Advancing Quantitative Proteomics and Protein Post-translational Modification Analyses by Multi-Dimensional Mass Spectrometric Approaches

By

Zihui Li

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2021

Date of final oral examination: October 13th, 2021

The dissertation is approved by the following members of the Final Oral Committee:

Lingjun Li, Professor, Pharmacy and Chemistry

Lloyd M. Smith, Professor, Chemistry

Seungpyo Hong, Professor, Pharmacy

Jon S. Odorico, Professor, Surgery

Acknowledgements

First of all, I would like to express my sincere gratitude to my advisor Dr. Lingjun Li, without whom none of this work would be possible. I met Dr. Li back to the winter of 2015 when she was giving a mass spectrometry short course in China and I was instantly intrigued by the magic of mass spec and the diverse research in Li lab. In 2016, I came to Madison and joined the Li lab as I planned, which turns out to be the most important and correct choice I have ever made. During my five years in the lab, Dr. Li has been always caring and supportive. She always tries to provide the best resources to everyone in the lab, and she always listens to our thoughts with great patience and encourages us to explore all the possibilities in science. She is a really excellent role model in all aspects and teaches us to be independent and qualified scientists. I will forever be thankful for this opportunity to join the lab, and work and grow under her mentorship.

I would also like to thank my committee members, Dr. Lloyd Smith, Dr. Seungpyo Hong and Dr. Jon Odorico, for their time, guidance and helpful suggestions on my thesis background examination, original research proposal, 4th year committee meeting, and finally, this dissertation. Dr. Odorico is also a terrific collaborator and I am really impressed by his attitude and deep insights into science. It has been a great honor to have all of them witness my growth and development throughout my PhD career.

Furthermore, I would like to acknowledge many fantastic collaborators and I really enjoyed learning from them and discussing with them about the sparkling ideas. Dan Tremmel and Dr. Sara Sackett are unbelievable scientists from Odorico lab. Dan and I worked together for a long time

on the pancreas ECM project. We discussed data, made figures, prepared and revised manuscript and finally generated a great piece of work together. I will miss our inspiring and informative monthly meetings on Tuesday mornings. Thank Kristal Gant from Dr. Manish Patankar's lab for involving me in the great work on ovarian cancer. Thank Dr. Louise Atkinson and Dr. Angela Mousley from Queen's University Belfast in UK for the work on nematode neuropeptidomics. Thank Dr. Xudong Shi, Dr. Hui Ye and Dr. Zhengqing Ye for the great help on citrullination projects. Only through the joint effort can we accomplish so many thrilling studies.

I have benefited a lot from my mentors and have been fortunate to work with many talented colleagues in the lab. I knew Dr. Yang (Sheila) Liu when I was an undergrad and she actually introduced me to the lab. And she helped me to settle in and get familiar with the life here when I first came to Madison. She was my mentor when I joined the lab and taught me a lot about neuropeptide studies (and crabs). Dr. Yatao Shi was another great mentor and he taught me so much about proteomics, mass spectrometry imaging, instrument maintenance and cleaning, and all the other things in the lab. We have been working together on citrullination related projects almost through my entire graduate career, and we have overcome so many hurdles together. I will miss the days when we worked until midnight with full enthusiasm even on weekends. He is also a good friend outside of lab especially when seeking delicious food in Madison. Dr. Fengfei Ma also taught me a lot on proteomics and instrument operation, and introduced me into the collaboration with Odorico lab. Bin Wang has been very helpful working with me on multiple projects. He will continue to work on several citrullination projects and also take the responsibility

of instrument maintenance after I graduate. He is a good friend in life and we often have dinners together, play basketball and joke about each other. Many thanks to Dr. Min Ma, Dr. Hua Zhang, Dr. Gongyu Li, Dr. Rui Liu, Dr. Dustin Frost, Dr. Zhengwei Chen, Dr. Amanda Buchberger Jones, Dr. Xiaofang Zhong, Dr. Qinjingwen Cao, Dr. Kellen DeLaney, Dr. Pingli Wei, Dr. Yusi Cui, Graham Delafield, Dylan Tabang, Chris Sauer, Miyang Li, Yuan Liu, Zexin Zhu and all the other Li lab members for their tremendous help. They have always been wonderful colleagues and friends, and I have been so lucky to work in such a big and warm family.

I would like to thank a lot of good friends outside of the lab as they have brought so much joy to my life and made my PhD study less stressful. Dr. Yue Gui and Xin Yao have been great friends. We discovered food, played sports and traveled together, and I believe our friendship will last forever. Xingyang Fu and Hongyuan Sheng joined the chemistry program the same year as me. We taught, took classes, played and traveled together. We also encouraged and supported each other to go through the difficult times.

My final thanks go to my family. I would like to thank my colleague and my girlfriend Qinying Yu who also joined the lab the same year as me. It's fantastic to have someone who is not only willing to help with scientific problems but also share my happiness and sadness. Thank you for making my PhD life much more pleasant and memorable. My parents are best teachers and my best friends, and they always accompany me in all stages of my life. I would never be in this place without their love and support. I hope they are proud of me and continue to be so in the future.

To all those I have mentioned: again, thank you!

Table of Contents

Advancing Quantitative Proteomics and Protein Post-translational Modification Analyses by Multi-Dimensional Mass Spectrometric Approaches

Zihui Li

Under the supervision of Professor Lingjun Li

At the University of Wisconsin-Madison

Abstract

Proteins are important molecules that are involved in numerous critical biological processes. Consequently, understanding the functional roles and dysregulation of proteins constitutes a vital task in both basic and clinical research. However, it is very challenging to define the full complement of proteins due to their large degree of variations, especially considering that many proteins can undergo numerous post-translational modifications (PTMs) which dramatically affect protein conformations and functions. Recently, mass spectrometry (MS) has evolved as a powerful tool to characterize proteins and their PTMs benefiting from its high speed, sensitivity and the ability to monitor thousands of analytes simultaneously.

This dissertation is devoted to the development and application of various MS-based strategies to facilitate the detection of proteins and their PTMs qualitatively and quantitatively. A portion of this dissertation describes the investigation of proteome-wide alterations with the development and maturation of human pancreas using custom-developed isobaric tags. This study focuses on the extracellular matrix proteins and provides information on the localization changes

of selected proteins in islet or acinar compartments as well. The normal baseline data across the lifespan of human pancreas will become a valuable resource and benefit future studies into the pancreatic pathologies. Switching gear to protein PTMs, this dissertation documents the design and development of a novel biotin thiol tag to greatly improve the in-depth profiling of two important PTMs, citrullination and homocitrullination, from complex biological samples. The utility of the approach is demonstrated by providing the first tissue-specific citrullination and homocitrullination atlas in mice. The first high-throughput quantitative analysis pipeline of protein citrullination is subsequently developed by the integration with isobaric labeling strategy. We also further improve the protein citrullination detection by using a combination of various digestion and fragmentation techniques. Citrullination alteration during the progression of Alzheimer's disease is further explored using human cerebrospinal fluid to shed light on the disease pathogenesis. This dissertation also highlights the application of a novel sub-atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry platform for the *in situ* imaging of N-glycans. Spatial distribution of N-glycans in ovarian cancer tissue section may indicate potential association of this modification with tumor progression.

Collectively, we expect the biological insights and technological advancements presented in this work will lead to improved understanding of microenvironment in both healthy and disease conditions, and the new knowledge gained from these studies will inspire and benefit future research effort in related fields.

Chapter 1

Introduction and research summary

Introduction and research summary

Proteins are basic building blocks of an organism that are deeply involved in the major processes of life.[1] Consequently, understanding the functional roles and dysregulation of proteins constitute a vital task in both basic and clinical research. However, it is very challenging to define the full complement of proteins due to their large degree of variations, especially considering that many proteins can undergo numerous site-specific changes including post-translational modifications (PTMs), which dramatically increases the potential number of functional proteins, or proteoforms.[1] PTMs refer to the covalent and generally enzymatic modification of proteins following protein biosynthesis. These modifications include phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation and proteolysis and influence almost all aspects of normal cell biology and pathogenesis.[2-4] Therefore, identifying and understanding PTMs, although challenging, can provide invaluable insights into the cellular functions underlying etiological processes. Mass spectrometry (MS) has become a powerful tool in biomolecular analysis and is widely used in proteomics and PTM studies owing to its high speed, high sensitivity and resolution, and the ability to detect thousands of analytes in a single run.[5-8] This dissertation is devoted to the development and application of MS-based strategies to facilitate the detection of proteins (**Chapter 2**) and their PTMs (**Chapter 3-6**) qualitatively and quantitatively. This dissertation also highlights the application of novel instrument platform to enhancing the mass spectrometry imaging (MSI) analysis of N-glycans, which provides information on the spatial distribution of certain analytes (**Chapter 7**).

Chapter 1 serves as a general introduction to the work in this dissertation and briefly introduces the background and major findings in each project.

Chapter 2 features a comprehensive investigation of proteome-wide alterations and localization changes of extracellular matrix proteins (ECM) across the lifespan of human pancreas.[9] ECM or the matrisome is the network of proteins and polysaccharides surrounding cells, which is a key component for maintaining optimal cell and tissue homeostasis through its involvement in many intracellular events.[10-12] The pancreas is a glandular organ which functions importantly in both digestive and endocrine systems. However, how the pancreatic ECM dynamically changes in humans over the lifespan has not been previously described. In this study, we apply MS-based quantitative proteomics strategies using *N,N*-dimethyl leucine (DiLeu) isobaric tags[13-14] to delineating proteome-wide and ECM-specific alterations in four age groups of healthy human pancreas: fetal (18-20 weeks gestation), juvenile (5-16 years old), young adults (21-29 years old) and older adults (50-61 years old). We also visualize specific ECM proteins of interest using immunofluorescent staining and investigate changes in ECM localization within islet or acinar compartments. This study presents new knowledge and the most comprehensive proteomics analysis across the life cycle of human pancreas development and maturation to date. For example, we reveal dynamic changes in pancreatic matrisome localization, and identify specific ECM proteins enriched in pancreatic islets (e.g., COL6A1, EMILIN1, FBN2, OGN), which provides new insights for studying islet development, function and disease. Our study generates a large dataset that could be utilized to examine other facets. For instance, we find that previously reported tumor biomarkers (e.g., THBS1)[15] are observed to display significant changes in protein expression levels with age, which could be important knowledge for studies in biomarker discovery and clinical implementation.

Beyond global proteomic analysis of ECM, another major focus of my PhD dissertation work has been on PTM studies. **Chapters 3-6** introduce my effort to improve the detection of two

important PTMs, citrullination and homocitrullination, by developing novel chemical proteomics approaches and utilizing various MS-based techniques. Protein citrullination results from the conversion of the guanidium group to ureido group on arginine residues and is catalyzed by a calcium-regulated family of enzymes called protein arginine deiminases (PADs).[16-19] Protein homocitrullination is another chemically related PTM that occurs on lysine side chains and is known as a nonenzymatic PTM.[20] These two types of PTMs lead to the loss of positive charges on the basic amino acid residues under physiological conditions, and therefore have profound effects on protein conformations, protein-protein interactions and protein functions.[16-17, 21] Recent accumulating evidence has revealed that dysregulation of citrullination and homocitrullination is associated with the development of diverse pathological states including rheumatoid arthritis[20, 22-28], multiple sclerosis[29-32], prion disease[33], psoriasis[34], Alzheimer's disease (AD)[35-37] and cancers[38-43], which raises a fast-growing interest in studying these two important PTMs. However, the underlying regulatory mechanisms remain poorly understood due to limited knowledge of their substrate proteins and modification sites, which is hampered by the lack of effective analytical tools. Antibody-based techniques are currently the most prevalent methods to detect these PTMs.[44-46] However, these approaches are neither suitable for high-throughput analysis nor able to pinpoint exact sites of the PTMs with confidence.[47-48] The application of MS to mapping the citrullination and homocitrullination proteome also suffers from several challenges.[47-51] 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 ^{13}C isotopic peaks (+ 1.0033 Da). To combat these issues, tremendous effort has been devoted to improving aspects of the analytical workflow such as careful manual examination of the spectra[28, 52-53], developing

delicate searching algorithms and statistical modelings[54-58], developing chemical derivatization methods[59-61] and chemical probes[24, 62-63]. Unfortunately, none of these methods can provide confident identification results with reasonable input of time and resource.

In **Chapter 3**, we design a novel biotin thiol tag that enables derivatization, enrichment and identification of these two PTMs simultaneously. We apply this method to mapping the citrullination and homocitrullination proteome from five brain regions and six body organs in mice. In total, we identify 691 citrullination sites and 81 homocitrullination sites from 432 and 63 proteins, respectively, which represents the largest dataset to date. Our results significantly expand our current understanding of these PTMs by mapping their wide distribution than hitherto anticipated and their participation in previously unknown biological processes. For example, we reveal their important roles in central nervous system and many metabolic processes including respiration and glycolysis. Correspondingly, modified proteins are observed to concentrate in axon, myelin sheath, synapse and mitochondria. Functional implications are exemplified by firstly reported citrullination or homocitrullination sites on proteins including myelin basic protein, glial fibrillary acidic protein, apolipoprotein E, microtubule-associated protein tau, pyruvate kinase and histones. We also achieve multiplexing quantitative analysis by the integration with isotopic labeling techniques. This study depicts a landscape of protein citrullination and homocitrullination and lays the foundation for further deciphering their physiological and pathological roles.

Continuing the work described in the previous chapter, **Chapter 4** describes the integration of this recently developed method with our custom-developed 12-plex DiLeu isobaric tags to achieve a high-throughput quantitative analysis of protein citrullination for the first time. We first demonstrate the efficacy of this pipeline using standard peptide and further extend its utility to provide accurate and precise quantitation when analyzing complex samples. We then apply this

strategy to investigating how protein citrullination is involved in DNA damage response in human cell lines and reveal the vital and complex roles of citrullination during this process. Specifically, we observe hypercitrullination on some DNA repair proteins (e.g., RAD50) and RNA binding proteins (e.g., HNRNPs) in the treated groups. And many of these citrullination sites are located in important domains such as RNA binding domain or nuclear targeting domain, suggesting that protein citrullination actively participates in DNA damage through interfering normal translational activities and DNA repair processes. Our study provides new insights into DNA damage-associated disease pathogenesis from the perspective of protein citrullination and offers a novel yet easy-to-adapt quantitative approach which will greatly facilitate future investigations into this important PTM.

Chapter 5 aims to further push the boundary of citrullination detection by incorporating different digestion methods (trypsin and LysargiNase) and fragmentation techniques (HCD and ETD). We first explore different digestion protocols and observe the optimal results when using a combination of LysN and LysargiNase enzyme mixture. We then demonstrate the distinct fragmentation patterns of citrullinated peptides from the LysargiNase digestion sample and its orthogonality in detecting protein citrullination compared with trypsin digestion. Benefiting from various digestion and fragmentation techniques, we identify 305 citrullination sites on 228 proteins from mouse brain, which is the most in-depth analysis from a single tissue type. Our results indicate intimate involvement of this PTM in many important functions such as regulation of structural integrity, neuronal signal transmission and metabolic processes including respiration. More importantly, our novel method provides an attractive complement to enhance the profiling of protein citrullination and will greatly facilitate future investigations to understand the critical roles of this PTM *in vivo*.

Chapter 6 describes the investigation of citrullination changes during the progression of AD. We perform citrullination profiling of human cerebrospinal fluid (CSF) from healthy individuals or dementia patients, and each group contains four specimens. We identify 135 citrullination sites from 68 citrullinated proteins, the majority of which have not been reported before. We observe an overall lower expression level of citrullination in AD compared to healthy individuals, which may provide new insights into the disease pathogenesis. We confidently identify nine citrullination sites on apolipoprotein E, a previously well-studied protein with elevated risk associated with AD. Three of these citrullination sites are only identified in healthy individuals and four of them show significantly decreased expression in AD patients, which indicates an altered PTM expression and function of this protein with the onset of AD. In another scenario, six citrullination sites are identified on clusterin and seven are identified on complement C3, which are other two important proteins related to AD pathogenesis and development. Collectively, this is a pilot study to investigate the association of citrullination with the progression of AD in a large-scale manner. This study pinpoints some potential citrullination biomarkers with the progression of AD and will reveal the pathogenesis and inspire future therapeutic approaches for treating this devastating disease.

Chapter 7 discusses the application of a novel sub-atmospheric pressure (SubAP)/matrix assisted laser desorption/ionization (MALDI) instrument platform for *in situ* imaging of N-linked glycans from formalin-fixed paraffin-embedded (FFPE) tissue sections. N-linked glycosylation is a type of PTMs which plays important roles in many diseases.[64-65] Although electrospray ionization (ESI)-based glycomic and glycoproteomic analyses have been applied to investigating N-glycan expression changes between healthy and disease specimens[66-67], the homogenization of tissue samples inevitably results in the loss of spatial information regarding N-glycan localization.

