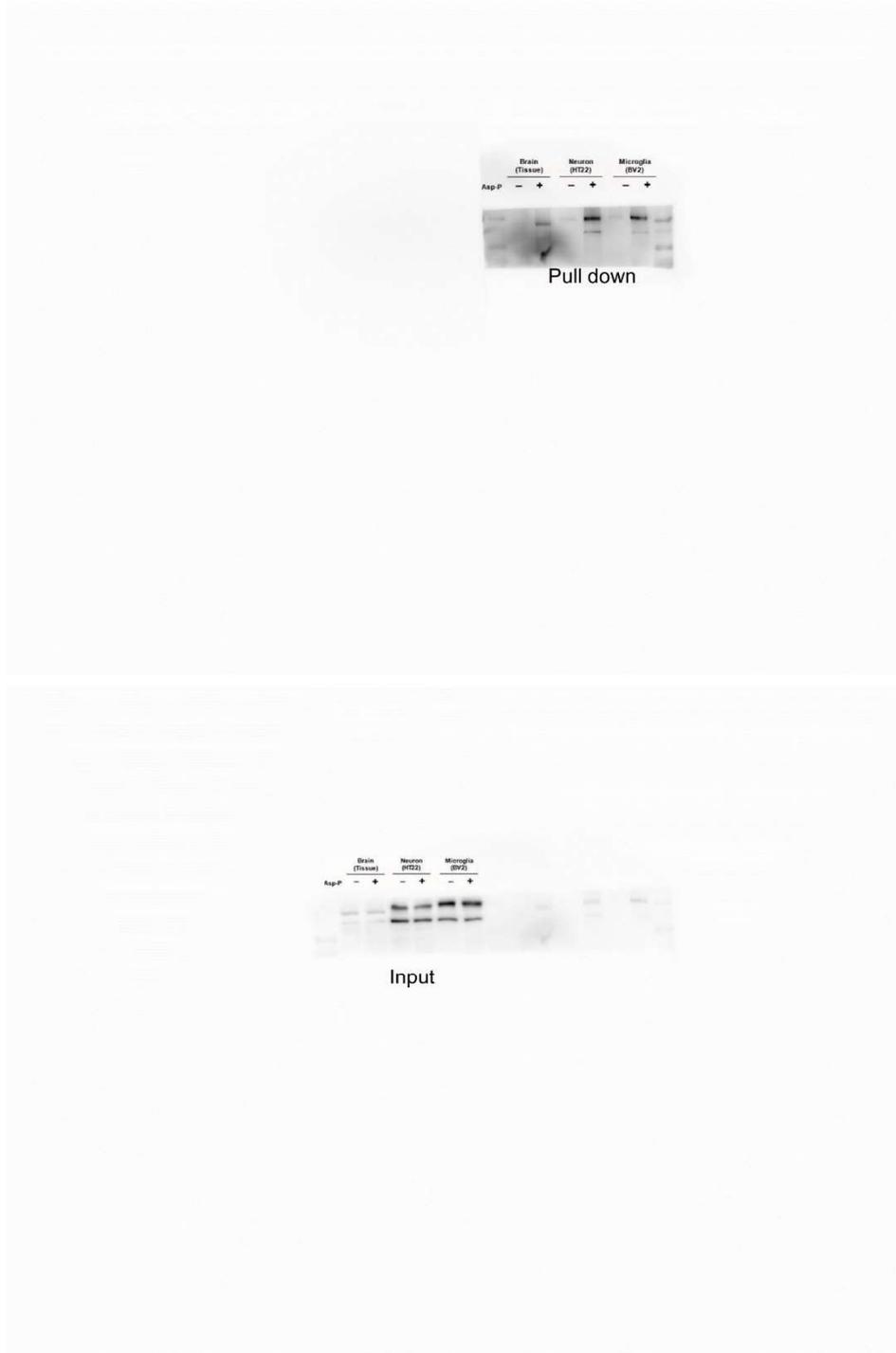
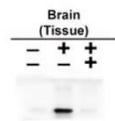
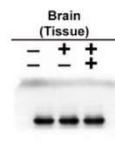


Fig. S5 CRMP4

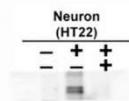
Pull down



Input



Pull down



Input

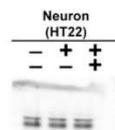
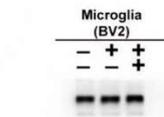
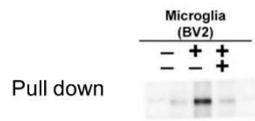
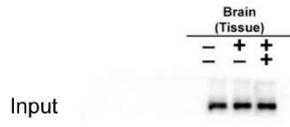
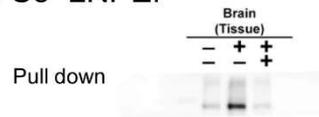


Fig. S5 LNPEP



Input

Fig. S5 SIRPa

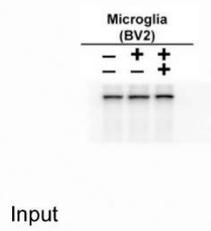
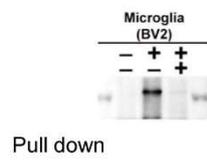
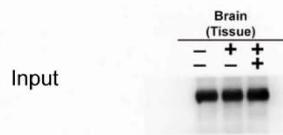


Fig. S7 PTGS1

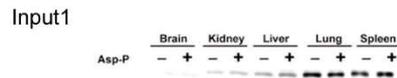
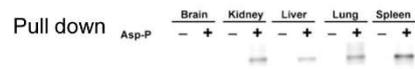


Fig. S7 PTGS2

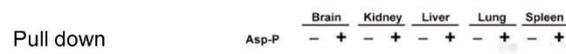


Fig. S7 β -actin

