

Enabling global analysis of protein citrullination and homocitrullination via biotin thiol tag-assisted mass spectrometry

Lingjun Li (✉ lingjun.li@wisc.edu)

University of Wisconsin-Madison <https://orcid.org/0000-0003-0056-3869>

Yatao Shi

University of Wisconsin-Madison

Zihui Li

University of Wisconsin-Madison <https://orcid.org/0000-0003-0865-4634>

Bin Wang

University of Wisconsin-Madison

Xudong Shi

University of Wisconsin-Madison

Hui Ye

China Pharmaceutical University

Daniel Delafield

University of Wisconsin - Madison <https://orcid.org/0000-0003-2989-9833>

Langlang Lv

China Pharmaceutical University

Zhengqing Ye

University of Wisconsin-Madison <https://orcid.org/0000-0002-7242-1504>

Zhengwei Chen

University of Wisconsin-Madison

Fengfei Ma

University of Wisconsin-Madison <https://orcid.org/0000-0003-4202-4724>

Article

Keywords: homocitrullination, citrullination, mass spectrometry

Posted Date: February 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-215281/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Citrullination and homocitrullination are key post-translational modifications (PTMs) that affect protein structures and functions. Although they have been linked to various biological processes and disease pathogenesis, the underlying mechanism remains poorly understood due to a lack of effective tools to enrich, detect, and localize these PTMs. Herein, we report the design and development of a biotin thiol tag that enables derivatization, enrichment, and confident identification of these two PTMs simultaneously via mass spectrometry. We perform global mapping of the citrullination and homocitrullination proteomes of mouse tissues. In total, we identify 1,198 citrullination sites and 108 homocitrullination sites from 619 and 79 proteins, respectively, representing the largest datasets to date. We discover novel distribution and functions of these two PTMs. We also perform multiplexing quantitative analysis via isotopic labeling techniques. This study depicts a landscape of protein citrullination and homocitrullination and lays the foundation to further decipher their physiological and pathological roles.

Introduction

Protein citrullination/deimination is an emerging post-translational modification (PTM) resulting from the conversion of peptidyl arginine to citrulline and is catalyzed by a calcium-regulated family of enzymes called protein arginine deiminases (PADs) (Fig. 1a)^{1,2}. Protein homocitrullination/carbamylation is another chemically related PTM that occurs on lysine side chains. However, it is known as a nonenzymatic PTM and its expression is highly associated with the level of cyanate *in vivo* (Fig. 1b)³. These two types of PTMs lead to the loss of positive charges on the basic amino acid residues under physiological conditions, and therefore have a profound effect on protein conformations, protein-protein interactions and protein functions^{1,2}.

The pathological involvement of these two PTMs was initially explored in rheumatoid arthritis in which pain in the joints is caused by PAD dysregulation. Proteins with aberrant citrullination and homocitrullination also stimulate the generation of anti-citrullinated protein antibodies that are related to atypical autoimmune and inflammatory responses³⁻⁹. In another extensively studied disease multiple sclerosis, excessive citrullination of myelin basic protein (MBP) is considered to be a major driver of partial unfolding of myelin sheath and the resultant impaired neuronal signal transduction¹⁰⁻¹². Moreover, recent accumulating evidence has revealed that citrullination and homocitrullination are associated with the development of diverse pathological states including prion disease¹³, psoriasis¹⁴, Alzheimer's disease (AD)¹⁵⁻¹⁷ and cancers¹⁸⁻²¹, which raises a fast-growing interest in studying these two important PTMs.

Despite the emerging interests, knowledge of the citrullination and homocitrullination proteome is still limited primarily due to the lack of effective analytical tools. Antibody-based techniques such as Western blotting and immunohistochemistry are currently the most prevalent methods to detect these PTMs²²⁻²⁴. However, these approaches are neither suitable for high-throughput analysis nor able to pinpoint exact

sites of the PTMs with confidence^{25,26}. Mass spectrometry (MS)-based strategies, on the other hand, are gaining popularity as powerful tools for large-scale characterization and localization of various PTMs. However, its application to mapping the citrullination and homocitrullination proteome suffers from several challenges^{26,27}. Firstly, signals of these low-abundance PTMs can be largely suppressed by other molecules in the sample and effective enrichment methods are lacking. Secondly, the small mass shift induced by citrullination (+ 0.984 Da) is easily confused with deamidation (+ 0.984 Da) and ¹³C isotopic peaks (+ 1.0033 Da). These limitations contribute to the poor-quality tandem MS spectra, which pose challenges for confident identification and localization of these PTMs. To combat these issues, significant effort has been devoted to improving aspects of the analytical workflow. However, none of the reported methods have overcome all the difficulties so far. For example, direct MS analysis is possible but often requires high mass accuracy of the instrument and time-consuming manual examination of the spectra^{28,29}. Delicate searching algorithms and statistical modeling have also been developed to aid in the direct analysis^{30,31}. Chemical derivatization of the PTMs prior to analysis is an alternative to enlarge the mass shift but usually suffers from incomplete reaction^{32,33}. The above-mentioned strategies did not address the intrinsic low abundance of these PTMs either. Alternative studies have sought the means of using chemical probes for simultaneous introduction of mass shift and enrichment groups. Nevertheless, previous designs led to unsatisfying fragmentation of the peptide backbones and thus limited the identified citrullination and homocitrullination sites^{6,34,35}.

Here, we design a novel biotin thiol tag that enables derivatization and enrichment of citrullinated and homocitrullinated peptides with high specificity and efficiency. We then develop a reliable and robust proteomics approach for large-scale characterization of these PTMs from complex samples. The utility of this pipeline is demonstrated by comprehensive profiling of the landscape of protein citrullination and homocitrullination from different mouse tissues. Furthermore, we combine this novel method with MS-based quantitation strategies, such as isotopic dimethyl labeling, to achieve multiplexed quantitative analysis of citrullination and homocitrullination from various biological samples.

Results

Development of a novel biotin thiol tag for citrullination analysis. Protein citrullination and homocitrullination both feature a ureido group on the side chains that can be used for chemical derivatization as previously reported^{6,22,34,35}. Here, we design a biotin thiol tag that can be easily synthesized with low cost (**Supplementary Fig. 1**) and can specifically react with citrulline or homocitrulline residues together with 2,3-butanedione (Fig. 1c). This derivatization not only increases the mass shift to allow more confident identification, but also introduces a biotin moiety that enables subsequent enrichment of the modified molecules.

We first performed a proof-of-principle test using a synthetic peptide standard containing one citrullination site within the sequence (SAVRACitSSVPGVR) (**Supplementary Fig. 2a**). After 6 h, the reaction was complete without any observable side products (**Supplementary Fig. 2b**), suggesting a high

specificity towards ureido group. The low-abundance peak at m/z 1392 corresponds to the loss of biotin moiety caused by in-source fragmentation when using a matrix-assisted laser desorption/ionization (MALDI) source. We then evaluated the enrichment performance by spiking the derivatized peptide standard into a complex peptide mixture (1:400, w/w) (**Supplementary Fig. 2c**) followed by enrichment with streptavidin beads (**Supplementary Fig. 2d**). The results indicate that derivatized citrullinated peptides can be enriched with excellent specificity and released from streptavidin beads for MS analysis. The peak at m/z 1392 is still present after enrichment which further proves that it originates from in-source fragmentation instead of incomplete derivatization.

Previously reported chemical probes for citrullination analysis had bulky structures that negatively impacted the solubility of analytes. Upon derivatization, extensive yet uninformative fragments were generated from the tag, which severely impeded the peptide backbone fragmentation and therefore led to low identification rates⁶. In contrast, our novel design of biotin thiol tag features a compact structure which only generates two fragment/diagnostic ions during higher-energy collisional dissociation (HCD) (Fig. 1d and **Supplementary Fig. 3a-c**). Consequently, peptide backbones can preserve good fragmentation efficiency and produce rich b/y or c/z ion series during HCD or electron-transfer dissociation (ETD) (**Supplementary Fig. 3a, d-g**), respectively. The collected tandem MS spectra of the derivatized peptide standard delivered nearly full sequence coverage under HCD (Fig. 1e), ETD (**Supplementary Fig. 3h**) or electron-transfer/higher-energy collision dissociation (ETHcD) (**Supplementary Fig. 3i**) fragmentation. Our results indicate that the biotin thiol tag derivatized citrullinated peptides can generate high-quality tandem MS spectra for sequence annotation, which enhances the identification confidence of citrullination sites when coupled with various fragmentation techniques.