MALDI-MS has been introduced for *in situ* biomolecular analysis due to its capability to retain spatial information and directly visualize biomolecules in tissue sections.[68-69] In this study, we take advantage of a novel SubAP/MALDI MS platform to achieve an improved imaging analysis of N-glycans. Compared with traditional vacuum MALDI source, the SubAP/MALDI source can be easily switched back and forth with an ESI source on the same MS instrument, providing greater flexibility to allow for complementary ESI- and MALDI-based analysis. Furthermore, this new platform offers MS imaging acquisition with higher spatial resolution and better sensitivity towards specific analytes. We apply this novel platform to imaging N-glycans on mouse FFPE brain sections and ovarian cancer tissue sections. High-resolution N-glycan images with different distribution patterns are collected to shed light on the involvement of the N-glycan modification in both physiological and pathological processes.

Chapter 8 summarizes results and findings from each previous chapter and presents either ongoing or potential future directions of our work.

References

(1) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schluter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlen, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschlager, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B. *Nat Chem Biol* **2018,** *14*, 206-214.

(2) Wang, Y. C.; Peterson, S. E.; Loring, J. F. *Cell research* **2014,** *24*, 143-60.

(3) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. *Nat Rev Nephrol* **2019,** *15*, 346-366.

(4) Cohen, P. *Eur. J. Biochem.* **2001,** *268*, 5001-10.

(5) Aebersold, R.; Mann, M. *Nature* **2016,** *537*, 347-55.

(6) Han, X.; Aslanian, A.; Yates, J. R., 3rd *Curr Opin Chem Biol* **2008,** *12*, 483-90.

(7) Doll, S.; Burlingame, A. L. *ACS chemical biology* **2015,** *10*, 63-71.

(8) Olsen, J. V.; Mann, M. *Mol Cell Proteomics* **2013,** *12*, 3444-52.

(9) Li, Z.; Tremmel, D. M.; Ma, F.; Yu, Q.; Ma, M.; Delafield, D. G.; Shi, Y.; Wang, B.; Mitchell, S. A.; Feeney, A. K.; Jain, V. S.; Sackett, S. D.; Odorico, J. S.; Li, L. *Nat. Commun.* **2021,** *12*, 1020.

(10) Bonnans, C.; Chou, J.; Werb, Z. *Nat Rev Mol Cell Biol* **2014,** *15*, 786-801.

(11) Kular, J. K.; Basu, S.; Sharma, R. I. *J Tissue Eng* **2014,** *5*, 2041731414557112.

(12) Phillip, J. M.; Aifuwa, I.; Walston, J.; Wirtz, D. *Annu Rev Biomed Eng* **2015,** *17*, 113-141.

(13) Frost, D. C.; Greer, T.; Li, L. *Anal. Chem.* **2015,** *87*, 1646-54.

(14) Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L. *Anal. Chem.* **2010,** *82*, 2817-25.

(15) Daubon, T.; Leon, C.; Clarke, K.; Andrique, L.; Salabert, L.; Darbo, E.; Pineau, R.; Guerit, S.; Maitre, M.; Dedieu, S.; Jeanne, A.; Bailly, S.; Feige, J. J.; Miletic, H.; Rossi, M.; Bello, L.; Falciani, F.; Bjerkvig, R.; Bikfalvi, A. *Nat. Commun.* **2019,** *10*, 1146.

(16) Fuhrmann, J.; Clancy, K. W.; Thompson, P. R. *Chem. Rev.* **2015,** *115*, 5413-61.

(17) Fuhrmann, J.; Thompson, P. R. *ACS chemical biology* **2016,** *11*, 654-68.

(18) Witalison, E.; Thompson, P.; Hofseth, L. *Current Drug Targets* **2015,** *16*, 700-710.

(19) Mondal, S.; Thompson, P. R. *Acc Chem Res* **2019,** *52*, 818-832.

(20) Pruijn, G. J. *Front Immunol* **2015,** *6*, 192.

(21) Gyorgy, B.; Toth, E.; Tarcsa, E.; Falus, A.; Buzas, E. I. *The international journal of biochemistry & cell biology* **2006,** *38*, 1662-77.

(22) Schellekens, G. A.; de Jong, B. A.; van den Hoogen, F. H.; van de Putte, L. B.; van Venrooij, W. J. *The Journal of clinical investigation* **1998,** *101*, 273-81.

(23) Elkon, K. B. *Sci Transl Med* **2013,** *5*, 209fs39.

(24) Tutturen, A. E.; Fleckenstein, B.; de Souza, G. A. *Journal of proteome research* **2014,** *13*, 2867-73.

(25) Turunen, S.; Huhtakangas, J.; Nousiainen, T.; Valkealahti, M.; Melkko, J.; Risteli, J.; Lehenkari, P. *Arthritis Res Ther* **2016,** *18*, 239.

(26) Tilvawala, R.; Nguyen, S. H.; Maurais, A. J.; Nemmara, V. V.; Nagar, M.; Salinger, A. J.; Nagpal, S.; Weerapana, E.; Thompson, P. R. *Cell Chem Biol* **2018,** *25*, 691-704 e6.

(27) Fert-Bober, J.; Darrah, E.; Andrade, F. *Immunol Rev* **2020,** *294*, 133-147.

(28) Raijmakers, R.; van Beers, J. J.; El-Azzouny, M.; Visser, N. F.; Bozic, B.; Pruijn, G. J.; Heck, A. J. *Arthritis Res Ther* **2012,** *14*, R114.

(29) Moscarello, M. A.; Mastronardi, F. G.; Wood, D. D. *Neurochem Res* **2007,** *32*, 251-6.

(30) Bradford, C. M.; Ramos, I.; Cross, A. K.; Haddock, G.; McQuaid, S.; Nicholas, A. P.; Woodroofe, M. N. *J Neuroimmunol* **2014,** *273*, 85-95.

(31) Yang, L.; Tan, D.; Piao, H. *Neurochem Res* **2016,** *41*, 1845-56.

(32) Gs Chirivi, R. *J. Clin. Cell. Immunol.* **2013,** *04*.

(33) Jang, B.; Kim, E.; Choi, J. K.; Jin, J. K.; Kim, J. I.; Ishigami, A.; Maruyama, N.; Carp, R. I.; Kim, Y. S.; Choi, E. K. *The American journal of pathology* **2008,** *173*, 1129-42.

(34) Ishida-Yamamoto, A.; Senshu, T.; Takahashi, H.; Akiyama, K.; Nomura, K.; Iizuka, H. *J Invest Dermatol* **2000,** *114*, 701-5.

(35) Ishigami, A.; Ohsawa, T.; Hiratsuka, M.; Taguchi, H.; Kobayashi, S.; Saito, Y.; Murayama, S.; Asaga, H.; Toda, T.; Kimura, N.; Maruyama, N. *Journal of neuroscience research* **2005,** *80*, 120-8.

(36) Acharya, N. K.; Nagele, E. P.; Han, M.; Coretti, N. J.; DeMarshall, C.; Kosciuk, M. C.; Boulos, P. A.; Nagele, R. G. *J Autoimmun* **2012,** *38*, 369-80.

(37) Ishigami, A.; Masutomi, H.; Handa, S.; Nakamura, M.; Nakaya, S.; Uchida, Y.; Saito, Y.; Murayama, S.; Jang, B.; Jeon, Y. C.; Choi, E. K.; Kim, Y. S.; Kasahara, Y.; Maruyama, N.; Toda, T. *Journal of neuroscience research* **2015,** *93*, 1664-74.

(38) Yuzhalin, A. E.; Gordon-Weeks, A. N.; Tognoli, M. L.; Jones, K.; Markelc, B.; Konietzny, R.; Fischer, R.; Muth, A.; O'Neill, E.; Thompson, P. R.; Venables, P. J.; Kessler, B. M.; Lim, S. Y.; Muschel, R. J. *Nat. Commun.* **2018,** *9*, 4783.

(39) Yuzhalin, A. E. *Cancer Res* **2019,** *79*, 1274-1284.

(40) Stadler, S. C.; Vincent, C. T.; Fedorov, V. D.; Patsialou, A.; Cherrington, B. D.; Wakshlag, J. J.; Mohanan, S.; Zee, B. M.; Zhang, X.; Garcia, B. A.; Condeelis, J. S.; Brown, A. M.; Coonrod, S. A.; Allis, C. D. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110*, 11851-6.

(41) Chang, X.; Han, J.; Pang, L.; Zhao, Y.; Yang, Y.; Shen, Z. *BMC Cancer* **2009,** *9*, 40.

(42) Wang, L.; Song, G.; Zhang, X.; Feng, T.; Pan, J.; Chen, W.; Yang, M.; Bai, X.; Pang, Y.; Yu, J.; Han, J.; Han, B. *Cancer Res* **2017,** *77*, 5755-5768.

(43) Thalin, C.; Lundstrom, S.; Seignez, C.; Daleskog, M.; Lundstrom, A.; Henriksson, P.; Helleday, T.; Phillipson, M.; Wallen, H.; Demers, M. *PloS one* **2018,** *13*, e0191231.

(44) Senshu, T.; Sato, T.; Inoue, T.; Akiyama, K.; Asaga, H. *Anal. Biochem.* **1992,** *203*, 94-100.

(45) Nicholas, A. P.; King, J. L.; Sambandam, T.; Echols, J. D.; Gupta, K. B.; McInnis, C.; Whitaker, J. N. *The Journal of comparative neurology* **2003,** *459*, 251-66.

(46) Moelants, E. A.; Van Damme, J.; Proost, P. *PloS one* **2011,** *6*, e28976.

(47) Verheul, M. K.; van Veelen, P. A.; van Delft, M. A. M.; de Ru, A.; Janssen, G. M. C.; Rispens, T.; Toes, R. E. M.; Trouw, L. A. *Autoimmun Rev* **2018,** *17*, 136-141.

(48) Hensen, S. M.; Pruijn, G. J. *Mol Cell Proteomics* **2014,** *13*, 388-96.

(49) Clancy, K. W.; Weerapana, E.; Thompson, P. R. *Curr Opin Chem Biol* **2016,** *30*, 1-6.

(50) Vitorino, R.; Guedes, S.; Vitorino, C.; Ferreira, R.; Amado, F.; Van Eyk, J. E. *Journal of proteome research* **2020**.

(51) De Ceuleneer, M.; Van Steendam, K.; Dhaenens, M.; Deforce, D. *Proteomics* **2012,** *12*, 752-60.

(52) Lee, C. Y.; Wang, D.; Wilhelm, M.; Zolg, D. P.; Schmidt, T.; Schnatbaum, K.; Reimer, U.; Ponten, F.; Uhlen, M.; Hahne, H.; Kuster, B. *Mol Cell Proteomics* **2018,** *17*, 1378-1391.

(53) Stobernack, T.; Glasner, C.; Junker, S.; Gabarrini, G.; de Smit, M.; de Jong, A.; Otto, A.; Becher, D.; van Winkelhoff, A. J.; van Dijl, J. M. *Journal of proteome research* **2016,** *15*, 4532-4543.

(54) Fert-Bober, J.; Venkatraman, V.; Hunter, C. L.; Liu, R.; Crowgey, E. L.; Pandey, R.; Holewinski, R. J.; Stotland, A.; Berman, B. P.; Van Eyk, J. E. *Journal of proteome research* **2019**.

(55) Wang, X.; Swensen, A. C.; Zhang, T.; Piehowski, P. D.; Gaffrey, M. J.; Monroe, M. E.; Zhu, Y.; Dong, H.; Qian, W. J. *Journal of proteome research* **2020,** *19*, 1863-1872.

(56) Villacres, C.; Spicer, V.; Krokhin, O. V. *Journal of proteome research* **2021**.

(57) Huh, S.; Hwang, D.; Kim, M. S. *Anal. Chem.* **2020,** *92*, 12975-12986.

(58) De Ceuleneer, M.; Van Steendam, K.; Dhaenens, M.; Elewaut, D.; Deforce, D. *Journal of proteome research* **2012,** *11*, 5245-51.

(59) Stensland, M.; Holm, A.; Kiehne, A.; Fleckenstein, B. *Rapid Commun Mass Spectrom* **2009,** *23*, 2754-62.

(60) De Ceuleneer, M.; De Wit, V.; Van Steendam, K.; Van Nieuwerburgh, F.; Tilleman, K.; Deforce, D. *Rapid Commun Mass Spectrom* **2011,** *25*, 1536-42.

(61) Choi, M.; Song, J. S.; Kim, H. J.; Cha, S.; Lee, E. Y. *Analytical biochemistry* **2013,** *437*, 62-7.

(62) Tutturen, A. E.; Holm, A.; Fleckenstein, B. *Anal Bioanal Chem* **2013,** *405*, 9321-31.

(63) Lewallen, D. M.; Bicker, K. L.; Subramanian, V.; Clancy, K. W.; Slade, D. J.; Martell, J.; Dreyton, C. J.; Sokolove, J.; Weerapana, E.; Thompson, P. R. *ACS chemical biology* **2015,** *10*, 2520-8.

(64) Lau, K. S.; Dennis, J. W. *Glycobiology* **2008,** *18*, 750-60.

(65) de Leoz, M. L.; Young, L. J.; An, H. J.; Kronewitter, S. R.; Kim, J.; Miyamoto, S.; Borowsky, A. D.; Chew, H. K.; Lebrilla, C. B. *Mol Cell Proteomics* **2011,** *10*, M110 002717.

(66) Sun, S.; Shah, P.; Eshghi, S. T.; Yang, W.; Trikannad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Hoti, N.; Zhang, Z.; Chan, D. W.; Zhang, H. *Nat Biotechnol* **2016,** *34*, 84-8.

(67) Chen, Z.; Yu, Q.; Hao, L.; Liu, F.; Johnson, J.; Tian, Z.; Kao, W. J.; Xu, W.; Li, L. *Analyst* **2018,** *143*, 2508-2519.

(68) Everest-Dass, A. V.; Briggs, M. T.; Kaur, G.; Oehler, M. K.; Hoffmann, P.; Packer, N. H. *Mol Cell Proteomics* **2016,** *15*, 3003-16.

(69) Holst, S.; Heijs, B.; de Haan, N.; van Zeijl, R. J.; Briaire-de Bruijn, I. H.; van Pelt, G. W.; Mehta, A. S.; Angel, P. M.; Mesker, W. E.; Tollenaar, R. A.; Drake, R. R.; Bovee, J. V.; McDonnell, L. A.; Wuhrer, M. *Anal. Chem.* **2016,** *88*, 5904-13.

Chapter 2

Proteome-wide and matrisome-specific alterations during human pancreas development and maturation

Adapted from: **Li, Z.**[#]; Tremmel, D. M.[#]; Ma, F.; Yu, Q.; Ma, M.; Delafield, D. G.; Shi, Y.; Wang, B.; Mitchell, S. A.; Feeney, A. K.; Jain, V. S.; Sackett, S. D.; Odorico, J. S.; Li, L., Proteome-wide and matrisome-specific alterations during human pancreas development and maturation. *Nature Communications* **2021,** *12* (1), 1020. DOI https://doi.org/10.1038/s41467-021-21261-w

Abstract

The extracellular matrix (ECM) is unique to each tissue and capable of guiding cell differentiation, migration, morphology, and function. The ECM proteome of different developmental stages has not been systematically studied in the human pancreas. In this study, we apply mass spectrometry -based quantitative proteomics strategies using *N*,*N*-dimethyl leucine isobaric tags to delineate proteome-wide and ECM-specific alterations in four age groups: fetal (18-20 weeks gestation), juvenile (5-16 years old), young adults (21-29 years old) and older adults (50-61 years old). We identify 3,523 proteins including 185 ECM proteins and quantify 117 of them. We detect previously unknown proteome and matrisome features during pancreas development and maturation. We also visualize specific ECM proteins of interest using immunofluorescent staining and investigate changes in ECM localization within islet or acinar compartments. This comprehensive proteomics analysis contributes to an improved understanding of the critical roles that ECM plays throughout human pancreas development and maturation.

Introduction

The extracellular matrix (ECM) is the network of proteins and polysaccharides surrounding cells, forming a niche in which cells reside and function. Every tissue has a unique composition and arrangement of ECM components, and ECM is known to be significantly remodeled throughout development, with aging, and during various disease states[1-3]. During development, the ECM changes in both composition and organization, guiding cell migration and influencing cell fate[1]. In cell physiology, ECM plays a multitude of roles, from providing structural and mechanical support, to modifying growth factor diffusion and binding and providing anchorage-dependent cell survival signals[4-5]. Modern approaches in regenerative medicine and tissue engineering aim to utilize or recapitulate ECM scaffolding and signaling[6-7], but only recently have studies been able to extensively define the matrisome of human tissues[8], and studies of ECM composition throughout human development are even more limited. Understanding the complexity of ECM in normal tissue, throughout distinct developmental time points, can provide important context for understanding healthy physiology, identifying and characterizing disease states, improving cell and tissue isolation methods, and recapitulating physiological tissues or development *in vitro*.

The pancreas is a glandular organ which functions in both digestive and endocrine systems. The exocrine cells in the acini secrete digestive enzymes into a ductal system eventually emptying into the duodenum. The islets of Langerhans are clusters of five endocrine cell types in close contact with capillary networks, which secrete hormones into the blood. Pancreatic ECM has been studied in many non-human species and across developmental time frames, and has been found to have significant compositional and structural differences among species and from fetal to adult tissues[9]. The ECM of the pancreas is also known to be heavily altered in states of fibrotic pancreatitis and with the progression of pancreatic cancers[10-12]. Islets are surrounded by a capsule

consisting of fibroblasts and ECM[13-14]; along the vasculature that penetrates the islets, there is a double basement membrane between endocrine cells and islet capillaries, which is unique to human islets[15]. Furthermore, islet and capillary ECM is altered in the progression of both Type 1 and Type 2 diabetes[16-19], and found to play important roles in the survival and function of insulin-secreting beta cells[9, 18, 20]. The role of the ECM in the pancreas, and specifically within islets, is of key interest in the fields of diabetes and beta cell replacement.

To better understand the usefulness of ECM for regenerative medicine and tissue engineering applications, it is essential to have a more complete and comprehensive understanding of the ECM within normal, native tissue. Previous data about the pancreatic matrisome is abundant, but often incomplete and inconsistent. Ample data has been derived from various animal species, but variation among species is well documented[14, 21] and even the architecture of the islets themselves differs dramatically between rodents and primates[15, 22-27]. Until recently, most data on human pancreas and islet ECM has been obtained through immunohistological staining, and thus has not comprehensively identified the compositional profile of ECM proteins. As mass spectrometry (MS) becomes an essential tool for detecting biomolecules, MS-based proteomics shows great potential for the sensitive and large-scale analysis of complex biological systems. Characterization of ECM proteins using MS, however, still faces challenges due to their unique biophysical and biochemical properties. The large dynamic range of the analytes renders it relatively intractable to in-depth analysis, and heavy crosslinking between ECM and other components results in low solubility and therefore decreased identification rates[28]. To overcome these difficulties, various orthogonal separation techniques were utilized prior to liquid chromatography (LC)-MS/MS to decrease sample complexity including off-gel electrophoresis[29], basic reversed-phase liquid chromatography[30] and strong cation exchange (SCX)[31]. Different

protein extraction, ECM enrichment and digestion methods have also been explored to improve ECM coverage[32-38]. As most of these approaches require extensive sample extraction or fractionation, the higher cost of additional sample handling and lengthy LC-MS/MS runs should be considered and balanced against improved proteome coverage[39].

To compare the relative abundance of ECM proteins from different samples, label-free quantification (LFQ), which is based on measuring precursor ion intensities or spectral counting, has been broadly used in related studies[8, 40]. Though easy to perform, LFQ methods demand longer instrument time and the quantification is less accurate due to run-to-run variation[41]. Label-based approaches, on the other hand, allow for accurate and multiplexing quantitative analysis. Among these labeling methods, commercially available isobaric tags such as tandem mass tag (TMT)[11, 42] or isobaric tags for relative and absolute quantitation (iTRAQ)[43-45] have been more widely used recently. Nevertheless, the utility of these commercial isobaric tags in large-scale, discovery proteomics studies is often hampered by the limited multiplexing capacity and the high price of the reagent kits. To address these limitations, we have developed a cost-effective alternative, based on a set of *N*,*N*-dimethyl leucine (DiLeu) isobaric tags that can be synthesized in-house at high yield, with three steps using commercially available reagents, at a fraction of the cost. DiLeu isobaric tags were initially demonstrated as a 4-plex set[46], and then expanded to an 8-plex set[47]. The multiplexing capacity was then increased three-fold from 4-plex to 12-plex by taking advantage of the mass-defect feature of stable isotopes to enable simultaneous quantification of 12 samples via 12 reporter ions spanning from 115-118 *m/z* [48].

Utilizing the aforementioned methods, a few studies have investigated ECM proteins in isolated mouse islets[45], decellularized rat pancreatic tissue[49], and human pancreas with pancreatic ductal adenocarcinoma and pancreatitis[11]. Additionally, we recently reported a MS-proteomic

ECM analysis comparing human fetal and adult pancreas using a sample preparation protocol known as the surfactant and chaotropic agent assisted sequential extraction/on pellet digestion (SCAD)[31, 50]. However, an in-depth quantitative study of pancreatic ECM throughout human development and maturation has not yet been reported. Herein, we present a relatively simple and effective workflow for quantitative ECM analysis with our custom developed 12-plex DiLeu isobaric tags and apply this pipeline to investigating matrisome alterations at various developmental stages of human pancreatic tissue. These results add to the knowledge base in understanding the significant alterations that occur in the ECM and associated extracellular molecules throughout the human life cycle. They also provide valuable information for future regenerative medicine strategies for beta cell replacement.

Results

In-depth proteome-wide and ECM quantification using 12-plex DiLeu isobaric tagging. We performed quantitative analysis on normal, non-diabetic, non-pancreatitis human pancreatic tissue from four developmental stages including fetal (18-20 weeks gestation), juvenile (5-16 years), young adults (21-29 years) and older adults (50-61 years) (**Figure 1**) (Donor Information available in **Table S1**). The SCAD method, which takes advantage of the solubilizing power of both surfactant (SDS) and chaotropic (urea) reagents, together with on-pellet digestion, renders a more comprehensive protein extraction and improves digestion efficiency and recovery of relatively insoluble pellets[31]. We used two sets of 12-plex DiLeu tags to label these samples while a shared sample was involved in each set to allow for better normalization and quantitation. We also replaced the previously utilized SCX[31] with off-line high pH (HpH) fractionation prior to LC-MS/MS which reduced sample complexity thus not only improving proteome coverage, but also strengthening detection of low-abundance proteins[51]. In total, we identified 3,523 proteins with

high confidence and quantified 2,064 which were present across all samples (**Tables S2, S3**). Among them, 185 proteins were categorized as ECM proteins based on Human Matrisome Database[29, 52-53] and 117 were quantifiable (**Tables S4, S5**), which makes it one of the largest datasets of human pancreas matrisome. The identified proteins show a large dynamic range spanning over six orders of magnitude (**Figure S1**) that demonstrates the ability of our method to detect low-abundance molecules.

Proteome-wide alterations at different developmental stages of human pancreas. We performed hierarchical clustering of all quantified proteins to explore their profiles at different stages (**Figure 2a**). This heatmap illustrates column-wise clustering of biological replicates in either the fetal or juvenile group, suggesting larger intergroup differences than intragroup variations. Samples from young adults and older adults, on the other hand, are mixed and grouped together, which suggests that the biological environment becomes relatively stable in adulthood. Our workflow shows the capability of accurate protein quantification and our samples are representative of a range of age groups to produce conclusive results. To examine the reproducibility of samples within each group more carefully, we performed pair-wise Pearson correlation analysis and the results were summarized as a heatmap (**Figure S2a**). Intragroup samples exhibit good correlation with one another, with an average coefficient over 0.9 while shallow-color regions show that samples from different groups are more distinct. We did not observe obvious gender disparities according to dendrograms generated from hierarchical clustering, but a confirmatory conclusion might need verification with larger sample cohorts (**Figure S2b**). We then conducted one-way ANOVA analysis to compare protein abundance across multiple stages and found 1,570 proteins were significantly changed (FDR 0.05) (**Table S3**). The profiles of these proteins are depicted by hierarchical clustering (**Figure 2b**). Six clusters can be

further generated based on their changing patterns across the four developmental stages and selected biological processes and pathways are annotated for each cluster. For better pair-wise comparison between groups, we performed Student's *t* test of all combinations and graphically displayed the results using volcano plots (**Figure 2c**). Colored dots refer to significantly changed proteins (p value < 0.05) identified with a fold change greater than two. We examined more closely two pairs of samples: juvenile versus fetal, which reflects postnatal maturation, and young adults versus juvenile, which reflects post-pubertal maturation of the pancreas. Gene set variation analysis (GSVA)[54-55] was performed to reveal expression changes of functionally related genes or gene sets. Heatmaps show the significantly changed cellular components and molecular functions while the color coding indicates normalized enrichment scores in each sample (**Figure S3**). Specifically, we found some terms related to the exocrine function of pancreas, such as lipase and exopeptidase activity, were highly expressed in older age groups, which is a clear sign of organ maturation. A full list of enriched cellular components, molecular functions, biological processes, and transcription factor targets in GSVA is also provided (**Table S6**). Biological processes enriched from significantly changed proteins also indicate highly distinct molecular features between each pair (**Figure S4a, S4c**). Each node refers to an enriched term and different terms are grouped into clusters based on their similarities while the most statistically significant term represents the cluster name[56]. We also generated a chord diagram[57] for each pair to further discern how proteins change within a few processes related to pancreas functions including response to glucose, response to peptide and regulated exocytosis (**Figure S4b, S4d**). Results indicate the complexity in the regulation of these biological processes with slightly more up-regulated proteins involved. We also found that many previously reported pancreatic cancer biomarkers showed

different expression levels at various developmental stages (**Figure S5**). These findings, along with other interesting correlations not presented, warrant further investigation.

ECM remodeling of the human pancreas throughout life. Hierarchical clustering of all quantified ECM proteins suggested matrisome features unique to different age groups including fetal, juvenile, and adult (**Figure 3a**). Pairwise comparisons showed significant ECM compositional changes throughout the fetal, juvenile, and adult stages although between young and older adults very few significant differences in ECM were found (**Figure 3b**). ANOVA analysis revealed that 84 ECM proteins were significantly changed accounting for 72% of the 117 quantified ECM proteins (**Table S5, Figure S6a**). Categorical differences of the significantly changed ECM proteins between age groups were also observed. While only 7 of 19 (37%) collagens that were quantified changed in abundance among the developmental groups, the majority of ECM glycoproteins (26/35, 74%), ECM regulators (25/28, 89%) and ECM-affiliated proteins (18/20, 90%) were found to change (**Figure S6b, S6c**). The expression levels across developmental stages were compared to select individual ECM proteins that exhibited high abundance in one or more developmental groups, or that showed significant expression changes (**Figure 3c** and **Figure S7**). This analysis highlights at least four patterns of changing expression levels. A subset of collagens shows similar abundance across all stages (e.g., COL1A1, COL3A1, COL5A1, COL6A1). In contrast, a few proteins are expressed at the highest level in the fetal samples and steadily decrease in the juvenile and adult developmental transitions (e.g., COL12A1, COL14A1, FBN2, POSTN, OGN). Others maintain an equal high expression level in the fetal and juvenile stages and only start to decrease at the adult stages (e.g., COL2A1, LAMA4, EMILIN1, FN1). The final group exhibits a lower level in fetal and higher levels in postnatal pancreata (e.g., COL4A1, COL16A1, LAMA5). For four collagen proteins (COL1, COL4, COL5 COL6), multiple

isoforms were detected. COL1, COL4, and COL6 have relatively consistent ratios of the measured isoforms throughout all developmental stages studied. Interestingly, the relative ratio of COL5 isoforms changed; COL5A1 and COL5A2 slightly decreased with age, and COL5A3 became more abundant in the adult groups (**Figure S7b**).

Visualizing ECM proteins throughout human pancreas development. To confirm the trends in core ECM protein expression from the MS study (**Table S7**), we performed immunofluorescent staining to visualize a select group of 16 ECM proteins. Representative images are shown in **Figure 4**, additional images are included in **Figure S8**, and higher magnification images in **Figure S9** for better visualization of subcellular localization. Proteins found to have non-significant differences in total expression by MS among the four developmental groups (COL1A1, COL3A1, COL5A1, COL6A1) were found to also qualitatively exhibit consistent expression by immunofluorescence across developmental ages. Likewise, proteins expressed at higher relative levels during the fetal stage than postnatal stages (COL2A1, COL12A1, COL14A1, EMILIN1, FBN2, FN1, LAMA4, OGN, POSTN) and proteins expressed at lower relative levels in the fetal stage compared to expression at later stages (COL4A1, COL16A1, LAMA5) consistently displayed similar patterns by immunofluorescence staining as they did by MS. Qualitatively, the immunofluorescence staining correlates well with and validates the matrisome data.

ECM abundance and localization changes across developmental time points. The visualization of these proteins within tissue sections provides an opportunity for more precise characterization of ECM localization among the various compartments of the pancreatic tissue. Of specific interest, we quantified the expression of these proteins in the acinar regions, which makes up the majority of the pancreas volume, and the islet regions, which constitute ~1-2% of the pancreas. The ratio of signal in the islet and acinar compartments (islet/acinar) was calculated for

each image. This ratio represents the enrichment of each particular ECM protein in islets compared to the acinar for each image, but due to the normalization does not account for the total abundance of the ECM protein itself (**Figure 5a**, purple bars). Total abundance as determined by MS is included for side-by-side comparison (**Figure 5a**, black bars). Heat maps are included (**Figure 5b**) to express overall trends in the data sets.

For many ECM proteins, both abundance and localization were found to change throughout development. Only two of the selected proteins (COL2A1, COL3A1) were expressed at similar ratios in the acinar and islet compartments at all four stages. A subset of ECM proteins studied is generally enriched within the acinar compartment (COL12A1, COL14A1, COL16A1, LAMA5, POSTN). Two proteins (COL12A1, COL14A1) are abundant in fetal acinar tissue, whereas COL16A1 is more strongly expressed within the postnatal acinar compartment. Other ECM proteins are enriched in and around the islets (EMILIN1, FBN2, LAMA4, OGN). Both OGN and EMILIN1 are expressed throughout fetal tissue, but become restricted to the islets in postnatal pancreata; EMILIN1 is concentrated intracellularly in the islets, while OGN is localized to the islet extracellular space and, to a lesser extent, in the nearby acini. FBN2 is intracellularly expressed in delta cells, but not alpha or beta cells (**Figure 4** and **Figure S11b**). COL4A1 and LAMA4 are both enriched in fetal islets with an islet/acinar ratio of around 2. COL4A1 and LAMA4 proteins are localized around the blood vessels of developing islets; however, beyond the fetal stage, COL4A1 is found to be expressed more evenly throughout acinar regions. LAMA4, on the other hand, becomes more enriched within islets in older adults, as do three other ECM proteins in our study (COL1A1, COL5A1, LAMA5). Although we found no significant changes in the total abundance of those proteins between younger and older adult pancreata, there does appear to be significant accumulation in and around the islets of the older donors. Three proteins (COL14A1,

FN1 and POSTN) displayed detectable protein expression in all four groups revealed by MS, but quantifiable staining in islets and acinar was only detectable at the fetal stage. For COL14A1 and POSTN, most expression was found in the fetal mesenchyme (**Figure S8**). In the adult donors, positive staining for both COL14A1 and POSTN was concentrated around blood vessels and ducts (**Figure S11a**), with minimal extracellular expression in the islet or acinar compartments, although COL14A1 had faint intracellular staining in the acinar regions. FN1 was found extracellularly in and around the islets in the fetal tissue, but in postnatal tissues was only found faintly and intracellularly in the islets and acinar regions, compared to more robust signal around ducts and vessels, as well as sporadic extracellular staining in the parenchyma, particularly along the boundaries of pancreatic lobules.