Improved *in vitro* protein citrullination analysis with biotin thiol tag. Following the initial experiments, we streamlined the citrullination and homocitrullination analysis using our biotin thiol tag and MS-based bottom-up proteomics approach (Fig. 2a). Proteins were extracted from biological samples and enzymatically digested to peptides. The biotin tag was incubated with the peptides under acidic conditions and reacted with citrulline or homocitrulline residues. Excess tag was removed by strong cation exchange (SCX), and derivatized citrullinated and homocitrullinated peptides were enriched by streptavidin resin. The enriched peptides were then released for liquid chromatography coupled with tandem MS (LC-MS/MS) analysis and data processing.

We tested this procedure using recombinant human histone H3 protein with or without *in vitro* PAD treatment. Although the recombinant protein, which is expressed in *Escherichia coli*, is supposed to bear no citrullination, we identified three citrullination sites in our experiment (Fig. 2b). This is likely due to the presence of unknown PAD isoenzymes in prokaryotes, which has been reported in recent literature^{36,37}. Despite this unexpected result, the other arginine residues are proven to be non-citrullinated which still makes this protein a good negative control. After *in vitro* PAD treatment, we found all the arginine residues were catalyzed to citrulline with abundant peptides confidently identified as citrullinated (Fig. 2c), indicating the high efficacy of our method. Interestingly, some peptides were identified with citrullination sites located at peptide C-termini and two representative tandem MS spectra of high quality

are shown (Fig. 2d, e). It remains controversial whether trypsin is able to cleave after citrulline residues. While some researchers believe citrulline is resistant to trypsin digestion due to its neutral-charge property and even use it as a rule to exclude their identifications^{28,38}, others have reported some C-terminal citrullination sites though manual inspection of the spectra is usually required^{18,31}. Our results suggest that citrulline residues could potentially be cleaved by trypsin.

Exploring different fragmentation techniques and enzymatic digestion methods for optimized citrullination analysis from complex biological samples. We moved forward to evaluate our method with complex biological samples. We first compared three MS fragmentation methods, including stepped HCD, HCD product ion-triggered ETD (HCD-pd-ETD) and HCD product ion-triggered EThcD (HCD-pd-EThcD), using mouse brain digest. All three methods were able to achieve in-depth citrullination analysis with decent numbers of identifications (**Supplementary Fig. 4a**) while stepped HCD method slightly outperformed the other two likely due to shorter cycle time. Different methods show certain overlaps but are also complementary to one another (**Supplementary Fig. 4b, c**), suggesting the importance of choosing an appropriate one depending on specific applications. When comparing the same citrullination site identified with various fragmentation techniques, we observed that they all produced high-quality spectra though EThcD showed even better sequence coverage as expected (**Supplementary Fig. 4d-f**). Thus, we conclude that stepped HCD confers optimal performance for citrullination analysis of complex samples due to its faster acquisition rate and shorter duty cycle while EThcD shines in providing more informative fragment ions and hence is more beneficial for relatively simple systems.

We then sought to optimize the enzymatic digestion methods. Lower identification numbers were observed when only using LysC to digest the samples (**Supplementary Fig. 5a**) probably because LysC digestion produces longer peptides, which results in lower fragmentation efficiency. Similar to the observation from *in vitro* protein analysis, we noticed that 51% of the citrullination sites were identified at peptide C-termini in LysC/trypsin digested samples and this percentage rose to 64% when using trypsin only (**Supplementary Fig. 5b, c**). These findings were consistent to the fact that trypsin digested samples provided slightly more identifications compared to LysC/trypsin digestion (**Supplementary Fig. 5a**) since higher missed cleavage rate of citrulline residues resulted in longer peptides in the latter. When searching the results of LysC digestion with tryptic peptide parameters, we found almost all of the citrullination sites were still identified in the middle of the peptide sequence (**Supplementary Fig. 5d**), which demonstrates that no artificial cleavage of citrulline residues happens after enzymatic digestion. Some citrullination sites were confidently identified with different digestion protocols (**Supplementary Fig. 5e-g**) that further supports our observations of trypsin cleavable C-terminal citrullinated arginine. Again, our results provide stronger evidence to support that some citrulline residues could be cleaved by trypsin though the mechanism needs further investigations.

Taking consideration of citrullination identification rate and economic cost, we determined that using LysC/trypsin digestion and stepped HCD fragmentation technique would be the optimal solution for processing the citrullination proteome. In addition, we evaluated the reproducibility by analyzing three biological replicates and the good overlap among these replicates indicates the robustness of our

optimized methods (**Supplementary Fig. 6**). All the detailed data described in this section are provided in **Supplementary Data 1**.

Large-scale citrullinome profiling of different mouse tissues. Next, we ask whether the developed method can delineate the citrullination landscape from biological samples, and holds potential to elucidate the regulatory mechanisms of citrullination in cells. We performed an in-depth citrullinome analysis of six body organs and five brain regions in mice, generating a first tissue-specific atlas of mouse citrullinome. In total, we identified 1,198 citrullination sites from 619 citrullinated proteins with high confidence (Fig. 3a and **Supplementary Data 2**), which is a dramatic increase compared to previous studies. More importantly, about 60% of these proteins were not reported to be proteins with PTMs retrievable from the UniProt database (**Supplementary Fig. 7**), which suggests that our results greatly expand the understanding of citrullination and how these substrate proteins are subjected to modulation via PTM. Intriguingly, we found each examined brain region doubles in the number of identifications compared to other organs (Fig. 3a); however, the total number of citrullinated proteins in the brain is lower than that in the body (**Supplementary Fig. 7**). To investigate the seemingly contradictory results, we generated two arcplots where the width of ribbons connecting two tissues is proportional to the number of overlapping proteins or sites between them (**Supplementary Fig. 8a, b**). We observed a larger degree of overlap between brain regions with many more shared proteins and sites in between (**Supplementary Fig. 8c-f**). This could indicate protein citrullination functions importantly and similarly across multiple brain regions, while in body organs it is involved in diverse biological processes. Our results greatly expand the knowledge of the substrate proteome for citrullination although the overlapped fraction with UniProt repository is negligible (Fig. 3b). This is likely because nearly 40% of the citrullination sites described in UniProt are based on similarity extrapolation without experimental evidence which are inconsistent with the identified *in vivo* citrullination proteome. In addition, many of those reported sites are located on histone proteins especially at protein termini that may escape detection with our bottom-up strategies (Fig. 3b). Figure 3c captures the prevalence of singly- and multiply-citrullinated proteins where 60% of the identified proteins were observed with only one citrullination site.

The newly discovered citrullination proteome serves as a precious reservoir to conjecture the regulatory mechanisms of citrullination. For instance, we identified ten citrullination sites on MBP while there are only four reported in UniProt database (Fig. 3d). Our results provided high-quality tandem MS spectra, which not only confirmed the presence of known modification sites (Fig. 3e), but also identified unknown sites with confidence (Fig. 3f). These findings may partially explain why MBP is more susceptible to hypercitrullination when PADs are dysregulated under pathological conditions and thus can help better understand the mechanisms in related diseases. Two citrullination sites described in UniProt were not detected in our study which could result from the complementarity of various analytical tools. But again, these sites from UniProt are all based on similarity extrapolation from human and our results might indeed indicate a species-specific profile of protein citrullination. Another interesting example is glial fibrillary acidic protein (GFAP), which is an astrocyte-specific protein marker and is involved in astrocyte-neuron interactions. Increased expression of citrullinated GFAP was also observed in brains from patients

with AD^{15,17}. In this study, we identified 14 citrullination sites on GFAP compared to four described in UniProt (**Supplementary Fig. 9**), which reveals the importance of citrullination in regulating GFAP functions and understanding the pathology of AD and possible other astrocyte disorders. In addition, we detected many novel citrullinated proteins for the first time. For example, we identified one citrullination site on apolipoprotein E (**Supplementary Fig. 10a**) and microtubule-associated protein tau (**Supplementary Fig. 10b**). These two proteins have been proven to be closely associated with the initiation and progression of AD³⁹⁻⁴¹ and our results suggest the possible roles of their citrullinated forms in the pathogenesis of such neurodegenerative diseases. We also identified two novel citrullination sites on NAD-dependent protein deacetylase sirtuin-2, which functions as an essential enzyme targeting histones, tubulin and many key transcription factors, and therefore plays a critical role in many biological processes (**Supplementary Fig. 10c**)⁴².