Discussion

Although dynamic proteome changes are known to occur with aging[2-3, 58], a comprehensive analysis of the proteome and matrisome across early and late developmental stages of human pancreatic tissue has not been reported. In this study, we present a simple and cost-efficient protocol to achieve in-depth proteome-wide and ECM-specific quantitative analysis which omits labor-intensive extraction or enrichment steps. We utilized this method to systematically examine the dynamics of proteomic composition in human pancreata among four developmental age groups, including fetal, juvenile, young adults, and older adults. We quantified 2,064 proteins and found 1,570 of them were significantly changed across multiple stages which suggests a large degree of proteome remodeling through the human life cycle. The identified yet unquantifiable proteins are the result of differences between two batches of DiLeu labeling, which could reflect the distinct protein profiles between different age groups. In addition, the stochastic nature of data-dependent acquisition (DDA) sampling may also lead to some proteins being below the limit of detection and

categorized as undetected across multiple batches[59]. To alleviate this issue, one may consider alternative highly multiplexed labeling strategies such as cPILOT[60] or 27-plex TMT[61]. However, these methods are associated with increased spectra complexity and decreased identification rates which can be emerging concerns. It is also worthwhile to note that we only collected four fractions and ran two technical replicates for each sample but still achieved a relatively deep analysis. Protein coverage can be further improved by simply increasing the number of fractions if instrument resources permit. Processes or pathways related to the functions of the pancreas are mainly activated with increasing age such as pancreatic secretion and metabolic pathways. On the other hand, some cancer-related pathways are more often deactivated with age, which is likely a sign of postnatal maturation of the organ. We observed many significant compositional changes from fetal to juvenile and juvenile to young adult pancreata. Young and older adults, on the other hand, are more similar in composition, with very few significantly altered proteins. We examined some interesting targets more closely. As an example, thrombospondin 1 (THBS1) is a glycoprotein that mediates cell-to-cell and cell-to-matrix interactions which is therefore involved in tissue genesis and remodeling[62]. In our study it showed a more than two-fold increase in abundance from juvenile to young adult that is consistent with organ maturation. In addition, THBS1 has been shown to be expressed at high levels during tumor progression[62-63] and numerous thrombospondin-based therapeutic approaches have been studied[64]. However, our data suggests that conclusions or strategies based on these studies could be further strengthened by taking into consideration the age of the patient. Along this direction we found that many previously reported pancreatic cancer biomarkers showed different expression levels at various developmental stages such as ANXA2[65], PLG[65], ITGB1[65], ACTN4[66], LGALS1[66] and LAMB1[66]. ACTN4 and LGALS1, for instance, were reported to show fold changes of 1.85 and 1.75, respectively, in cancerous versus

normal tissue. Nevertheless, they also have more than 1.3-fold higher expression in young adults compared to juvenile and this difference is even more dramatic in older adults. Though few studies mentioned an age-dependent effect on biomarker protein expression, our results indicate age-related alterations and the necessity of paying particular attention to patient age in biomarker discovery or disease treatment studies.

In regard to pancreas and islet ECM, previous studies mainly focused on a subset of collagens or laminins, using techniques to quantify specific ECM proteins through gene expression and immunofluorescent staining[15, 67]. In our study, we have identified 185 ECM proteins and quantified 117 throughout four developmental stages of the human pancreas. 84 of those quantified ECM proteins were significantly changed in total abundance among the four age groups and displayed distinct expression patterns, which revealed the complexity of ECM remodeling during development and maturation. Very few previous studies reported on the changes in ECM abundance or localization among human fetal and adult pancreas tissue. Otonkoski et al. observed the distribution of laminins throughout the acinar and islet regions of fetal and adult human pancreas by immunofluorescent staining[15]. The authors found that the profiles were quite similar between the two age groups, with the major differences being that LAMA1 was expressed in fetal but not in adult islets, and LAMB2 was only expressed in adult islets. Likewise, we found that LAMB2 was expressed at significantly lower levels in the fetal tissue than in adult tissue, but not absent, while LAMA1 was only detected in fetal and juvenile tissues (**Table S4**). Extending these findings to other laminins, our results indicate a general shift in the composition of laminins between fetal and adult stages, with some laminins increasing in abundance from fetal to adult (LAMA2, LAMA5, LAMB2, LAMC3) and only LAMA4 decreasing. LAMB1 and LAMC1 did not significantly change throughout the developmental stages we measured. We further identified

ECM proteins only present in some developmental groups, which could reflect the most drastic differences between developmental stages, and these proteins were not quantifiable due to only being detected in a subset of the donors. 59 ECM-related proteins were exclusively detected in the fetal and juvenile groups and 9 were found only in adult samples (**Table S4**). For instance, FMOD, which has critical roles in the extracellular matrix organization as well as the initiation and progression of several malignancies[68], is present exclusively in the fetal and juvenile groups. On the other hand, MXRA5 that functions in an anti-inflammatory and anti-fibrotic role[69] is only detected in the adult samples. These observations can provide deeper insights into ECM dynamics during pancreas development. Overall, we have found that as a group, collagens change less dramatically throughout development, while glycoproteins and proteoglycans, which contain ECM proteins that contribute to the basement membrane (e.g. laminins, perlecan) exhibit much more dynamic changing patterns throughout development (**Figure S6c**).

A primary means of studying islet function is to use islets isolated from the pancreas through enzymatic treatment, which destroys much of the native ECM and disrupts islet architecture[70-71]. This fact has prompted many studies to supplement ECM for improved islet culture, and the limitations to the longevity of islet culture highlight the necessity of ECM for islet survival and function[9, 20, 72-73]. Studies using human fetal isolated islets have also identified a role for ECM in the differentiation, proliferation, and function of islet endocrine cells during *in utero* development[67, 74-76]. As many look to use stem cell-derived islets as a future diabetes therapy, some question the role ECM may have in the differentiation, function, and transplantation of these cells[77-82]. To better understand the islet ECM environment, we investigated the localization of selected ECM proteins in addition to total abundance in the pancreatic tissue.

Four of the ECM proteins we studied were enriched in adult islets relative to the exocrine tissue (COL6A1, EMILIN1, FBN2, OGN). Importantly, the relative abundance and distribution of these proteins was found to also change throughout development. OGN for example, is expressed throughout the acinar and islet regions of fetal and juvenile pancreata, whereas in adult pancreata it is more concentrated in islets. One previous study has found a role for OGN in islet function; whole-body OGN knockout mice have impaired glucose tolerance and reduced beta cell mass, while OGN supplemented *in vitro* or at a systemic level *in vivo*, improves beta cell function and glucose tolerance[83]. While this study focused on circulating OGN produced from osteoblastic cells, our finding that OGN is concentrated around human pancreatic islets suggests that OGN may also play a localized role in islet function. Furthermore, FBN2 has been reported as one of 13 methylated genes silenced in human pancreatic cancers[84], but to our knowledge no other studies have reported on the normal function of FBN2 in the pancreas or in islets. Interestingly, somatostatin was another of the 13 genes identified in the study[84]; we now additionally show that FBN2 is exclusively expressed within somatostatin-expressing delta cells of the islet. In addition, a role for EMILIN1 in pancreas or islet physiology has not been reported, but the protein has been identified in previous pancreas and islet matrisome studies[45, 85]. Collagen 6 co-culture has recently been shown to improve human islet survival and function[86], but it is not clear if collagen 6 is more potent than other purified collagens in enhancing islet *in vitro* function.

Our present study also revealed unexpected differences in the islet ECM between the younger adult (21-29 years) and older adult (50-61 years) groups. Four ECM proteins (COL1A1, COL5A1, LAMA4, LAMA5) were enriched within the islets in the older adults compared to younger adults, while the total abundance of these proteins in the pancreas was insignificantly different between the two groups. This suggests that the adult islet matrisome changes with age,

consistent with recent findings that younger and older adult donors have localized differences in islet ECM, which may impact the effectiveness of enzymatic digestion during islet isolation[87-88].

Research in animal models has indicated that functional changes in islet maturation occur postnatally[89], but the possible correlation with ECM structural changes in human islets has not been studied. A recent study by Arda et al. comparing gene expression and function of isolated human islets from juvenile and adult donors indicated that there are over 500 genes differentially expressed between the two age groups, with distinct histone-mediated changes in regulation of gene expression. Furthermore, the authors observed that age was correlated with functional changes. Adult islets secreted more insulin than juvenile islets under both basal and glucose-stimulated conditions, while total insulin content did not change between juvenile and adult islets[90]. Our study suggests that the islet ECM environment also changes between childhood and adulthood. Further studies into how maturing islets interact with ECM during this key transition are likely to be informative toward better defining the mechanisms of maturation, and risks of autoimmune diabetes.

Likewise, ECM studies using animal tissues and isolated islets have previously been reported, but whether these findings are relevant to human tissue is uncertain. In their 2017 study, Naba et al. identified 120 ECM proteins in isolated mouse islets, including 66 core ECM proteins[45]. The methodology employed in their study does not appear to translate to the larger and denser human pancreas, as effective efforts to isolate islets from the human pancreas inherently destroy the ECM[91]. In the present study, we have identified 61 core ECM proteins in human pancreas across the four developmental stages. Interestingly, many of the significant proteins highlighted in our study were also identified in the Naba study, indicating previously unknown similarities of important ECM components between both species. A total of 80 core ECM proteins are identified

between the two studies. Though there were similarities between species, 33 proteins, or 41% of all core ECM proteins, only appear in one of the species' datasets. This difference highlights species-specific characteristics in the pancreatic and islet matrisome that may play relevant roles in health and disease. In a separate unpublished study, all 117 ECM-associated genes that were identified in our proteomic study were also detected at the RNA level in isolated adult human islets, which may drive future efforts to quantify the human islet-specific matrisome and investigate islet-specific changes throughout development.

Overall, we present a quantitative proteomic analysis of human pancreas to delineate molecular changes, particularly ECM composition, throughout pancreas development, maturation, and aging. We found that many previously reported pancreas tumor biomarkers displayed significant changes in protein expression across multiple developmental stages and this age-dependent effect could be important knowledge for studies in biomarker identification and clinical implementation. Our study revealed dynamic changes in pancreatic ECM composition and localization throughout life cycle, and have identified specific ECM proteins enriched in pancreatic islets (e.g. COL6A1, EMILIN1, FBN2, OGN) which may provide insight for studying islet development, function and disease. In addition, our protocol can be easily adapted to achieve large-scale and in-depth quantitative analysis of ECM-containing proteome in other biological systems with relatively simple operations and low cost. Our data interpretations are far from exhaustive but instead, we expect our results to serve as a valuable resource for broad audiences with various interests and will provide a foundation for more in-depth investigations.

Methods

Human pancreas tissue preparation. Human fetal pancreas tissue was obtained from secondary sources (Advanced Biosystems Resources, Inc.) under approved Material Transfer Agreements

and with protocols approved by the University of Wisconsin's Institutional Animal Care and Use Committee (IACUC) and Institutional Review Board (IRB) (IRB Study #2013-141). ABR, Inc obtains consent in accordance with Uniform Anatomical Gift Act (UAGA) and National Organ Transplant Act (NOTA) guidelines. ABR, Inc warrants that appropriate consent for tissue donation is obtained and adequate records of such consents are maintained. In addition, that tissues are obtained with local, state, and federal laws and regulations governing the procurement of human tissue. Within 24 hours of recovery, the organs were received and cleaned of surrounding connective tissue. Small pieces of tissue were removed and fixed with 4% PFA for histology, and the majority of the pancreas was homogenized in sterile water for 3 seconds. The homogenized tissue was pelleted (16,100 xg, 5 min) and the translucent supernatant was discarded. The pellet was flash frozen and stored at -80 °C prior to further processing for MS analysis.

Juvenile and adult human pancreas tissue was procured by the University of Wisconsin Organ and Tissue Donation Services from donors with no indication of diabetes or pancreatitis, with consent obtained for research from next of kin and authorization by the University of Wisconsin-Madison Health Sciences Institutional Review Board (waiver as exempt due to research on deceased donors not being considered human subjects research). Following organ harvest, pancreata were allocated for research if deemed unfit for transplantation due to vascular damage during organ recovery, no suitable recipient, and non-ideal age or BMI. The organs were received within 24 hours of recovery and trimmed of extra-pancreatic connective tissues, including duodenum, large arteries and veins. The parenchyma was cut into 1 cm^3 cubes and frozen at -80 °C for future use, some pieces were also immediately fixed with 4% PFA for histology. One piece of frozen pancreas per donor was thawed and rinsed with 1x PBS followed by sterile water, and then manually chopped into small pieces. The pieces were immersed in sterile water and homogenized

for 3 seconds, then pelleted (16,100 xg, 5 min). Any floating lipids were removed, and the translucent supernatant was discarded. The pellet was flash frozen and stored at -80 °C. Donor information can be found in **Table S1**.

Protein extraction and digestion. Slightly modified SCAD method[50] was used to prepare all pancreas samples. Each sample was dissolved in 150 μL of extraction buffer solution (4 % SDS, 50 mM Tris buffer) and sonicated using a probe sonicator (Thermo Fisher Scientific). 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 ℃) 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. 8 M urea was added to dissolve the pellets and 50 mM Tris buffer was used to dilute the samples to a urea concentration < 1 M. On-pellet digestion was performed with LysC/trypsin (Promega) in a 50:1 ratio (protein:enzyme, w/w) at 37 ℃ 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 Fisher Scientific). 100 μg peptide was aliquoted for each sample, dried in vacuo and reconstituted in 0.5M triethylammonium bicarbonate prior to DiLeu labeling.

12-plex DiLeu labeling. Synthesis of DiLeu tags and labeling process were performed according to protocols previously described[46, 48]. Briefly, L-Leucine or isotopic L-leucine and sodium cyanoborohydride or sodium cyanoborodeuteride (2.5× molar excess to leucine) were suspended in H_2O or D_2O, and the mixture was cooled in an ice-water bath. Formaldehyde (CH_2O, 37% w/w) or isotopic formaldehyde (CD_2O or $^{13}CH_2O$, 20% w/w) (2.5× molar excess to leucine) was added

dropwise, and the mixture was stirred in an ice-water bath for 30 min. The target product was purified by flash column chromatography (MeOH/DCM) and dried in vacuo. Each isotopologue of reporter 115 and 116 requires ^{18}O exchange prior to reductive dimethylation. Leucine or isotopic leucine was dissolved in 1 N HCl $H_2{}^{18}O$ solution (pH 1) and stirred on a hot plate at 65 °C for 4 h. Following evaporation of HCl from the solution in vacuo, trace amounts of acid were removed with StratoSpheres PL-HCO_3 MP resin (Agilent) to obtain ^{18}O leucine in free base form. The identity and purity of DiLeu tags were confirmed with MS before all experiments. 1 mg of each DiLeu tag was dissolved in 100 μL of anhydrous *N*,*N*-dimethylformamide and combined with 4-(4,6-dimethoxy1,3,5-triazin-2-yl)-4-methylmorpholinium tetrafluoroborate and *N*-methylmorpholine at 0.7× molar ratios. The activation was performed by vortexing the mixture for 45 min at room temperature and supernatant was added to each sample for peptide labeling. After vortexing at room temperature for 2 h, the labeling reaction was quenched by addition of hydroxylamine to a concentration of 0.25%. The samples were then dried in vacuo, combined, and cleaned with SCX SpinTips (PolyLC) according to manufacturer's protocols.

HpH fractionation. HpH fractionation was performed on a Waters Alliance e2695 HPLC using a C18 reverse phase column (2.1 × 150 mm, 5 μm, 100 Å, PolyLC) operating at 0.2 mL/min. Mobile phase A consisted of 10 mM ammonium formate at pH 10 adjusted with ammonium hydroxide and mobile phase B consisted of 90% ACN and 10 mM ammonium format at pH 10. Separation was achieved with a gradient as following: 1 % B (0-5 min), 1-40% B (5-50 min), 40-60% B (50-54 min), 60-70% B (54-58 min), and 70-100 % B (58-59 min). Fractions were collected every 4 min and non-adjacent fractions were concatenated into four samples before being dried in vacuo for LC-MS/MS analysis.

LC-MS/MS analysis. Samples were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to a Dionex UltiMate 3000 UPLC system. Each sample was dissolved in 3% ACN, 0.1% formic acid in water before loaded onto a 75 μm inner diameter homemade microcapillary column which is packed with 15 cm of Bridged Ethylene Hybrid C18 particles (1.7 μm, 130 Å, Waters) and fabricated with an integrated emitter tip. Mobile phase A was composed of water and 0.1% formic acid while mobile phase B was composed of ACN and 0.1% formic acid. LC separation was achieved across a 100-min gradient elution of 3% to 30% mobile phase B at a flow rate of 300 nL/min. Survey scans of peptide precursors from 300 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. The top 20 precursors were then selected for HCD fragmentation with a normalized collision energy of 30, an isolation width of 1.0 Da, a resolving power of 60k, an AGC target of 5×10^4, a maximum injection time of 118 ms, and a lower mass limit of 110 m/z. Precursors were subject to dynamic exclusion for 45 s with a 10 ppm tolerance. Each sample was acquired in technical duplicates.

Data analysis. Protein identifications and quantifications were performed using Proteome Discoverer (version 2.1, Thermo Scientific). Raw files were searched against the Uniprot *H. sapiens* reviewed database (September 2018) using Sequest HT algorithm with trypsin selected as the enzyme and three missed cleavages allowed. Precursor mass tolerance of 20 ppm and a fragment mass tolerance of 0.02 Da were set for the searching. DiLeu labeling on peptide N-termini and lysine residue (+145.12801), and carbamidomethylation of cysteine residues (+57.02146 Da) were chosen as static modifications. Dynamic modifications included oxidation of methionine residues (+15.99492 Da), deamidation of asparagine and glutamine residues (+0.98402 Da) and hydroxylation on proline residues (+15.99492 Da). Search results were filtered

to 1% false discovery rate (FDR) at both peptide and protein levels. Quantitation was performed in Proteome Discoverer with a reporter ion integration tolerance of 10 ppm for the most confident centroid. Protein quantitative ratios were determined using a minimum of one quantified peptide. Reporter ion intensities were normalized through equal total peptide amount. ECM proteins were identified and classified by matching the results to Human Matrisome dataset[29]. Missing intensities were replaced using the "replace missing values from normal distribution" feature in Perseus (version 1.6.0.7)[92] prior to further processing. Two-sample Student's *t* test with a two-tailed distribution for binary comparison and one-way ANOVA analysis were conducted using Perseus. All *p* values were further adjusted by Benjamini-Hochberg correction for multiple testing. Bioinformatics analyses including Pearson correlation analysis, hierarchical clustering, protein intensity profiling, volcano plots, GSVA analysis, chord diagram[57] and box plots were achieved using R packages. Biological process network was generated using Metascape (version 3.5)[56] and exported using Cytoscape (version 3.7.1). For protein intensity profiling, biological processes and pathways were enriched using DAVID bioinformatics resources[93] with a FDR cutoff of 0.05. Four selected terms with lower *p* values were shown for each cluster and the same term was only shown once where it appeared the most significant.

To compare abundance of different proteins, iBAQ method was used embedded in MaxQuant (version 1.5.2.8). Raw files were searched against the Uniprot *H. sapiens* reviewed database (September 2018) with trypsin/P selected as the enzyme and three missed cleavages allowed. DiLeu labeling on peptide N-termini and lysine residue (+145.12801), and carbamidomethylation of cysteine residues (+57.02146 Da) were chosen as fixed modifications. Variable modifications included oxidation of methionine residues (+15.99492 Da), deamidation

of asparagine and glutamine residues (+0.98402 Da) and hydroxylation on proline residues (+15.99492 Da). The iBAQ method was enabled and all other parameters were set as default.

Immunofluorescent staining and quantification. All donor tissue was fixed with 4% paraformaldehyde and embedded in paraffin. 5-micron sections were cut and deparaffinized with xylene and ethanol. Antigen retrieval was performed for 2.5 hours at 80 °C in 10 mM sodium citrate buffer. Following washing with PBS-T (1x PBS/ 0.05% Triton X-100) and blocking for 35 min (1x PBS/ 10% BSA) at room temperature, antibody-specific staining was performed following **Table S8**. Cover slips were mounted with Fluoromount (Sigma, #F4680). Images were taken at 20x magnification using a Zeiss Axiovert 200 M microscope with AxioVision version 4.8.2.0, and analyzed using ImageJ software (ImageJ 1.53c). The color channel representing the ECM protein of interest was converted to a binary image using an auto-threshold adjustment. Images were analyzed by tracing the islets (Ins^+ regions, red) and a neighboring acinar section of about the same size and location as each islet (Ins^- regions, using nuclear arrangement to define acinar clusters); the ECM protein (green) was quantified for each image and normalized by calculating the ratio of intensity within the islet divided by the acinar compartments (islet/acinar ratio). This method is visually outlined in **Figure S10**. For every combination of antibodies, images were taken for N=3 donors in each developmental group. 1-6 islets were quantified per image, and 4-5 images per donor, for a total of 9-22 islets quantified per donor, per stain. Statistical analysis was performed with Prism 6 for Windows (version 6.07) (GraphPad Software, Inc.). Results were reported as mean values across biological replicates ± the standard deviation of the mean. For immunofluorescent staining, statistical comparisons between two groups were determined using two-tailed unpaired Student's *t* tests. A *p* value of less than 0.05 was considered significant, and Prism's recommended classification for significance was followed ($p < 0.0001$ = extremely

significant (****), $0.0001 < p < 0.001$ = extremely significant (***), $0.001 < p < 0.01$ = very significant (**), and $0.01 < p < 0.05$ = significant (*)).

Acknowledgements

This study was supported in part by grant funding from the NIH (R21AI126419, R01DK071801, RF1AG052324, P41GM108538, and 1F31DK125021-01), and Juvenile Diabetes Research Foundation (1-PNF-2016-250-S-B and SRA-2016-168-S-B). Data presented here was also in part obtained through support from an NIH/NCATS UL1TR002373 award through the University of Wisconsin Institute for Clinical and Translational Research. 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. We would also like to acknowledge the generous support of the University of Wisconsin Organ and Tissue Donation Organization who provided human pancreas for research. Our research team would like to give special thanks to the families who donated tissues for this study. We also acknowledge Sierra Raglin and the University of Wisconsin Department of Surgery Histology Core for help processing and embedding tissues. 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.

References

(1) Bonnans, C.; Chou, J.; Werb, Z. *Nat Rev Mol Cell Biol* **2014,** *15*, 786-801.

(2) Kular, J. K.; Basu, S.; Sharma, R. I. *J Tissue Eng* **2014,** *5*, 2041731414557112.

(3) Phillip, J. M.; Aifuwa, I.; Walston, J.; Wirtz, D. *Annu Rev Biomed Eng* **2015,** *17*, 113-141.

(4) Rozario, T.; DeSimone, D. W. *Dev Biol* **2010,** *341*, 126-40.

(5) Zvibel, I.; Smets, F.; Soriano, H. *Cell Transplant* **2002,** *11*, 621-30.

(6) Ernst, A. U.; Bowers, D. T.; Wang, L. H.; Shariati, K.; Plesser, M. D.; Brown, N. K.; Mehrabyan, T.; Ma, M. *Adv Drug Deliv Rev* **2019,** *139*, 116-138.

(7) Prince, E.; Kumacheva, E. *Nature Reviews Materials* **2019,** *4*, 99-115.

(8) Taha, I. N.; Naba, A. *Essays Biochem* **2019,** *63*, 417-432.

(9) Stendahl, J. C.; Kaufman, D. B.; Stupp, S. I. *Cell Transplant* **2009,** *18*, 1-12.

(10) Gress, T. M.; Menke, A.; Bachem, M.; Muller-Pillasch, F.; Ellenrieder, V.; Weidenbach, H.; Wagner, M.; Adler, G. *Digestion* **1998,** *59*, 625-37.

(11) Tian, C.; Clauser, K. R.; Ohlund, D.; Rickelt, S.; Huang, Y.; Gupta, M.; Mani, D. R.; Carr, S. A.; Tuveson, D. A.; Hynes, R. O. *Proceedings of the National Academy of Sciences of the United States of America* **2019,** *116*, 19609-19618.

(12) Weniger, M.; Honselmann, K. C.; Liss, A. S. *Cancers (Basel)* **2018,** *10*.

(13) Nikolova, G.; Jabs, N.; Konstantinova, I.; Domogatskaya, A.; Tryggvason, K.; Sorokin, L.; Fassler, R.; Gu, G.; Gerber, H. P.; Ferrara, N.; Melton, D. A.; Lammert, E. *Dev Cell* **2006,** *10*, 397-405.

(14) van Deijnen, J. H.; Hulstaert, C. E.; Wolters, G. H.; van Schilfgaarde, R. *Cell Tissue Res* **1992,** *267*, 139-46.

(15) Otonkoski, T.; Banerjee, M.; Korsgren, O.; Thornell, L. E.; Virtanen, I. *Diabetes Obes Metab* **2008,** *10 Suppl 4*, 119-27.

(16) Bogdani, M.; Korpos, E.; Simeonovic, C. J.; Parish, C. R.; Sorokin, L.; Wight, T. N. *Curr Diab Rep* **2014,** *14*, 552.

(17) Choong, F. J.; Freeman, C.; Parish, C. R.; Simeonovic, C. J. *Am J Transplant* **2015,** *15*, 2851-64.

(18) Lammert, E.; Thorn, P. *J Mol Biol* **2020,** *432*, 1407-1418.

(19) Simeonovic, C. J.; Popp, S. K.; Starrs, L. M.; Brown, D. J.; Ziolkowski, A. F.; Ludwig, B.; Bornstein, S. R.; Wilson, J. D.; Pugliese, A.; Kay, T. W. H.; Thomas, H. E.; Loudovaris, T.; Choong, F. J.; Freeman, C.; Parish, C. R. *PLoS One* **2018,** *13*, e0191360.

(20) Townsend, S. E.; Gannon, M. *Endocrinology* **2019,** *160*, 1885-1894.

(21) Van Deijnen, J. H.; Van Suylichem, P. T.; Wolters, G. H.; Van Schilfgaarde, R. *Cell Tissue Res* **1994,** *277*, 115-21.

(22) Virtanen, I.; Banerjee, M.; Palgi, J.; Korsgren, O.; Lukinius, A.; Thornell, L. E.; Kikkawa, Y.; Sekiguchi, K.; Hukkanen, M.; Konttinen, Y. T.; Otonkoski, T. *Diabetologia* **2008,** *51*, 1181-91.

(23) Kilimnik, G.; Jo, J.; Periwal, V.; Zielinski, M. C.; Hara, M. *Islets* **2012,** *4*, 167-72.

(24) Brissova, M.; Fowler, M. J.; Nicholson, W. E.; Chu, A.; Hirshberg, B.; Harlan, D. M.; Powers, A. C. *J Histochem Cytochem* **2005,** *53*, 1087-97.

(25) Cabrera, O.; Berman, D. M.; Kenyon, N. S.; Ricordi, C.; Berggren, P. O.; Caicedo, A. *Proc Natl Acad Sci U S A* **2006,** *103*, 2334-9.

(26) Dolensek, J.; Rupnik, M. S.; Stozer, A. *Islets* **2015,** *7*, e1024405.

(27) Steiner, D. J.; Kim, A.; Miller, K.; Hara, M. *Islets* **2010,** *2*, 135-45.

(28) Byron, A.; Humphries, J. D.; Humphries, M. J. *Int J Exp Pathol* **2013,** *94*, 75-92.

(29) Naba, A.; Clauser, K. R.; Hoersch, S.; Liu, H.; Carr, S. A.; Hynes, R. O. *Mol Cell Proteomics* **2012,** *11*, M111 014647.

(30) Naba, A.; Pearce, O. M. T.; Del Rosario, A.; Ma, D.; Ding, H.; Rajeeve, V.; Cutillas, P. R.; Balkwill, F. R.; Hynes, R. O. *Journal of proteome research* **2017,** *16*, 3083-3091.

(31) Ma, F.; Tremmel, D. M.; Li, Z.; Lietz, C. B.; Sackett, S. D.; Odorico, J. S.; Li, L. *Journal of proteome research* **2019,** *18*, 3156-3165.

(32) Schiller, H. B.; Mayr, C. H.; Leuschner, G.; Strunz, M.; Staab-Weijnitz, C.; Preisendorfer, S.; Eckes, B.; Moinzadeh, P.; Krieg, T.; Schwartz, D. A.; Hatz, R. A.; Behr, J.; Mann, M.; Eickelberg, O. *Am J Respir Crit Care Med* **2017,** *196*, 1298-1310.

(33) Hill, R. C.; Calle, E. A.; Dzieciatkowska, M.; Niklason, L. E.; Hansen, K. C. *Mol Cell Proteomics* **2015,** *14*, 961-73.

(34) Schiller, H. B.; Fernandez, I. E.; Burgstaller, G.; Schaab, C.; Scheltema, R. A.; Schwarzmayr, T.; Strom, T. M.; Eickelberg, O.; Mann, M. *Mol Syst Biol* **2015,** *11*, 819.

(35) Li, Q.; Uygun, B. E.; Geerts, S.; Ozer, S.; Scalf, M.; Gilpin, S. E.; Ott, H. C.; Yarmush, M. L.; Smith, L. M.; Welham, N. V.; Frey, B. L. *Biomaterials* **2016,** *75*, 37-46.

(36) Hansen, K. C.; Kiemele, L.; Maller, O.; O'Brien, J.; Shankar, A.; Fornetti, J.; Schedin, P. *Mol Cell Proteomics* **2009,** *8*, 1648-57.

(37) Onnerfjord, P.; Khabut, A.; Reinholt, F. P.; Svensson, O.; Heinegard, D. *The Journal of biological chemistry* **2012,** *287*, 18913-24.

(38) Bonvillain, R. W.; Danchuk, S.; Sullivan, D. E.; Betancourt, A. M.; Semon, J. A.; Eagle, M. E.; Mayeux, J. P.; Gregory, A. N.; Wang, G.; Townley, I. K.; Borg, Z. D.; Weiss, D. J.; Bunnell, B. A. *Tissue Eng Part A* **2012,** *18*, 2437-52.

(39) Wilson, R.; Diseberg, A. F.; Gordon, L.; Zivkovic, S.; Tatarczuch, L.; Mackie, E. J.; Gorman, J. J.; Bateman, J. F. *Mol Cell Proteomics* **2010,** *9*, 1296-313.

(40) Zhu, W.; Smith, J. W.; Huang, C. M. *J Biomed Biotechnol* **2010,** *2010*, 840518.

(41) Megger, D. A.; Pott, L. L.; Ahrens, M.; Padden, J.; Bracht, T.; Kuhlmann, K.; Eisenacher, M.; Meyer, H. E.; Sitek, B. *Biochimica et biophysica acta* **2014,** *1844*, 967-76.

(42) Gocheva, V.; Naba, A.; Bhutkar, A.; Guardia, T.; Miller, K. M.; Li, C. M.; Dayton, T. L.; Sanchez-Rivera, F. J.; Kim-Kiselak, C.; Jailkhani, N.; Winslow, M. M.; Del Rosario, A.; Hynes, R. O.; Jacks, T. *Proceedings of the National Academy of Sciences of the United States of America* **2017,** *114*, E5625-E5634.

(43) Tian, Y.; Li, H.; Gao, Y.; Liu, C.; Qiu, T.; Wu, H.; Cao, M.; Zhang, Y.; Ding, H.; Chen, J.; Cai, H. *Clin Proteomics* **2019,** *16*, 6.

(44) Caldeira, J.; Santa, C.; Osorio, H.; Molinos, M.; Manadas, B.; Goncalves, R.; Barbosa, M. *Scientific reports* **2017,** *7*, 11629.

(45) Naba, A.; Clauser, K. R.; Mani, D. R.; Carr, S. A.; Hynes, R. O. *Scientific reports* **2017,** *7*, 40495.

(46) Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L. *Anal. Chem.* **2010,** *82*, 2817-25.

(47) Frost, D. C.; Greer, T.; Xiang, F.; Liang, Z.; Li, L. *Rapid Commun Mass Spectrom* **2015,** *29*, 1115-24.

(48) Frost, D. C.; Greer, T.; Li, L. *Anal. Chem.* **2015,** *87*, 1646-54.

(49) Bi, H.; Ye, K.; Jin, S. *Biomaterials* **2020,** *233*, 119673.

(50) Ma, F.; Liu, F.; Xu, W.; Li, L. *Journal of proteome research* **2018,** *17*, 2744-2754.

(51) Yang, F.; Shen, Y.; Camp, D. G., 2nd; Smith, R. D. *Expert Rev Proteomics* **2012,** *9*, 129-34.

(52) Shao, X.; Taha, I. N.; Clauser, K. R.; Gao, Y. T.; Naba, A. *Nucleic acids research* **2020,** *48*, D1136-D1144.

(53) Naba, A.; Hoersch, S.; Hynes, R. O. *Matrix Biol* **2012,** *31*, 371-2.

(54) Hanzelmann, S.; Castelo, R.; Guinney, J. *BMC Bioinformatics* **2013,** *14*, 7.

(55) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. *Proc Natl Acad Sci U S A* **2005,** *102*, 15545-50.

(56) Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. *Nat. Commun.* **2019,** *10*, 1523.

(57) Walter, W.; Sanchez-Cabo, F.; Ricote, M. *Bioinformatics* **2015,** *31*, 2912-4.

(58) Sandovici, I.; Hammerle, C. M.; Cooper, W. N.; Smith, N. H.; Tarry-Adkins, J. L.; Dunmore, B. J.; Bauer, J.; Andrews, S. R.; Yeo, G. S.; Ozanne, S. E.; Constancia, M. *Diabetologia* **2016,** *59*, 502-11.

(59) Koopmans, F.; Ho, J. T. C.; Smit, A. B.; Li, K. W. *Proteomics* **2018,** *18*.

(60) Frost, D. C.; Rust, C. J.; Robinson, R. A. S.; Li, L. *Anal. Chem.* **2018,** *90*, 10664-10669.