We then performed a motif analysis and found there was no conserved amino acid sequence patterns flanking identified citrullination sites (Fig. 3g), which is consistent with the observation that PAD treatment universally citrullinates arginine residues on histone H3 *in vitro* (Fig. 2c). To better discern the general functions that citrullinated proteins are involved in, we generated heatmaps showing multi-organ gene ontology (GO) analyses. Twenty most significantly enriched cellular components (Fig. 3h) or biological processes (Fig. 3i) are shown where the color coding indicates the *p* values of a certain term in different tissues. We found that there are clear disparities between brain and body while citrullinated proteins are more involved in brain functions. Specifically, citrullinated proteins are concentrated in axon, myelin sheath, dendrite and synapse, and consequently function importantly in the central nervous system. Furthermore, they also participate in many critical metabolic processes including respiration and are observed to enrich in mitochondria. In accordance with this, we also identified eight citrullination sites on an essential glycolytic enzyme pyruvate kinase (PKM). Interestingly, many of these sites on PKM are located in the proximity of its substrate binding pockets (**Supplementary Fig. 11a**), which raises the likelihood that citrullination can influence the kinase activity and supports a recent study concluding that citrullination regulates glycolysis⁴³. For instance, R120 and R294 are located near the catalytic pocket (**Supplementary Fig. 11b**) while R455 and R461 are close to the allosteric center (**Supplementary Fig. 11c, d**). R399 was also shown to be very important in stabilizing the highly active tetrameric form (**Supplementary Fig. 11e, f**)⁴⁴. Our results greatly expand current understandings of protein citrullination by demonstrating its widespread distribution (**Supplementary Fig. 12**) and involvement in many other biological processes (**Supplementary Fig. 13**), molecular functions (**Supplementary Fig. 14**) and Kyoto encyclopedia of genes and genomes (KEGG) pathways (**Supplementary Fig. 15**).

Additionally, we noticed that 30 citrullination sites are colocalized with other arginine modifications especially omega-n-methylarginine (**Supplementary Fig. 16a**). For example, we identified five citrullination sites on heterogeneous nuclear ribonucleoproteins A2/B1 (Hnrnpa2b1) and four of them were also reported as arginine methylation sites (**Supplementary Fig. 16b**). Hnrnpa2b1 was shown to influence RNA metabolism and transport, and arginine methylation could regulate the nucleocytoplasmic distribution of

this protein⁴⁵. Our results raise the possibility that citrullination indirectly participates in biological processes through an interplay with other protein modifications such as arginine methylation.

Profiling of protein homocitrullination in different mouse tissues. Homocitrullination is highly similar to citrullination structurally though it occurs on lysine residues. Therefore, current methods using antibodies to detect protein homocitrullination suffer from poor specificity while MS-based approaches also result in unsatisfying identification rate due to its low abundance²⁵. In contrast, our biotin thiol tag takes advantage of its high specificity towards ureido groups on both citrulline and homocitrulline, which allows for simultaneous enrichment and characterization of these two PTMs. We identified 108 homocitrullination sites from 79 proteins across all the tissues (**Supplementary Data 3**), which fills in the blank of protein homocitrullination database. Similarly, more sites and proteins were identified in brain regions compared to body organs that suggests its intimate association with brain functions (**Fig. 4a**). We also observed relatively high identification numbers in heart, which may indicate that this PTM is associated with processes such as transporting oxygenated blood and hormones to the body (**Fig. 4a**). Many other PTMs are described in UniProt on these homocitrullination sites detected in our study, indicating again a potential PTM crosstalk. For instance, we identified two homocitrullination sites with high confidence on histone H4 (**Fig. 4b, c**) while both are colocalized with several lysine modifications (**Fig. 4b**). These modifications were shown to play critical roles, which could modulate the packaging of chromatin by either directly altering chemical structures of histones or recruiting PTM-specific binding proteins^{46–51}. Our findings of competing homocitrullination sites on histones provide new insights into the complex regulatory mechanisms in dynamic chromatin-templated processes. No obvious sequence patterns surrounding homocitrullination sites were observed either, though there is higher propensity that the identified sites are located near protein C-termini (**Fig. 4d**). As expected, homocitrullinated proteins are concentrated in myelin sheath and may function importantly in the nervous system (**Fig. 4e**). They also participate in functions related to oxygen binding (**Fig. 4e**), which is consistent to more homocitrullinated proteins being identified in heart (**Fig. 4a**). Interestingly, we found that they are more likely to locate in mitochondria and correspondingly associated with processes such as tricarboxylic acid cycle (**Fig. 4e**). Homocitrullinated proteins may interfere with cell-cell interactions as well which can be discerned from their prevalence in extracellular matrix components (**Fig. 4e**).

Multiplexed quantitative citrullination analysis using chemical labeling strategies. We then sought to achieve multiplexed quantitative analysis by combining our methods with chemical labeling strategies. In theory, samples can be differentially labeled and combined before being derivatized and enriched using our biotin thiol tag (**Fig. 5a**). For isotopic labeling such as reductive dimethylation, quantification can be achieved during survey scans⁵². While for isobaric labeling approaches such as tandem mass tag (TMT)⁵³ or *N,N*-dimethyl leucine (DiLeu)⁵⁴, quantitative information can be obtained from reporter ions upon fragmentation (**Fig. 5a**).

In this study, we explored the quantitation capability of duplex dimethyl labeling which introduces a 4 Da mass difference between heavy isotopic labeling and light labeling (**Supplementary Fig. 17**). We first

tested this pipeline with citrullinated peptide standard and found the standard could be completely dimethylated without showing any observable side reactions (**Supplementary Fig. 18a, b**), which ensures no interference to the following steps. When differentially labeling the standard and mixing with known ratios, accurate quantitation was achieved (**Supplementary Fig. 18c-e**) and reliable results were obtained after biotin thiol tag derivatization (**Fig. 5b**). We moved forward to evaluate this strategy using complex biological samples (**Supplementary Fig. 19a**). Although we observed lower identification and quantification numbers which was likely due to increased complexity of the spectra (**Supplementary Fig. 19b**), the quantified citrullinated peptides exhibited both great accuracy and precision compared to theoretical ratios (**Fig. 5c**). In addition, the identification and quantification rates can be easily improved by increasing the amount of starting material or utilizing longer separation gradient. It is also worthwhile to note that dimethyl labeling conditions need to be carefully controlled to quantify homocitrullinated peptides since dimethylated lysine may affect further biotin thiol tag reaction.

Discussion

Herein, we report the design and development of a biotin thiol tag that specifically reacts with citrulline and homocitrulline and allows for enrichment of target molecules. After demonstrating its efficacy using standard peptide and recombinant protein, we streamline the workflow to detect these two PTMs from complex biological samples. We then apply this protocol to profile protein citrullination and homocitrullination of five brain regions and six body organs in mice. In total, we identify 1,198 citrullination sites and 108 homocitrullination sites from 619 and 79 proteins, respectively, which is the largest dataset to date. Our study reveals the critical roles these two PTMs may play in the nervous system and indicate they also function importantly in many metabolic processes including respiration and glycolysis. Despite a few intrinsic drawbacks with the mass difference isotopic labeling techniques, we demonstrate that reductive dimethylation can be utilized in conjunction with our method to achieve simultaneous high-throughput quantitative analysis. We will also integrate isobaric labeling strategies to alleviate these shortcomings and further increase the multiplexing capability for quantitative PTM analyses in the future. Collectively, our results expand current understanding of protein citrullination and homocitrullination by mapping their widespread distribution in different tissues and participation in various biological processes than hitherto anticipated. More importantly, we envision our method can serve as a simple yet powerful tool for unambiguous identification and quantification of these modifications, which will also inspire and benefit future investigations into their functional roles under physiological and pathological conditions.