(61) Wang, Z.; Yu, K.; Tan, H.; Wu, Z.; Cho, J. H.; Han, X.; Sun, H.; Beach, T. G.; Peng, J. *Anal. Chem.* **2020**.

(62) Kazerounian, S.; Yee, K. O.; Lawler, J. *Cellular and molecular life sciences : CMLS* **2008,** *65*, 700-12.

(63) Daubon, T.; Leon, C.; Clarke, K.; Andrique, L.; Salabert, L.; Darbo, E.; Pineau, R.; Guerit, S.; Maitre, M.; Dedieu, S.; Jeanne, A.; Bailly, S.; Feige, J. J.; Miletic, H.; Rossi, M.; Bello, L.; Falciani, F.; Bjerkvig, R.; Bikfalvi, A. *Nat. Commun.* **2019,** *10*, 1146.

(64) Zhang, X.; Lawler, J. *Microvascular research* **2007,** *74*, 90-9.

(65) Chen, R.; Brentnall, T. A.; Pan, S.; Cooke, K.; Moyes, K. W.; Lane, Z.; Crispin, D. A.; Goodlett, D. R.; Aebersold, R.; Bronner, M. P. *Mol Cell Proteomics* **2007,** *6*, 1331-42.

(66) Pan, S.; Chen, R.; Reimel, B. A.; Crispin, D. A.; Mirzaei, H.; Cooke, K.; Coleman, J. F.; Lane, Z.; Bronner, M. P.; Goodlett, D. R.; McIntosh, M. W.; Traverso, W.; Aebersold, R.; Brentnall, T. A. *Electrophoresis* **2009,** *30*, 1132-44.

(67) Riopel, M.; Wang, R. *Front Biosci (Landmark Ed)* **2014,** *19*, 77-90.

(68) Pourhanifeh, M. H.; Mohammadi, R.; Noruzi, S.; Hosseini, S. A.; Fanoudi, S.; Mohamadi, Y.; Hashemzehi, M.; Asemi, Z.; Mirzaei, H. R.; Salarinia, R.; Mirzaei, H. *Cancer Cell Int* **2019,** *19*, 157.

(69) Poveda, J.; Sanz, A. B.; Fernandez-Fernandez, B.; Carrasco, S.; Ruiz-Ortega, M.; Cannata-Ortiz, P.; Ortiz, A.; Sanchez-Nino, M. D. *J Cell Mol Med* **2017,** *21*, 154-164.

(70) Cross, S. E.; Vaughan, R. H.; Willcox, A. J.; McBride, A. J.; Abraham, A. A.; Han, B.; Johnson, J. D.; Maillard, E.; Bateman, P. A.; Ramracheya, R. D.; Rorsman, P.; Kadler, K. E.; Dunne, M. J.; Hughes, S. J.; Johnson, P. R. *Am J Transplant* **2017,** *17*, 451-461.

(71) Lavallard, V.; Armanet, M.; Parnaud, G.; Meyer, J.; Barbieux, C.; Montanari, E.; Meier, R.; Morel, P.; Berney, T.; Bosco, D. *FASEB J* **2016,** *30*, 748-60.

(72) Jiang, K.; Chaimov, D.; Patel, S. N.; Liang, J. P.; Wiggins, S. C.; Samojlik, M. M.; Rubiano, A.; Simmons, C. S.; Stabler, C. L. *Biomaterials* **2019,** *198*, 37-48.

(73) Stephens, C. H.; Orr, K. S.; Acton, A. J.; Tersey, S. A.; Mirmira, R. G.; Considine, R. V.; Voytik-Harbin, S. L. *Am J Physiol Endocrinol Metab* **2018,** *315*, E650-E661.

(74) Beattie, G. M.; Rubin, J. S.; Mally, M. I.; Otonkoski, T.; Hayek, A. *Diabetes* **1996,** *45*, 1223-8.

(75) Cirulli, V.; Beattie, G. M.; Klier, G.; Ellisman, M.; Ricordi, C.; Quaranta, V.; Frasier, F.; Ishii, J. K.; Hayek, A.; Salomon, D. R. *J Cell Biol* **2000,** *150*, 1445-60.

(76) Hisaoka, M.; Haratake, J.; Hashimoto, H. *Differentiation* **1993,** *53*, 163-72.

(77) Chaimov, D.; Baruch, L.; Krishtul, S.; Meivar-Levy, I.; Ferber, S.; Machluf, M. *J Control Release* **2017,** *257*, 91-101.

(78) Higuchi, Y.; Shiraki, N.; Yamane, K.; Qin, Z.; Mochitate, K.; Araki, K.; Senokuchi, T.; Yamagata, K.; Hara, M.; Kume, K.; Kume, S. *J Cell Sci* **2010,** *123*, 2733-42.

(79) Llacua, L. A.; Faas, M. M.; de Vos, P. *Diabetologia* **2018,** *61*, 1261-1272.

(80) Narayanan, K.; Lim, V. Y.; Shen, J.; Tan, Z. W.; Rajendran, D.; Luo, S. C.; Gao, S.; Wan, A. C.; Ying, J. Y. *Tissue Eng Part A* **2014,** *20*, 424-33.

(81) Salvatori, M.; Katari, R.; Patel, T.; Peloso, A.; Mugweru, J.; Owusu, K.; Orlando, G. *J Diabetes Sci Technol* **2014,** *8*, 159-169.

(82) Tremmel, D. M.; Odorico, J. S. *Organogenesis* **2018,** *14*, 163-168.

(83) Lee, N. J.; Ali, N.; Zhang, L.; Qi, Y.; Clarke, I.; Enriquez, R. F.; Brzozowska, M.; Lee, I. C.; Rogers, M. J.; Laybutt, D. R.; Center, J. R.; Baldock, P. A.; Herzog, H. *Molecular metabolism* **2018,** *13*, 30-44.

(84) Hagihara, A.; Miyamoto, K.; Furuta, J.; Hiraoka, N.; Wakazono, K.; Seki, S.; Fukushima, S.; Tsao, M. S.; Sugimura, T.; Ushijima, T. *Oncogene* **2004,** *23*, 8705-10.

(85) Sackett, S. D.; Tremmel, D. M.; Ma, F.; Feeney, A. K.; Maguire, R. M.; Brown, M. E.; Zhou, Y.; Li, X.; O'Brien, C.; Li, L.; Burlingham, W. J.; Odorico, J. S. *Scientific reports* **2018,** *8*, 10452.

(86) Llacua, L. A.; Hoek, A.; de Haan, B. J.; de Vos, P. *Islets* **2018,** *10*, 60-68.

(87) Spiers, R. M.; Cross, S. E.; Brown, H. L.; Bateman, P. A.; Vaughan, R. H.; Hughes, S. J.; Johnson, P. R. V. *Cell Transplant* **2018,** *27*, 1039-1046.

(88) Spiers, R. M.; Marzi, J.; Brauchle, E. M.; Cross, S. E.; Vaughan, R. H.; Bateman, P. A.; Hughes, S. J.; Schenke-Layland, K.; Johnson, P. R. V. *Acta Biomater* **2019,** *99*, 269-283.

(89) Blum, B.; Hrvatin, S.; Schuetz, C.; Bonal, C.; Rezania, A.; Melton, D. A. *Nat Biotechnol* **2012,** *30*, 261-4.

(90) Arda, H. E.; Li, L.; Tsai, J.; Torre, E. A.; Rosli, Y.; Peiris, H.; Spitale, R. C.; Dai, C.; Gu, X.; Qu, K.; Wang, P.; Wang, J.; Grompe, M.; Scharfmann, R.; Snyder, M. S.; Bottino, R.; Powers, A. C.; Chang, H. Y.; Kim, S. K. *Cell metabolism* **2016,** *23*, 909-20.

(91) Meier, R. P. H.; Meyer, J.; Muller, Y. D.; Szot, G. L.; Bedat, B.; Andres, A.; Masse, N.; Lablanche, S.; Puppa, G.; Bosco, D.; Berney, T. *Transpl Int* **2020**.

(92) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. *Nature methods* **2016,** *13*, 731-40.

(93) Huang da, W.; Sherman, B. T.; Lempicki, R. A. *Nat Protoc* **2009,** *4*, 44-57.

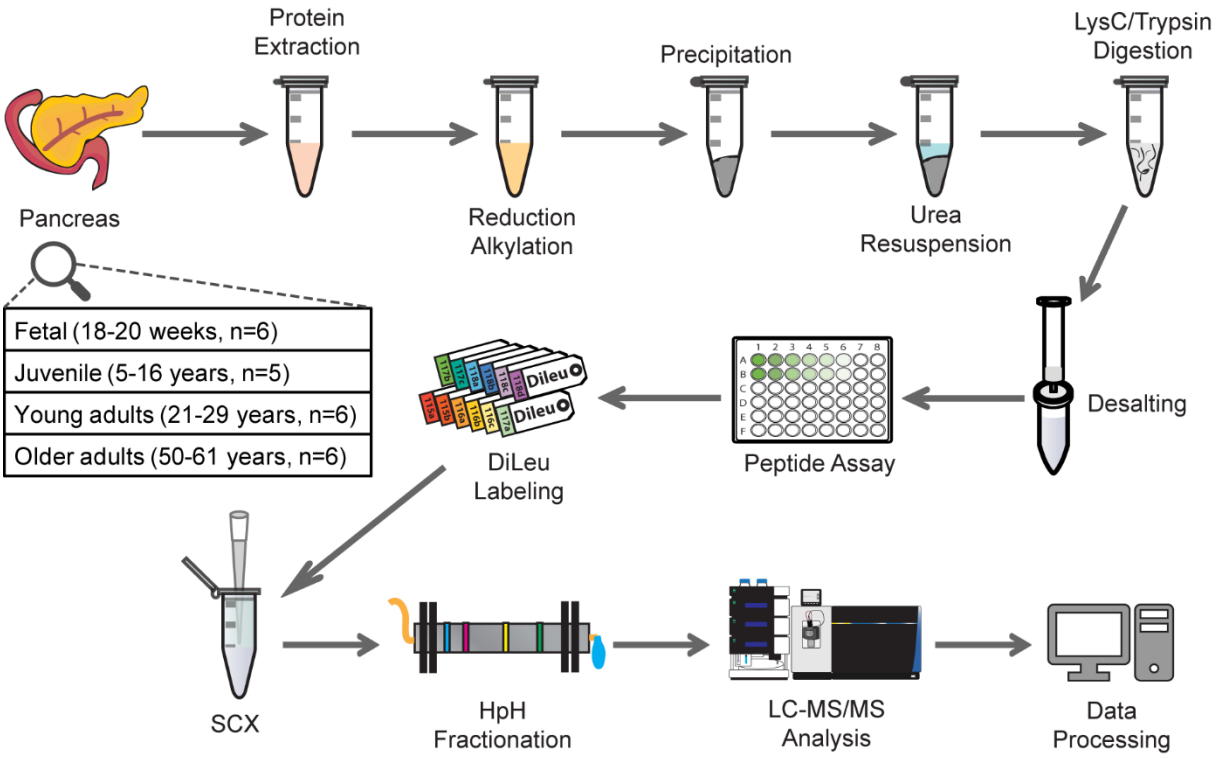

Figure 1. Experimental workflow of quantitative analysis using 12-plex DiLeu isobaric labeling strategy. Pancreas from four age groups were subjected to protein extraction with SDS buffer, precipitation, and digestion according to a slightly modified SCAD method. 12-plex DiLeu isobaric labeling was used to achieve quantitative proteomics analysis. Sample clean-up and fractionation steps were performed prior to LC-MS/MS analysis and data processing was done with commercially available software packages. SCX, strong cation exchange; HpH, high pH.

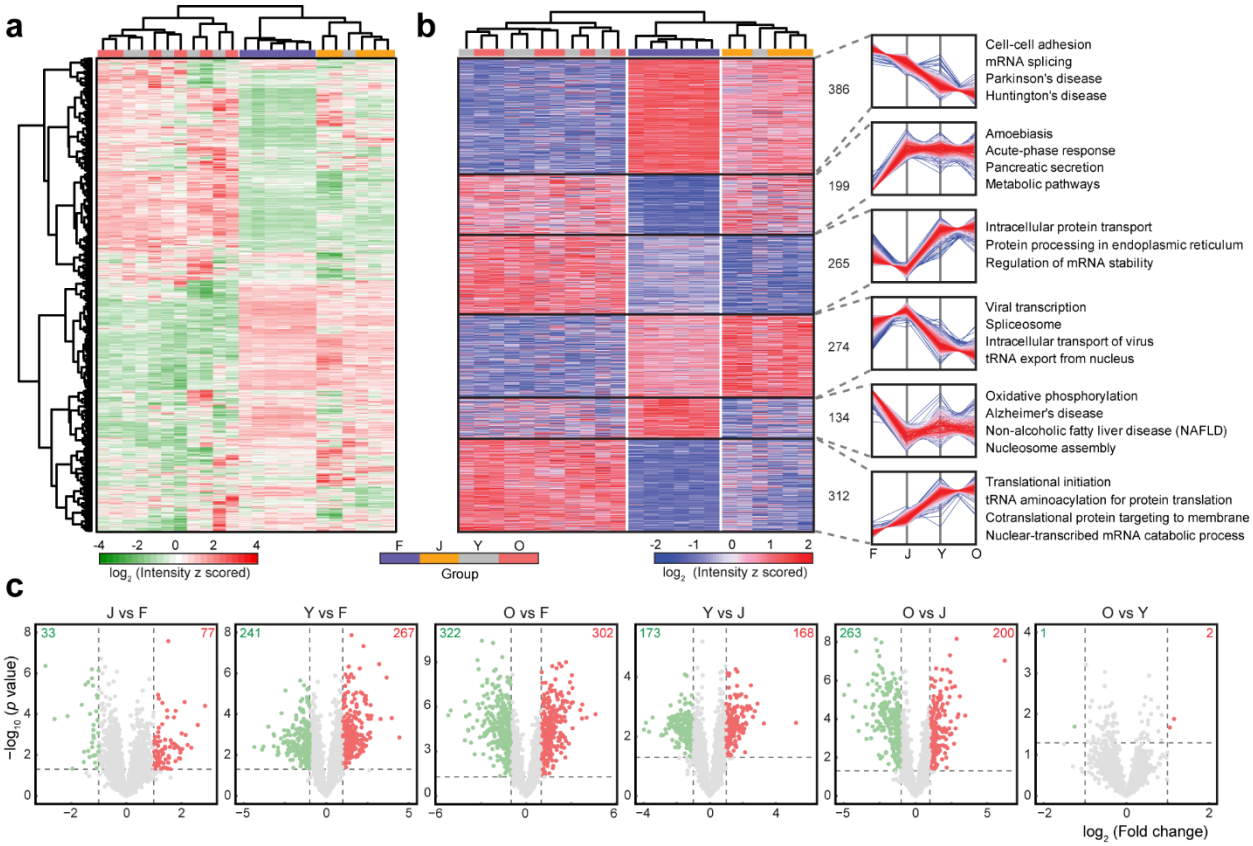

Figure 2. Protein profiles in pancreatic tissue and alterations across multiple developmental stages. a, Hierarchical clustering of DiLeu reporter ion intensities of 2064 quantified proteins. **b,** Hierarchical clustering of DiLeu reporter ion intensities of 1570 significantly changed proteins (one-way ANOVA, FDR 0.05). Six clusters of these proteins are depicted based on different intensity profiles across four developmental stages. Color of each line in clusters is based on its distance from the center, red (close) to blue (far). Number of proteins and selected enriched biological processes and pathways are indicated for each cluster. **c,** Volcano plots showing pairwise comparisons of protein expression levels between various stages. Points above horizontal dash lines represent significantly altered proteins (two-sided *t* test, *p* value < 0.05, *p* values were adjusted by Benjamini-Hochberg correction for multiple comparisons). Significantly down-regulated proteins (e.g., lower abundance in J in the first plot) are shown in green (protein fold change < 0.5) and up-regulated ones are shown in red (protein fold change > 2). F, fetal; J, juvenile; Y, young adult; O, older adult.

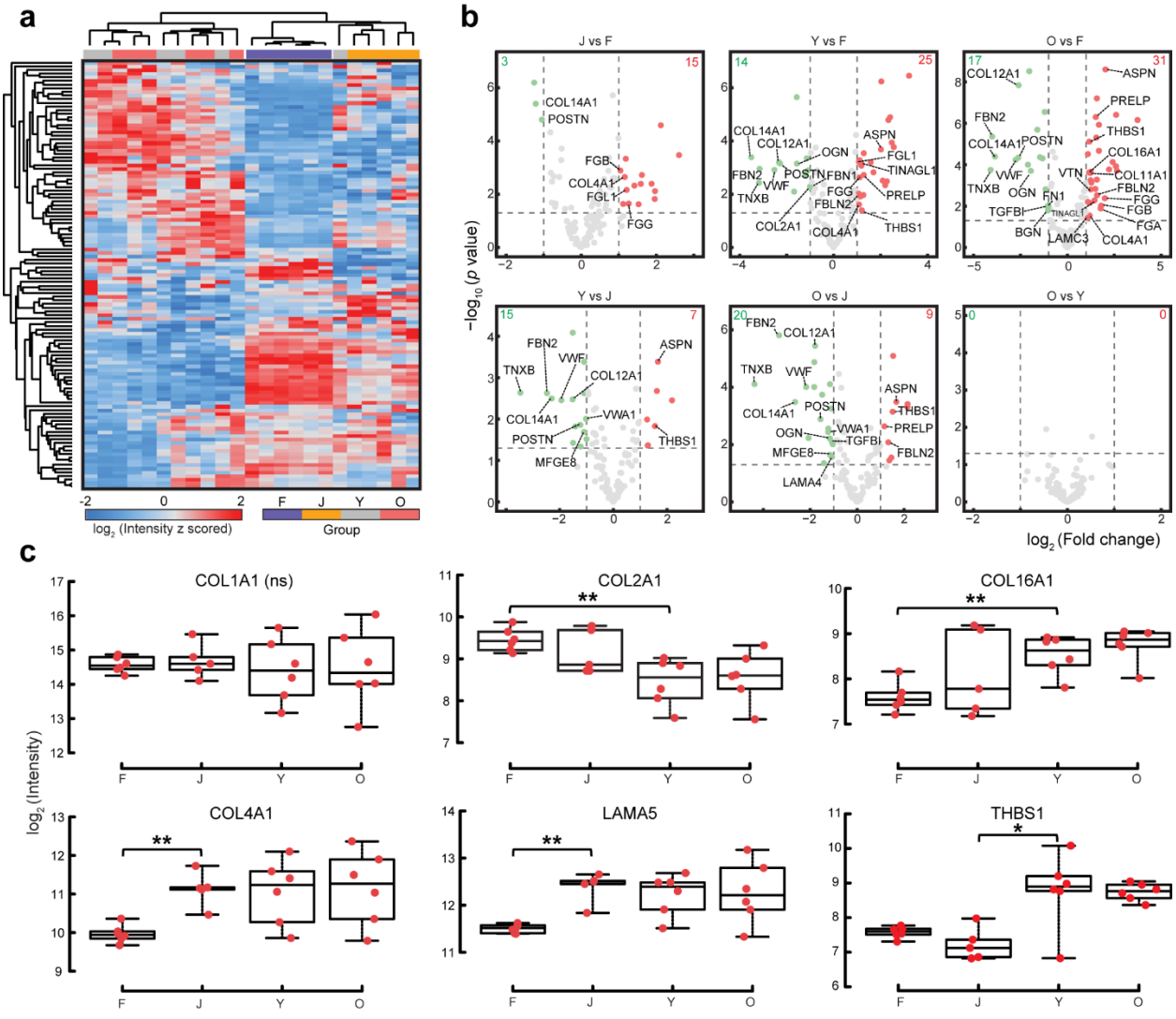

Figure 3. ECM remodeling during fetal and postnatal stages of human pancreas development. **a,** Hierarchical clustering of DiLeu reporter ion intensities of 117 quantified ECM proteins. **b,** Volcano plots showing pairwise comparisons of ECM protein expression levels between various stages. Points above horizontal dash lines represent significantly altered proteins (two-sided *t* test, *p* value < 0.05, *p* values were adjusted by Benjamini-Hochberg correction for multiple comparisons). Significantly down-regulated proteins are shown in green (protein fold change < 0.5) and up-regulated ones are shown in red (protein fold change > 2). Core matrisome proteins are annotated in each figure. **c,** Box plots showing expression levels of selected proteins at different developmental stages. Red dots indicate replicate data points. All boxplots indicate median (center line), 25th and 75th percentiles (bounds of box), and minimum and maximum (whiskers). Significance level is marked with an asterisk (two-sided *t* test, *p* values were adjusted by

Benjamini-Hochberg correction for multiple comparisons, *p value < 0.05, **p value < 0.01; ns, not significant).

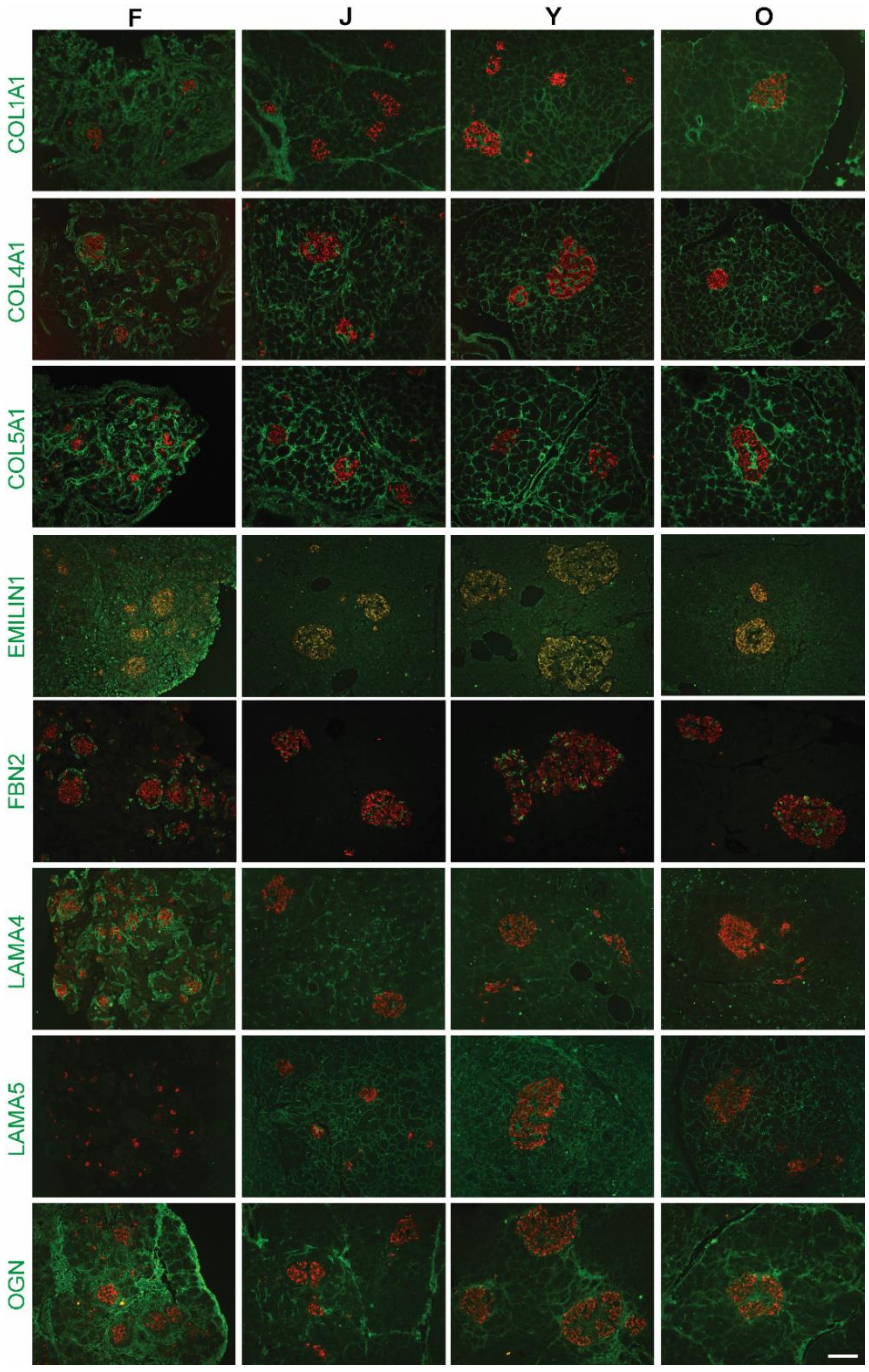

Figure 4. Visualizing ECM proteins across multiple developmental stages. Immunofluorescent images of selected ECM proteins (green) co-stained with insulin (red) in fetal (F), juvenile (J), young adult (Y) and older adult (O) pancreata. Qualitative trends in protein levels corroborate MS data. Representative images are shown, images were taken for N=3 donors per developmental group. An expanded panel of images is available in **Figure S8**. Scale bar = 100 microns.

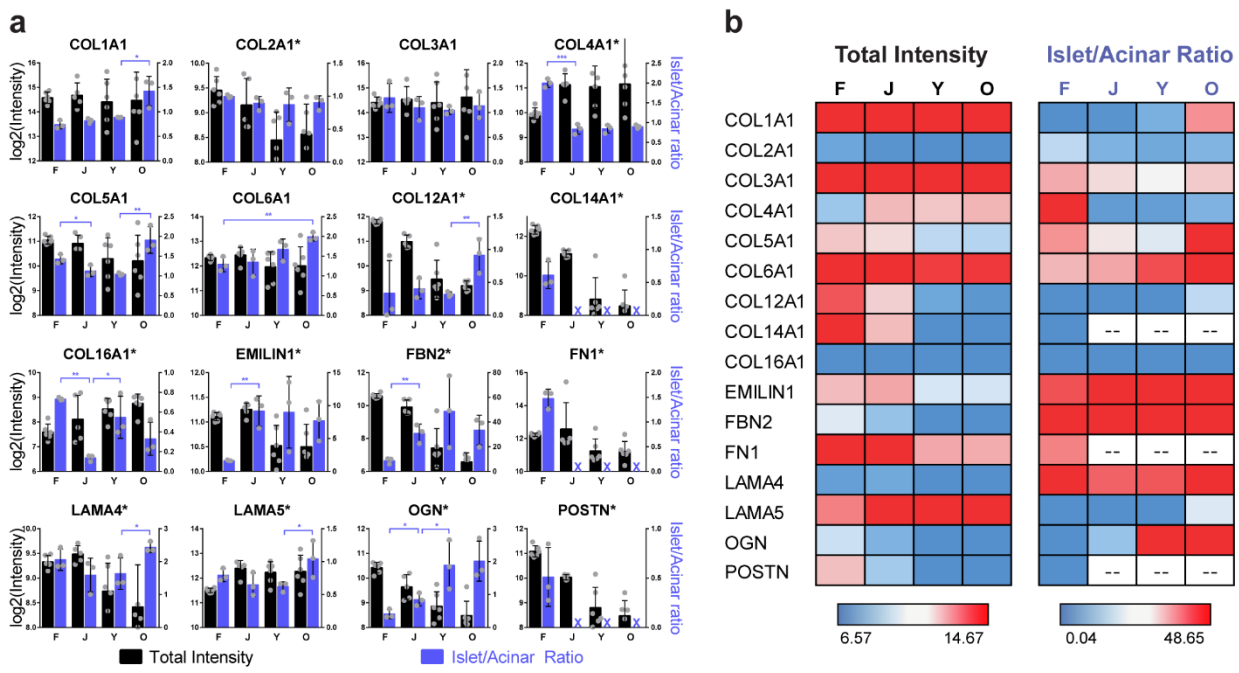

Figure 5. ECM proteins change in abundance and localization across developmental time points. a, Representation of total intensity of selected ECM proteins as measured by MS in the whole pancreas (black bars) (F,Y,O: N=6 donors per group) (J: N=5 donors). Quantification of immunofluorescence (IF) staining in pancreatic acinar and islet regions, depicted as the ratio of islet/acinar (purple bars), in fetal (F), juvenile (J), young adult (Y) and older adult (O) pancreata, graphed as mean ± the standard deviation (purple bars) (N=3 donors per group). Significant difference for islet/acinar ratio was determined by a two-tailed *t* test (**p* < 0.05, ***p* < 0.01, ****p* < 0.001). An asterisk next to the protein name indicates a significant difference in total protein abundance by ANOVA among the four groups (FDR 0.05). A purple 'X' indicates IF staining was too low in the islets and acinar to calculate the islet/acinar ratio. Exact *p* values are included in the Source Data file. **b,** Heat maps depicting the same data as **a**, with emphasis on relative changes throughout the four groups. Blue color indicates the lower 25th percentile of values, red color indicates the top 75th percentile of values, with a gradient toward white at the 50th percentile.

Supplemental Information

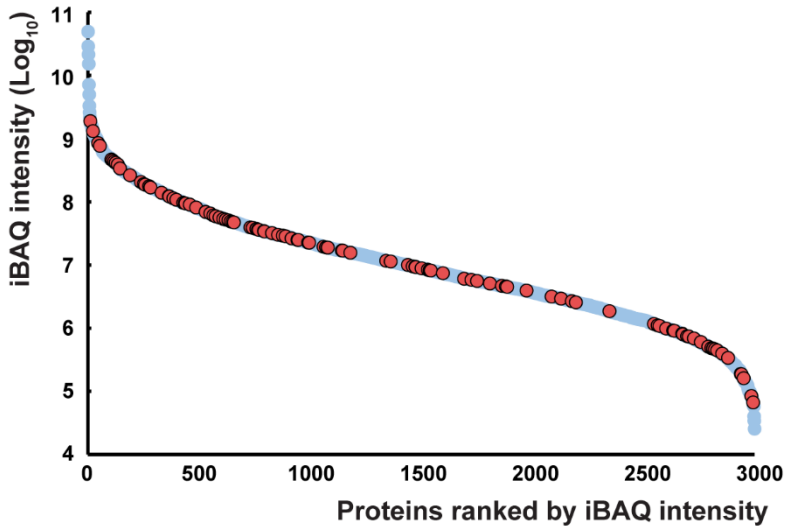

Figure S1. Dynamic intensity range of identified proteins in adult groups. Proteins identified in young and older adult groups are ranked and plotted from high to low based on iBAQ (Intensity Based Absolute Quantification) intensities. All quantified ECM proteins are highlighted in closed red circles.

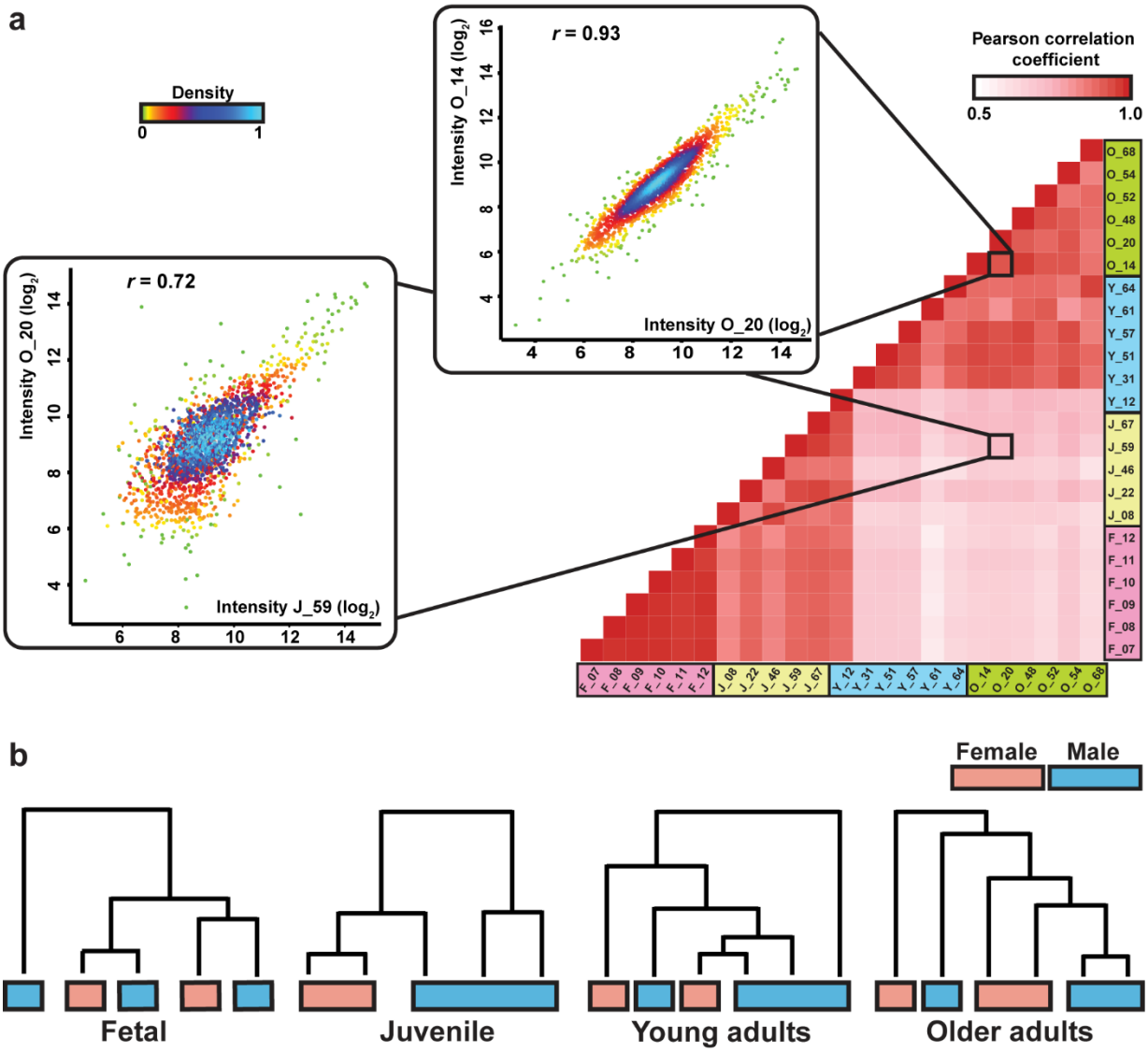

Figure S2. Correlation between samples and gender differences across developmental groups.

a, Pearson correlation analysis of protein intensity in all samples from four age groups. Color coding of each box in the heatmap indicates the Pearson correlation coefficient between the column- and row-indexed samples. Density plots illustrate the intensity correlation of two representative pairs with a good correlated one from two samples in the same age group and a poor correlated one from two samples at different stages. **b,** Dendrograms generated from hierarchical clustering in different groups.