Methods

Synthesis of biotin thiol tag. *N,N*-diisopropylethylamine (0.88 mM) was added to a solution of biotin-NHS ester (0.29 mM) and cysteamine (0.44 mM) in CH₂Cl₂ (5 mL) and stirred at 40°C for 24 h. The crude product was purified using a CombiFlash system with a gradient of dichloromethane from 0 to 20% in methanol. Fractions containing pure product (as detected by UV) were collected (68% yield). ¹H NMR data

was obtained from a Varian Inova 500 MHz NMR spectrometer. ^{13}C NMR data was obtained from a Bruker Avance III HD 400 MHz NMR spectrometer. The spectra were recorded in $10\text{ mg cm}^{-3}\text{ CD}_3\text{OD}$ solutions with a probe temperature of 300 K and referenced to internal standard tetramethylsilane. ^1H NMR (500 MHz, CD_3OD) δ 8.17 (m, 1H), 4.56 (dd, $J=7.7, 5.0$ Hz, 1H), 4.37 (dd, $J=7.8, 4.5$ Hz, 1H), 3.57–3.53 (m, 1H), 3.42–3.40 (m, 2H), 3.28 (dt, $J=9.9, 5.3$ Hz, 1H), 3.00 (dd, $J=12.7, 5.0$ Hz, 1H), 2.89 (q, $J=6.4$ Hz, 1H), 2.77 (d, $J=12.7$ Hz, ^1H), 2.67 (t, $J=6.8$ Hz, 2H), 2.29 (t, $J=7.4$ Hz, 2H), 1.85–1.63 (m, 4H), 1.55–1.49 (m, 2H). ^{13}C NMR (101 MHz, CD_3OD) δ 174.8, 163.2, 62.0, 60.2, 55.6, 42.5, 39.6, 35.3, 28.4, 28.1, 25.4, 23.1. Formula: $\text{C}_{12}\text{H}_{22}\text{N}_3\text{O}_2\text{S}_2$; $[\text{M} + \text{H}]^+$: m/z 304.1153 Da. Tris(2-carboxyethyl)phosphine (Millipore) was added into the biotin thiol tag solution to a final concentration of 10 mM before drying to prevent oxidation and the tag was stored at -80°C for long-term storage.

Derivatization of citrullinated peptide standard using biotin thiol tag. Citrullinated peptide standard SAVRACitSSVPGVR (Genscript) was dissolved in water to a concentration of 1 mg/mL. A solution of 2,3-butanedione was prepared by mixing 1 μL of 2,3-butanedione with 114 μL 12.5% trifluoroacetic acid (TFA). Three hundred microgram of biotin thiol tag was dissolved with 40 μL 12.5% TFA. One microliter of citrullinated peptide standard and 10 μL 2,3-butanedione solution were subsequently added to initiate the derivatization reaction. The mixture was vortexed in dark at 37°C for 6 h and then dried in vacuo. To remove the excess tag, SCX was performed using TopTips (Poly LC) containing PolySULFOETHYL A beads following the manufacturer's protocol. Briefly, SCX tips were equilibrated with 100 μL loading buffer containing 50% acetonitrile (ACN), 0.2% formic acid (FA) and 10 mM ammonium formate for three times. The derivatized citrullinated peptide standard was then resuspended in 200 μL loading buffer and added to SCX tips followed by washing with 100 μL loading buffer for 10 times. Peptide was finally eluted with 50 μL 25% ACN and 0.4 M ammonium formate for 3 times. Flowthrough was collected and dried in vacuo. All centrifugation steps were performed at 400 g for 2 min.

Enrichment of derivatized citrullinated peptide standard. The enrichment process was performed as previously described with slight modifications⁵⁵. Briefly, 75 μL streptavidin agarose (Sigma) was washed with 1 mL $1\times$ phosphate-buffered saline (PBS) for 5 times. Each time the tube containing beads was vortexed and centrifuged at 3,000 g for 2 min, and supernatant was removed. Peptide sample was resuspended in 1 mL PBS and loaded onto the streptavidin agarose followed by incubation at room temperature for 2 h with rotation. The agarose was subsequently washed with 1 mL PBS for 3 times, 1 mL 5% ACN in PBS for 3 times, and 1 mL water for 10 times. Peptides were finally released with 300 μL 80% ACN, 0.2% TFA and 0.1% FA for four times. The first release was performed in room temperature for 5 min, while the other three release processes were conducted at 95°C for 5 min with shaking. The eluents were combined and dried in vacuo.

MALDI-MS analysis of citrullinated peptide standard and its derivatized form. Samples were resuspended in 50 μL 50% ACN and prepared by premixing 1 μL of them with 1 μL of 2,5-dihydroxybenzoic acid matrix (150 mg/mL in 50% methanol, 0.1% FA). One microliter of each matrix/sample mixture was spotted onto the MALDI target plate and detected on a MALDI-LTQ Orbitrap XL mass spectrometer (Thermo).

Ionization was performed using a laser energy of 15 μJ . Spectra were acquired with a mass range of m/z 1000 – 2000 at a resolution of 30k (at m/z 400).

Fragmentation of derivatized citrullinated peptide standard. The derivatized citrullinated peptide standard was resuspended in 1 mL 0.1% FA, 50% ACN and directly injected into an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo). Full MS scan was performed with a mass range of m/z 300–1500 using a resolution of 60k and RF lens of 30. AGC target was set to 2×10^5 and the maximum injection time was 100 ms. The precursor ion was isolated in quadruple for HCD, ETD and EThcD fragmentation. Tandem MS spectra were collected, and fragment ions were manually annotated based on their accurate mass.

PAD treatment and digestion of histone H3. Ten microgram of recombinant human histone H3 (New England Biolabs) was incubated with recombinant human PAD2/PAD4 enzyme (Cayman Chemical) overnight at room temperature at a ratio of 2 μg enzyme per mg of histone. Histone with or without PAD treatment was then diluted with Tris buffer to a final concentration of 100 mM Tris and 5 mM CaCl_2 (pH 7.5). LysC/trypsin mixture (Promega) was added in a 50:1 ratio (protein:enzyme, w/w) and incubated overnight at 37°C. Digestion was quenched by adding TFA to a pH < 3 and desalted using Omix Tips (Agilent) before drying in vacuo.

Protein extraction and digestion of mouse tissues. For method optimization, brain was collected from one mouse. For tissue-specific citrullination and homocitrullination profiling, five brain regions and six body organs were collected: Bcortex (cerebral cortex), Scortex (hippocampus and thalamus), hypothalamus, cerebellum, medulla, spleen, pancreas, kidney, lung, heart, and liver. Each tissue was collected as triplicates from three mice. Tissues were dissolved in 150 μL of extraction buffer solution (4 % SDS, 50 mM Tris buffer) and sonicated using a probe sonicator (Thermo). Protein extracts were reduced with 10 mM dithiothreitol (DTT) for 30 min at room temperature and alkylated with 50 mM iodoacetamide for another 30 min in dark before quenched with DTT. Proteins were then precipitated with 80% (v/v) cold acetone (-20 °C) overnight. Samples were centrifuged at 14,000 g for 15 min after which supernatant containing SDS (in the extraction buffer) was discarded. Pellets were rinsed with cold acetone again and air-dried at room temperature. Five moles of guanidine hydrochloride (GuHCl) were added to dissolve the pellets and 50 mM Tris buffer was used to dilute the samples to a GuHCl concentration < 0.5 M. On-pellet digestion was performed with either trypsin, LysC or LysC/trypsin mixture (Promega) in a 50:1 ratio (protein:enzyme, w/w) at 37 °C overnight. The digestion was quenched with 1% TFA and samples were desalted with Sep-Pak C18 cartridges (Waters). Concentrations of peptide mixture were measured by peptide assay (Thermo). Four hundred microgram of peptide was aliquoted for each sample and dried in vacuo.

Duplex isotopic dimethyl labeling. Forty microliters of H_2O were added to dissolve peptide samples. After diluted to 1% (v/v) with H_2O , 20 μL formaldehyde or formaldehyde-d2 solution was added to samples for light or heavy labeling, respectively. To each sample 20 μL of borane pyridine (30 mM) was then added to initiate the labeling reaction. Following incubation at 37 °C for 20 min, labeling was quenched by addition of 20 μL ammonium bicarbonate solution (200 mM). Labeled peptides were then combined in

1:1, 2:1 or 5:1 ratio (v/v, light/heavy). Samples were acidified with FA to pH < 3, desalted with Sep-Pak C18 cartridges and dried in vacuo for later biotin thiol tag derivatization.

Derivatization and enrichment of citrullinated peptides in histone and mouse tissues. Three hundred microgram of biotin thiol tag was added to each sample tube containing peptides from mouse tissues or histone and resuspended in 40 μ L 12.5% TFA solution. Ten microliters of 2,3-butanedione solution prepared as mentioned before was added to initiate the reaction. The rest of derivatization, SCX and enrichment steps were the same as those for citrullinated peptide standard.

LC-MS/MS analysis. Samples were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo) coupled to a Dionex UltiMate 3000 UPLC system. Each sample was dissolved in 3% ACN, 0.1% FA in water before loaded onto a 75 μ m inner diameter homemade microcapillary column which is packed with 15 cm of Bridged Ethylene Hybrid C18 particles (1.7 μ m, 130 Å, Waters) and fabricated with an integrated emitter tip. Mobile phase A was composed of water and 0.1% FA while mobile phase B was composed of ACN and 0.1% FA. LC separation was achieved across a 100-min gradient elution of 3–30% mobile phase B at a flow rate of 300 nL/min. Survey scans of peptide precursors from 350 to 1500 m/z were performed at a resolving power of 60k (at m/z 200) with an AGC target of 2×10^5 and maximum injection time of 100 ms. For histone samples, survey scans of peptide precursors were performed from 200 to 1500 m/z . For stepped HCD method, precursors were selected for fragmentation for continuous 3 s with a stepped normalized collision energy of 27, 30 and 33. Tandem MS acquisition was performed with an isolation window of 1.6 Da, a resolving power of 30k, an AGC target of 5×10^4 , a maximum injection time of 54 ms, and a lower mass limit of 120 m/z . Precursors were subject to dynamic exclusion for 45 s with a 10-ppm tolerance. For HCD-pd-ETD and HCD-pd-ETHcD methods, m/z 227.0848 and 304.1147 were listed as HCD product ions to trigger the subsequent ETD or ETHcD fragmentation. In ETD acquisition, ETD reaction time was set to 100 ms for charge state 2 and 40 ms for charge state 3–6. ETD reagent target was set as 2×10^5 and max ETD reagent injection time as 200 ms. Other parameters were the same as in stepped HCD method. In ETHcD acquisition, each ETD scan was accompanied with a supplemental activation collision energy of 40 and all other settings remained the same. Each sample was acquired in technical duplicates for fragmentation method comparison and triplicates for all the other samples.

Data analysis. Raw files were searched against the UniProt *Mus musculus* reviewed database (December 2018) using MaxQuant (version 1.5.2.8) with trypsin/P selected as the enzyme and three missed cleavages allowed. Histone data files were searched against the UniProt *Homo sapiens* reviewed database (February 2020). Carbamidomethylation of cysteine residues (+ 57.02146 Da) were chosen as fixed modifications and variable modifications included oxidation of methionine residues (+ 15.99492 Da), biotin tag-labeled citrullination of arginine (+ 354.10718 Da) and biotin tag-labeled homocitrullination of lysine (+ 396.12898 Da). A neutral loss of biotin tag (303.10752 Da) and two diagnostic ions of 227.08487 Da and 304.11479 Da were included in the search. For dimethyl labeled samples, multiplicity was set to two with dimethLys0/dimethNter0 specified as light labels, and dimethLys4/dimethNter4 as heavy labels. Search results were filtered to 1% false discovery rate (FDR) at

both peptide and protein levels. Peptides that were found as reverse or potential contaminant hits were filtered out and citrullination or homocitrullination site localization probability threshold was set to 0.75. All other parameters were set as default. Bioinformatic analyses including Sankey diagram, arcplots and stacked bar graphs were performed using R packages. Heatmaps showing multi-tissue GO analyses were generated using Metascape⁵⁶ (version 3.5) while GO analysis for homocitrullination was accomplished using DAVID bioinformatics resources⁵⁷ with a FDR cutoff of 0.05. Sequence motif analyses were done using WebLogo⁵⁸. For homology modeling, the 3D structure of mouse PKM2 (residues 14–531) was modeled according to the crystal structures of human PKM2, which delivered a sequence identity of 97.7% and represented the most similar crystal structures to mouse PKM2 retrievable from the Protein Data Bank. The homology model module of Discovery Studio 2016 was used for the multi-templates structure construction, and the ligands including PYR, SER, FBP were copied from the input templates. The output model with the lowest PDF total energy and DOPE score was adopted, and energy minimization was conducted on the adopted structure using CHARMM (version 40.1). PyMOL (version 2.4.0a0) was used to measure the Euclidean distances between the atoms of the selected arginine residues and atoms of annotated ligands (such as substrates and allosteric activator) in Å.

Declarations

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange⁵⁹ Consortium via the PRIDE⁶⁰ partner repository with the dataset identifier PXD023733. Public release of the data will be made on time for online publication of the paper. For anonymous access to peer-reviewers please use the following account information: Username: reviewer_pxd023733@ebi.ac.uk, Password: nGAPJrm. *Mus musculus* and *Homo sapiens* databases used for database searching were downloaded from UniProt (<https://www.uniprot.org/>).

Code availability

All online-available software or R packages to perform data analysis or generate the figures are indicated throughout the manuscript.

Acknowledgements

This study was supported in part by grant funding from the NIH (R21AG060242, U01CA231081, RF1AG052324, and R01DK071801). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. L.L. acknowledges a Vilas Distinguished Achievement Professorship and the Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Author contributions

Y.S. and Z.L. contributed equally to this work. Y.S., Z.L. and L.L. designed the study. Y.S. and Z.L. performed the experiments and analyzed the data. B.W. was involved in sample preparations. X.S. collected the tissues. Z.Y. synthesized the biotin thiol tag. X.S., D.G.D., B.W., H.Y., Z.C. and F.M. were involved in data analysis and interpretation. Y.S., Z.L. and L.L. prepared the manuscript and all authors provided editorial feedback.

Competing interests

The authors declare the following competing financial interests: The University of Wisconsin–Madison is in the process of filing a provisional patent application P190302US01, based on this work. L.L., Y.S. and Z.L. are named as inventors on this provisional patent application.

References

1. Fuhrmann, J., Clancy, K.W. & Thompson, P.R. Chemical biology of protein arginine modifications in epigenetic regulation. *Chem. Rev.* **115**, 5413-5461 (2015).
2. Fuhrmann, J. & Thompson, P.R. Protein Arginine Methylation and Citrullination in Epigenetic Regulation. *ACS chemical biology* **11**, 654-668 (2016).
3. Pruijn, G.J. Citrullination and carbamylation in the pathophysiology of rheumatoid arthritis. *Front Immunol* **6**, 192 (2015).
4. Schellekens, G.A., de Jong, B.A., van den Hoogen, F.H., van de Putte, L.B. & van Venrooij, W.J. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. *The Journal of clinical investigation* **101**, 273-281 (1998).
5. Elkon, K.B. Poking holes in rheumatoid joints. *Sci Transl Med* **5**, 209fs239 (2013).
6. Tutturen, A.E., Fleckenstein, B. & de Souza, G.A. Assessing the citrullinome in rheumatoid arthritis synovial fluid with and without enrichment of citrullinated peptides. *Journal of proteome research* **13**, 2867-2873 (2014).
7. Turunen, S., *et al.* Rheumatoid arthritis antigens homocitrulline and citrulline are generated by local myeloperoxidase and peptidyl arginine deiminases 2, 3 and 4 in rheumatoid nodule and synovial tissue. *Arthritis Res Ther* **18**, 239 (2016).
8. Tilvawala, R., *et al.* The Rheumatoid Arthritis-Associated Citrullinome. *Cell Chem Biol* **25**, 691-704 e696 (2018).
9. Fert-Bober, J., Darrah, E. & Andrade, F. Insights into the study and origin of the citrullinome in rheumatoid arthritis. *Immunol Rev* **294**, 133-147 (2020).
10. Moscarello, M.A., Mastronardi, F.G. & Wood, D.D. The role of citrullinated proteins suggests a novel mechanism in the pathogenesis of multiple sclerosis. *Neurochem Res* **32**, 251-256 (2007).

11. Bradford, C.M., *et al.* Localisation of citrullinated proteins in normal appearing white matter and lesions in the central nervous system in multiple sclerosis. *J Neuroimmunol* **273**, 85-95 (2014).
12. Yang, L., Tan, D. & Piao, H. Myelin Basic Protein Citrullination in Multiple Sclerosis: A Potential Therapeutic Target for the Pathology. *Neurochem Res* **41**, 1845-1856 (2016).
13. Jang, B., *et al.* Accumulation of citrullinated proteins by up-regulated peptidylarginine deiminase 2 in brains of scrapie-infected mice: a possible role in pathogenesis. *The American journal of pathology* **173**, 1129-1142 (2008).
14. Ishida-Yamamoto, A., *et al.* Decreased deiminated keratin K1 in psoriatic hyperproliferative epidermis. *J Invest Dermatol* **114**, 701-705 (2000).
15. Ishigami, A., *et al.* Abnormal accumulation of citrullinated proteins catalyzed by peptidylarginine deiminase in hippocampal extracts from patients with Alzheimer's disease. *Journal of neuroscience research* **80**, 120-128 (2005).
16. Acharya, N.K., *et al.* Neuronal PAD4 expression and protein citrullination: possible role in production of autoantibodies associated with neurodegenerative disease. *J Autoimmun* **38**, 369-380 (2012).
17. Ishigami, A., *et al.* Mass spectrometric identification of citrullination sites and immunohistochemical detection of citrullinated glial fibrillary acidic protein in Alzheimer's disease brains. *Journal of neuroscience research* **93**, 1664-1674 (2015).
18. Yuzhalin, A.E., *et al.* Colorectal cancer liver metastatic growth depends on PAD4-driven citrullination of the extracellular matrix. *Nat. Commun.* **9**, 4783 (2018).
19. Yuzhalin, A.E. Citrullination in Cancer. *Cancer Res* **79**, 1274-1284 (2019).
20. Stadler, S.C., *et al.* Dysregulation of PAD4-mediated citrullination of nuclear GSK3beta activates TGF-beta signaling and induces epithelial-to-mesenchymal transition in breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11851-11856 (2013).
21. Chang, X., *et al.* Increased PADI4 expression in blood and tissues of patients with malignant tumors. *BMC Cancer* **9**, 40 (2009).
22. Senshu, T., Sato, T., Inoue, T., Akiyama, K. & Asaga, H. Detection of citrulline residues in deiminated proteins on polyvinylidene difluoride membrane. *Anal. Biochem.* **203**, 94-100 (1992).
23. Nicholas, A.P., *et al.* Immunohistochemical localization of citrullinated proteins in adult rat brain. *The Journal of comparative neurology* **459**, 251-266 (2003).
24. Moelants, E.A., Van Damme, J. & Proost, P. Detection and quantification of citrullinated chemokines. *PloS one* **6**, e28976 (2011).
25. Verheul, M.K., *et al.* Pitfalls in the detection of citrullination and carbamylation. *Autoimmun Rev* **17**, 136-141 (2018).
26. Hensen, S.M. & Pruijn, G.J. Methods for the detection of peptidylarginine deiminase (PAD) activity and protein citrullination. *Mol Cell Proteomics* **13**, 388-396 (2014).
27. Clancy, K.W., Weerapana, E. & Thompson, P.R. Detection and identification of protein citrullination in complex biological systems. *Curr Opin Chem Biol* **30**, 1-6 (2016).

28. Lee, C.Y., *et al.* Mining the Human Tissue Proteome for Protein Citrullination. *Mol Cell Proteomics* **17**, 1378-1391 (2018).
29. Raijmakers, R., *et al.* Elevated levels of fibrinogen-derived endogenous citrullinated peptides in synovial fluid of rheumatoid arthritis patients. *Arthritis Res Ther* **14**, R114 (2012).
30. Fert-Bober, J., *et al.* Mapping Citrullinated Sites in Multiple Organs of Mice Using Hypercitrullinated Library. *Journal of proteome research* (2019).
31. Wang, X., *et al.* Accurate Identification of Deamidation and Citrullination from Global Shotgun Proteomics Data Using a Dual-Search Delta Score Strategy. *Journal of proteome research* **19**, 1863-1872 (2020).
32. Stensland, M., Holm, A., Kiehne, A. & Fleckenstein, B. Targeted analysis of protein citrullination using chemical modification and tandem mass spectrometry. *Rapid Commun Mass Spectrom* **23**, 2754-2762 (2009).
33. De Ceuleneer, M., *et al.* Modification of citrulline residues with 2,3-butanedione facilitates their detection by liquid chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* **25**, 1536-1542 (2011).
34. Tuttunen, A.E., Holm, A. & Fleckenstein, B. Specific biotinylation and sensitive enrichment of citrullinated peptides. *Anal Bioanal Chem* **405**, 9321-9331 (2013).
35. Choi, M., Song, J.S., Kim, H.J., Cha, S. & Lee, E.Y. Matrix-assisted laser desorption ionization-time of flight mass spectrometry identification of peptide citrullination site using Br signature. *Analytical biochemistry* **437**, 62-67 (2013).
36. Santa Maria, J., Vallance, P., Charles, I.G. & Leiper, J.M. Identification of microbial dimethylarginine dimethylaminohydrolase enzymes. *Mol. Microbiol.* **33**, 1278-1279 (1999).
37. Kosgodage, U.S., *et al.* Peptidylarginine Deiminase Inhibitors Reduce Bacterial Membrane Vesicle Release and Sensitize Bacteria to Antibiotic Treatment. *Front Cell Infect Microbiol* **9**, 227 (2019).
38. Bennike, T., *et al.* Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids. *Journal of Proteomics & Bioinformatics* **6**(2013).
39. Yamazaki, Y., Zhao, N., Caulfield, T.R., Liu, C.-C. & Bu, G. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. *Nature Reviews Neurology* **15**, 501-518 (2019).
40. Liu, C.C., Liu, C.C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106-118 (2013).
41. Iqbal, K., Liu, F., Gong, C.X. & Grundke-Iqbal, I. Tau in Alzheimer disease and related tauopathies. *Curr Alzheimer Res* **7**, 656-664 (2010).
42. de Oliveira, R.M., Sarkander, J., Kazantsev, A.G. & Outeiro, T.F. SIRT2 as a Therapeutic Target for Age-Related Disorders. *Front. Pharmacol.* **3**, 82 (2012).
43. Coassolo, S., *et al.* Citrullination of pyruvate kinase by PADI1 and PADI3 regulates glycolysis and cancer cell proliferation. *bioRxiv*, 718486 (2020).

44. Gao, X., Wang, H., Yang, J.J., Liu, X. & Liu, Z.R. Pyruvate kinase M2 regulates gene transcription by acting as a protein kinase. *Mol Cell* **45**, 598-609 (2012).
45. Friend, L.R., *et al.* Arginine methylation of hnRNP A2 does not directly govern its subcellular localization. *PLoS one* **8**, e75669 (2013).
46. Ye, J., *et al.* Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly. *Mol Cell* **18**, 123-130 (2005).
47. Yan, Q., *et al.* BBAP monoubiquitylates histone H4 at lysine 91 and selectively modulates the DNA damage response. *Mol Cell* **36**, 110-120 (2009).
48. Bao, X., *et al.* Glutarylation of Histone H4 Lysine 91 Regulates Chromatin Dynamics. *Mol Cell* **76**, 660-675 e669 (2019).
49. Xie, Z., *et al.* Lysine succinylation and lysine malonylation in histones. *Mol Cell Proteomics* **11**, 100-107 (2012).
50. Dai, L., *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat Chem Biol* **10**, 365-370 (2014).
51. Chen, Y., *et al.* Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol Cell Proteomics* **6**, 812-819 (2007).
52. Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A.J. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **4**, 484-494 (2009).
53. Li, J., *et al.* TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nature methods* **17**, 399-404 (2020).
54. Frost, D.C., Greer, T. & Li, L. High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Anal. Chem.* **87**, 1646-1654 (2015).
55. Schiapparelli, L.M., *et al.* Direct detection of biotinylated proteins by mass spectrometry. *Journal of proteome research* **13**, 3966-3978 (2014).
56. Zhou, Y., *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
57. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
58. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).
59. Deutsch, E.W., *et al.* The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic acids research* **48**, D1145-D1152 (2020).
60. Perez-Riverol, Y., *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic acids research* **47**, D442-D450 (2019).

Figures

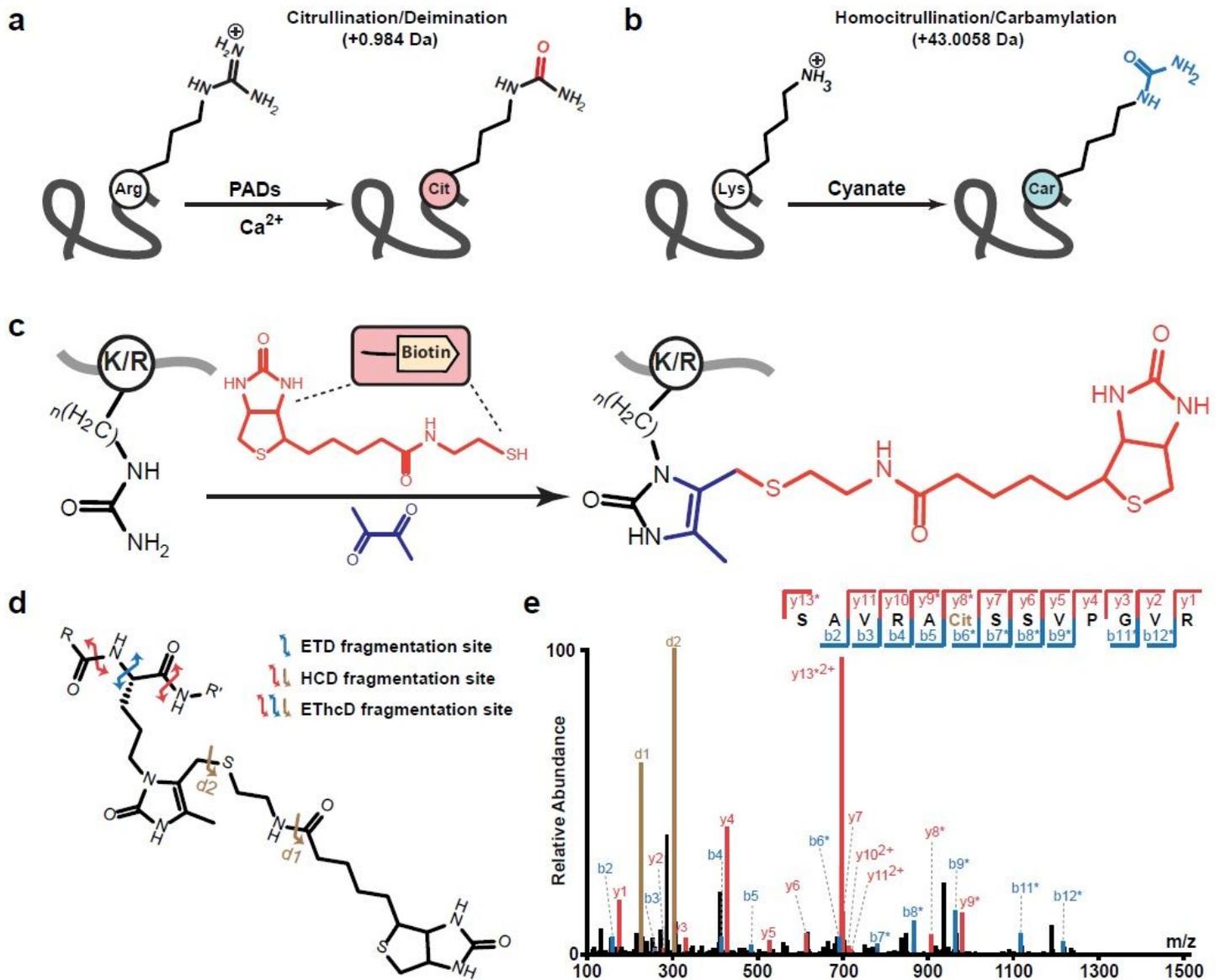


Figure 1

Design of biotin thiol tag for citrullination and homocitrullination analysis. a, PAD enzymes catalyze the hydrolytic conversion of peptidyl arginine residues into citrulline. b, Schematic showing homocitrullination process resulting from the chemical modification of lysine residues. c, Derivatization of citrullinated or homocitrullinated peptides using biotin thiol tag and 2,3-butanedione. d, Fragmentation sites of biotin thiol tag-derivatized citrullinated peptides upon HCD, ETD, or ETHcD fragmentations. e, Tandem MS spectrum of the biotin thiol tag-derivatized citrullinated peptide standard upon HCD fragmentation.

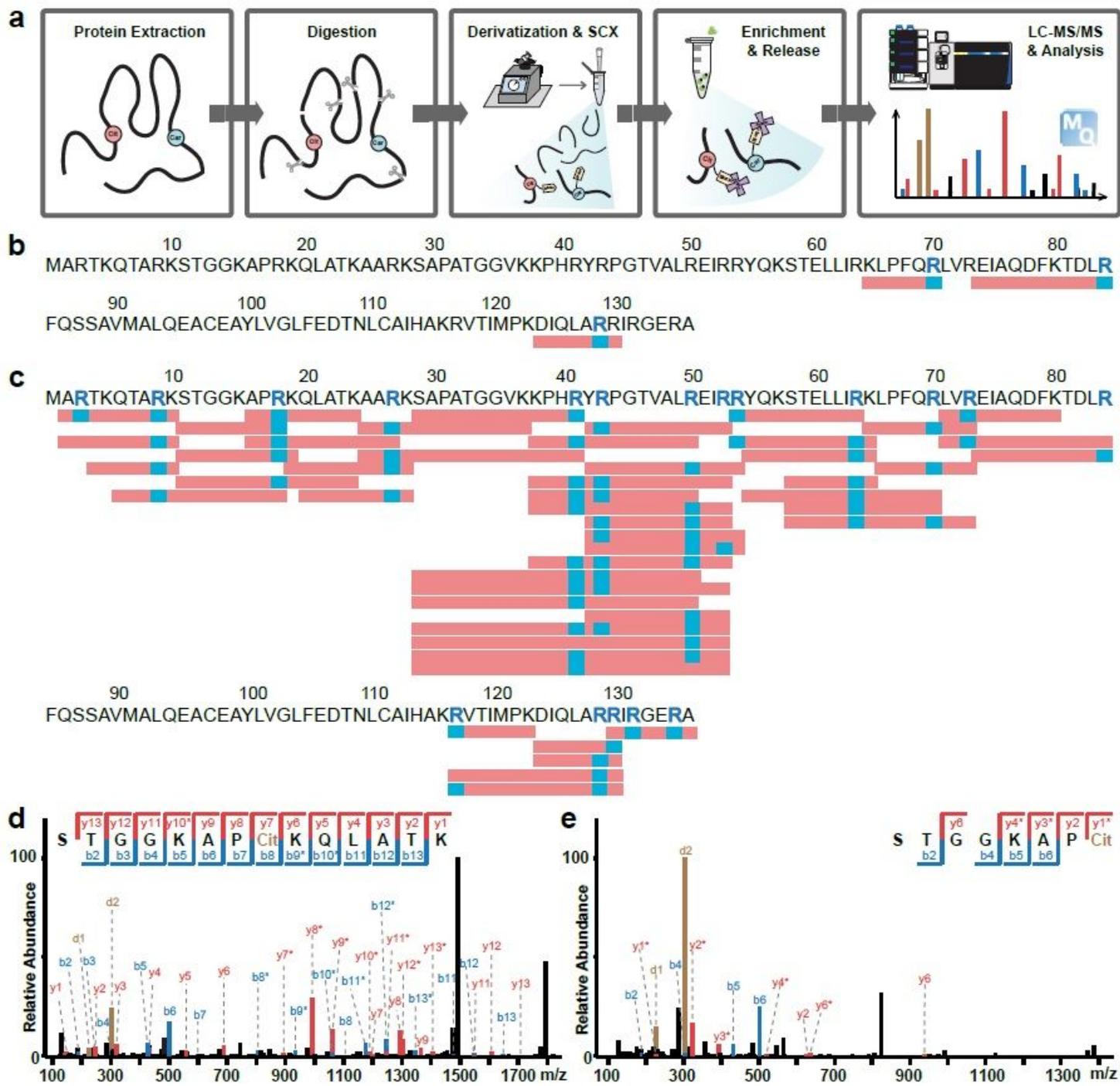


Figure 2

Improved in vitro protein citrullination analysis with biotin thiol tag. **a**, Experimental workflow of protein citrullination analysis with biotin thiol tag. **b,c**, Citrullination analysis on histone H3 protein before (**b**) and after (**c**) in vitro PAD treatment. Identified citrullination sites are highlighted as blue letters within the sequence. Red rectangles below the sequence indicate confidently identified citrullinated peptides while citrullination sites are shown in blue. **d**, Example tandem MS spectrum of an identified citrullinated peptide from PAD treated histone H3 with the citrullination site (R18Cit) in the middle of the peptide. **e**,

Example tandem MS spectrum of the same citrullination site (R18Cit) identified at the C-terminal of another citrullinated peptide from PAD treated histone H3.

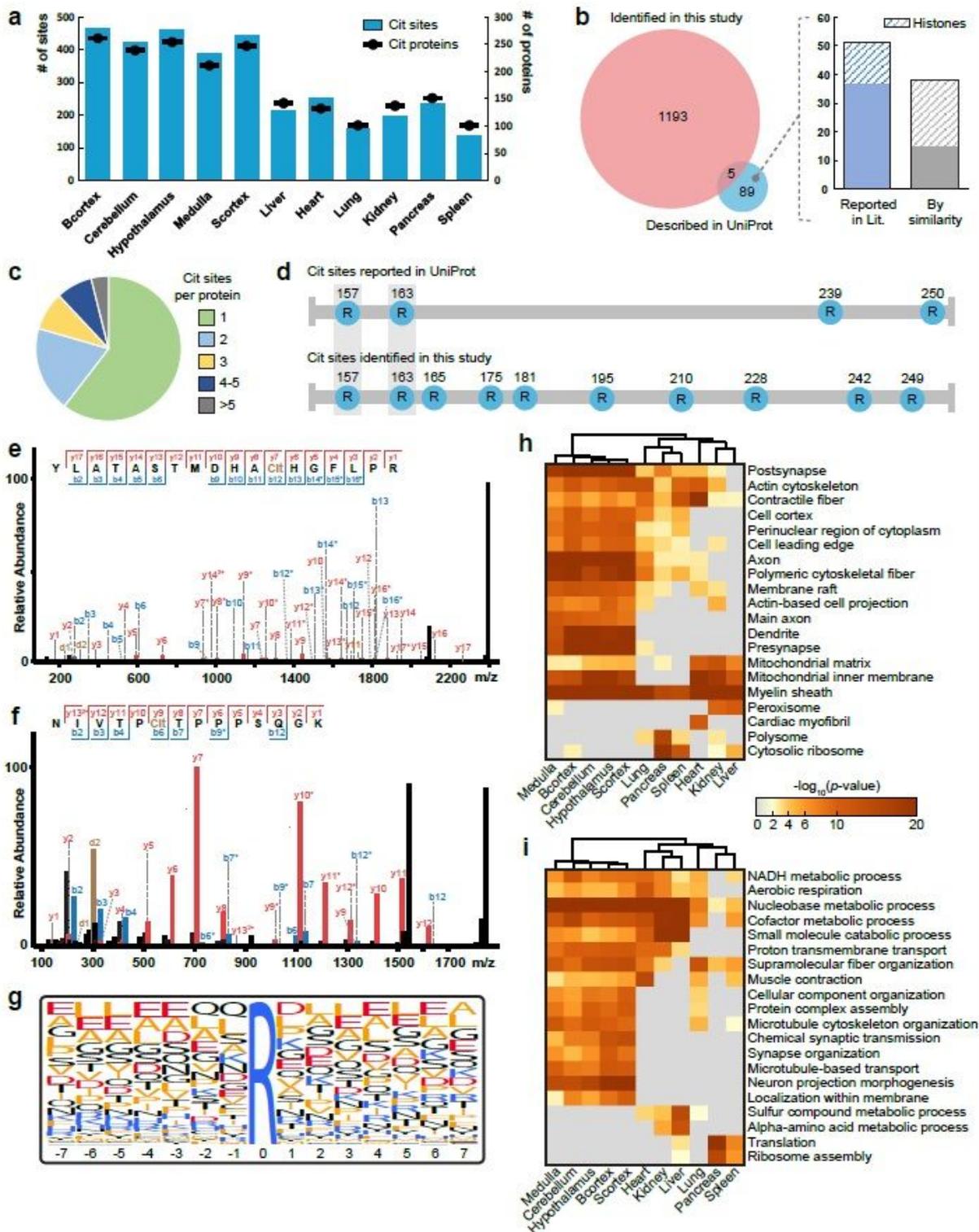


Figure 3

Large-scale citrullinome profiling of different mouse tissues. a, Number of identified citrullinated proteins and citrullination sites in different mouse tissues. Identification numbers are presented as a total from three mice. Cit, citrullination; Bcortex, cerebral cortex; Scortex, hippocampus and thalamus. b, Overlap of

sequence indicate vacancy due to peptides being identified at protein termini. e, Gene ontology (GO) analysis of all identified homocitrullinated proteins showing the significantly enriched terms (Fisher's exact test, p value < 0.05, p values were adjusted with Benjamini-Hochberg correction). CC, cellular component; BP, biological process; MF, molecular function.

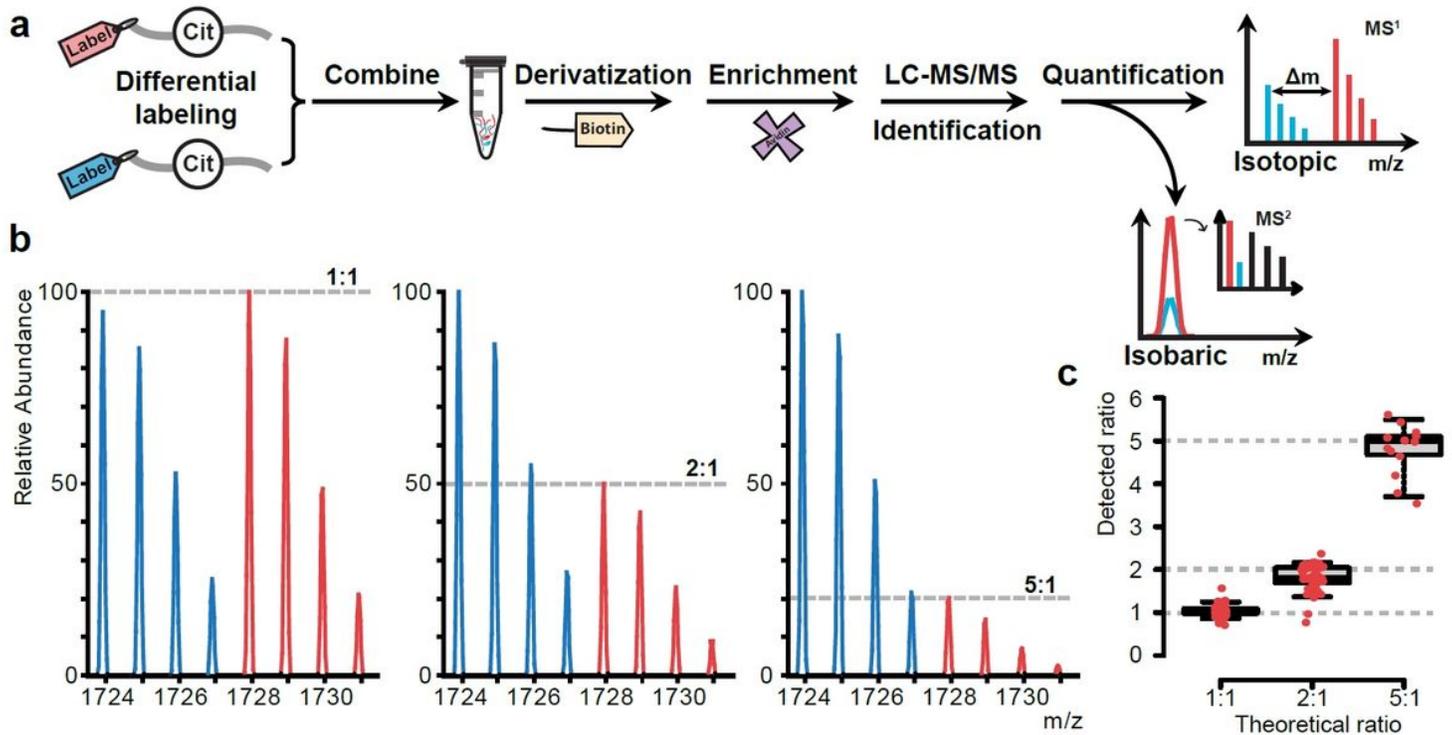


Figure 5

Multiplexed quantitative citrullination analysis using chemical labeling strategies. a, Schematic showing the pipeline for simultaneous qualitative and quantitative analysis of citrullination using isotopic or isobaric labeling. Relative quantification can be achieved during survey scans or tandem MS scans, respectively. b, Spectra showing quantitation accuracy in duplex dimethyl labeling using citrullinated peptide standard. Peptide standard was dimethylated by either heavy isotope or light isotope labeling, resulting in a 4 Da mass difference. Heavy (red) and light (blue) labels were mixed in three known ratios (1:1, 2:1 and 5:1) and subjected to derivatization using biotin thiol tag. c, Boxplots showing quantitation accuracy and precision in duplex dimethyl labeling using mouse brain digest. Red dots indicate the detected ratios for each quantified citrullinated peptides. Top and bottom of boxes indicate 3rd and 1st quartile, respectively, and whiskers extend to 95th and 5th quartile. Horizontal lines within boxes denote median.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData1.xlsx](#)

- [SupplementaryData2.xlsx](#)
- [SupplementaryData3.xlsx](#)
- [DescriptionofAdditionalSupplementaryFilesJan31.docx](#)
- [SupplementaryInformationFeb06Final.docx](#)