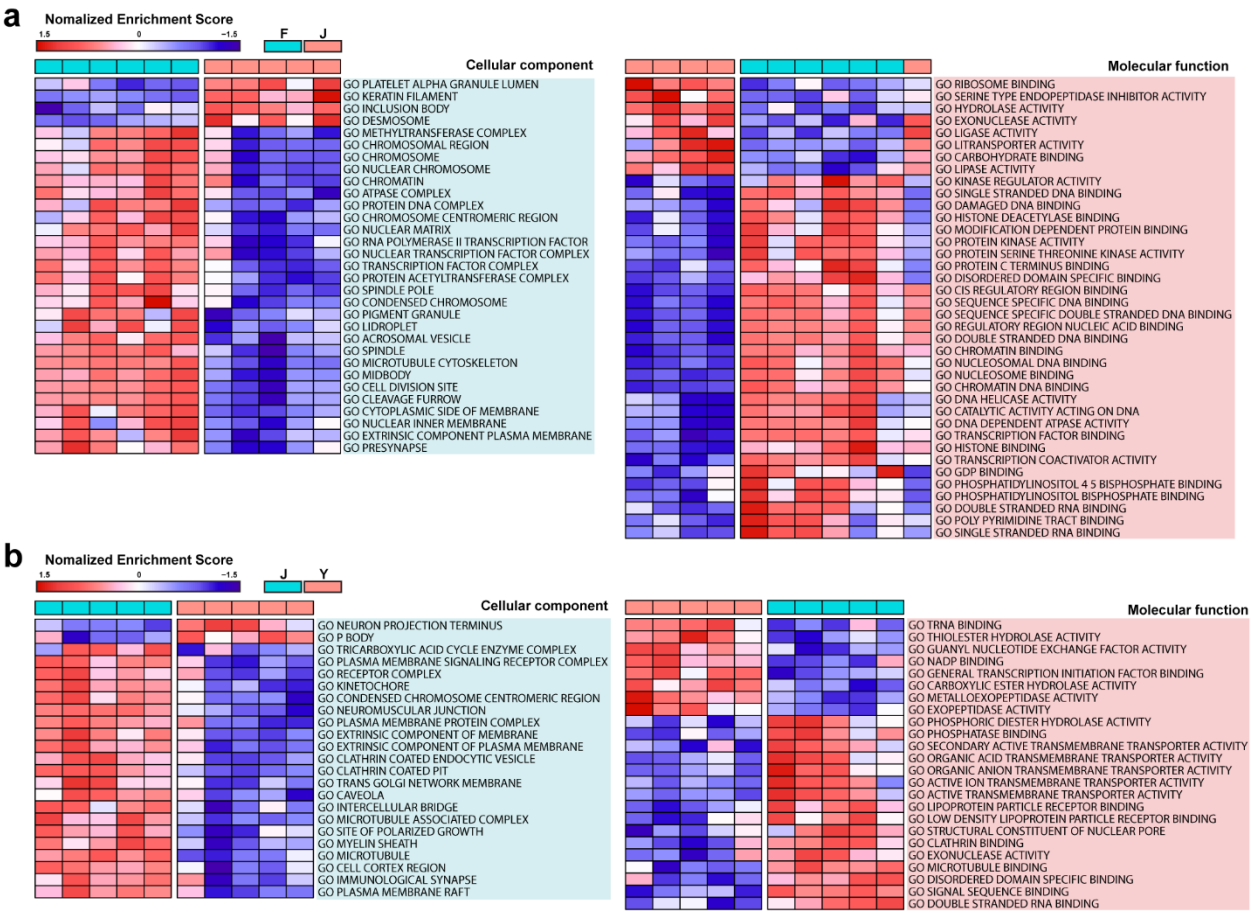

Figure S3. GSVA analysis in juvenile compared to fetal and young adult compared to juvenile. GSVA analysis showing the significantly changed (two-sided *t* test, *p* value < 0.05, *p* values were adjusted by Benjamini-Hochberg correction for multiple comparisons) cellular components and molecular functions in juvenile versus fetal (**a**) and young adult versus juvenile (**b**). Color coding of the heatmaps indicates normalized enrichment score in each sample. A full list of enriched terms including biological processes and transcription factor targets is available in **Table S6**.

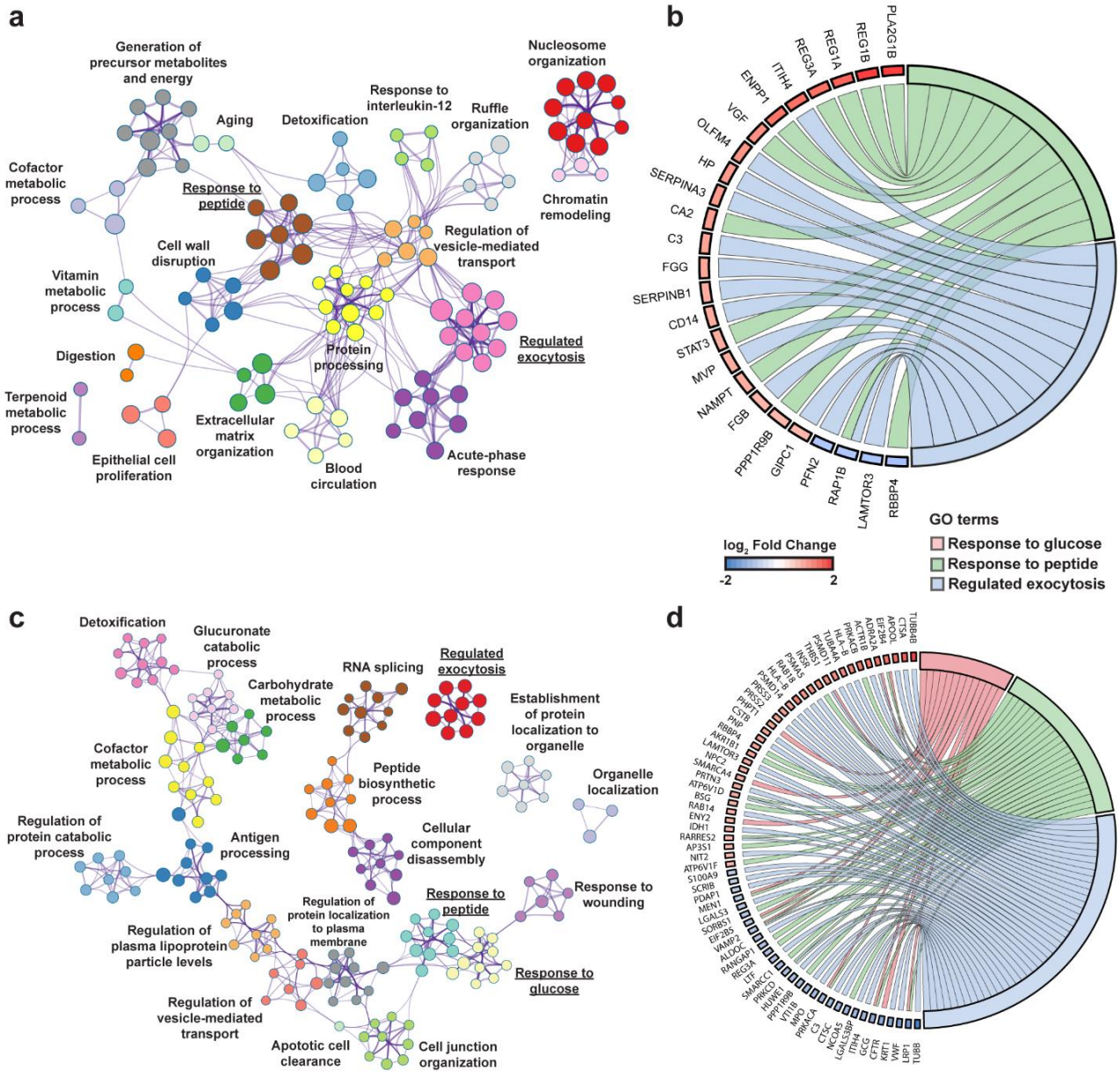

Figure S4. Profiling differences in juvenile compared to fetal and young adult compared to juvenile. a, Network of biological processes enriched from significantly changed proteins (two-sided *t* test, p value < 0.05, p values were adjusted by Benjamini-Hochberg correction for multiple comparisons) in juvenile versus fetal. Each node refers to an enriched term and different terms are grouped into clusters based on their similarities while the most statistically significant term represents the cluster name. **b,** Proteins to biological process linkages in juvenile versus fetal.

Genes are linked via ribbons to their assigned terms and blue-to-red coding next to the selected genes indicates their fold changes. **c,** Network of biological processes enriched from significantly changed proteins (two-sided *t* test, *p* value < 0.05, *p* values were adjusted by Benjamini-Hochberg correction for multiple comparisons) in young adult versus juvenile. **d,** Proteins to biological process linkages in young adult versus juvenile.

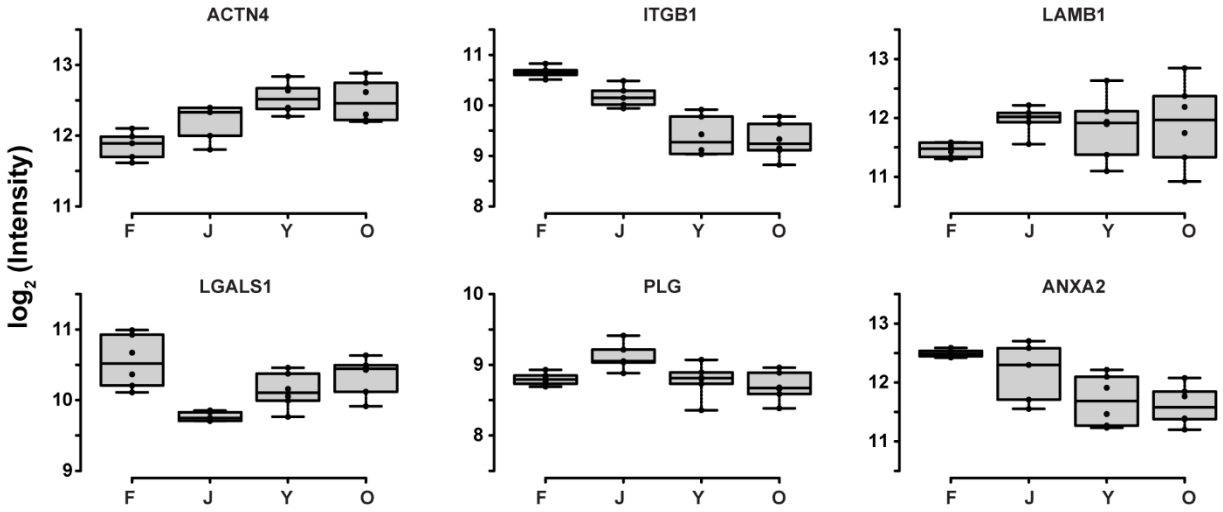

Figure S5. Protein expression level changes of previously reported pancreatic cancer biomarkers. Box plots showing expression levels of selected proteins at different developmental stages. Dots within boxes indicate replicate data points. All boxplots indicate median (center line), 25th and 75th percentiles (bounds of box), and minimum and maximum (whiskers).

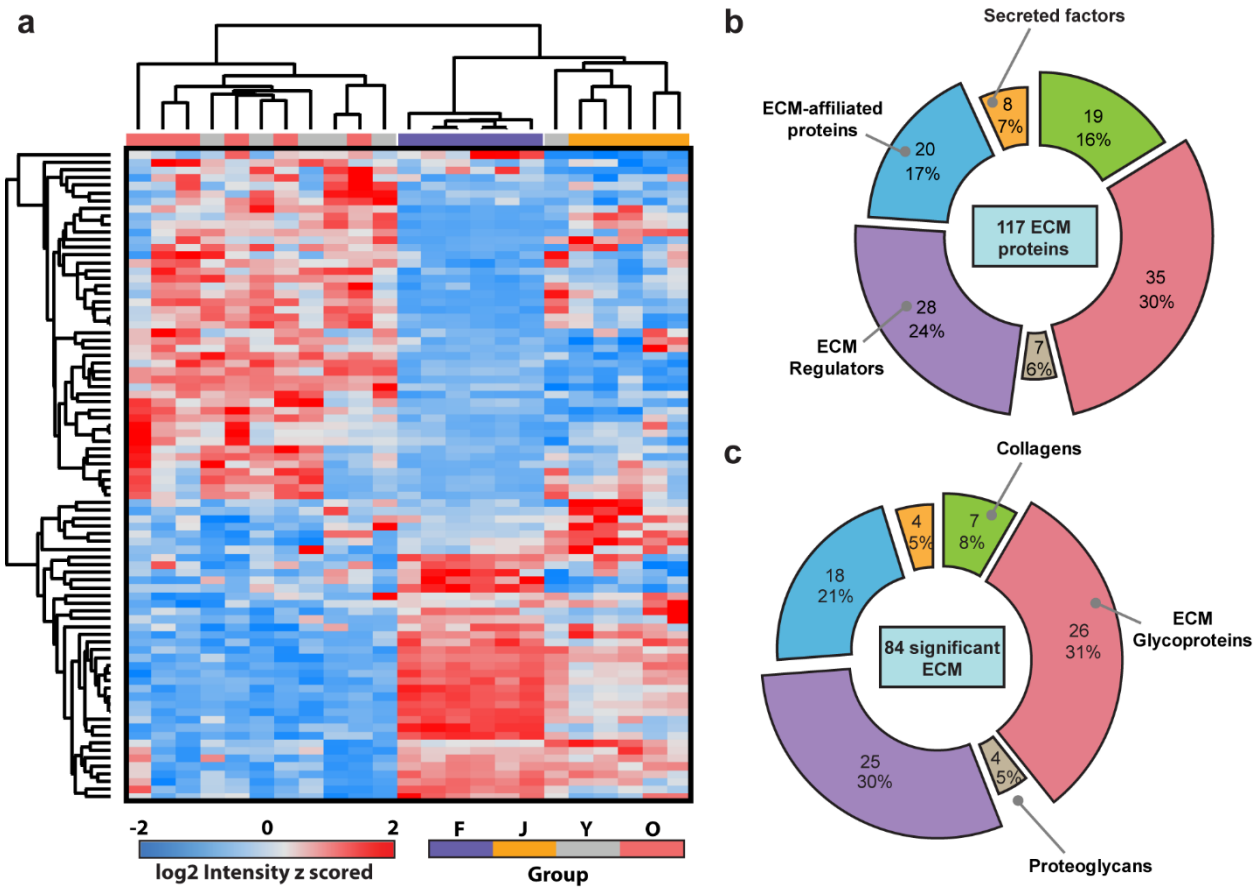

Figure S6. ECM profile alterations across various developmental stages. a, Hierarchical clustering of DiLeu reporter ion intensities of 84 significantly changed ECM proteins (one-way ANOVA, FDR 0.05). **b,** Pie chart showing the number and proportion of each category of all quantified ECM proteins. **c,** Pie chart showing the number and proportion of each category of significantly changed ECM proteins in **a**.

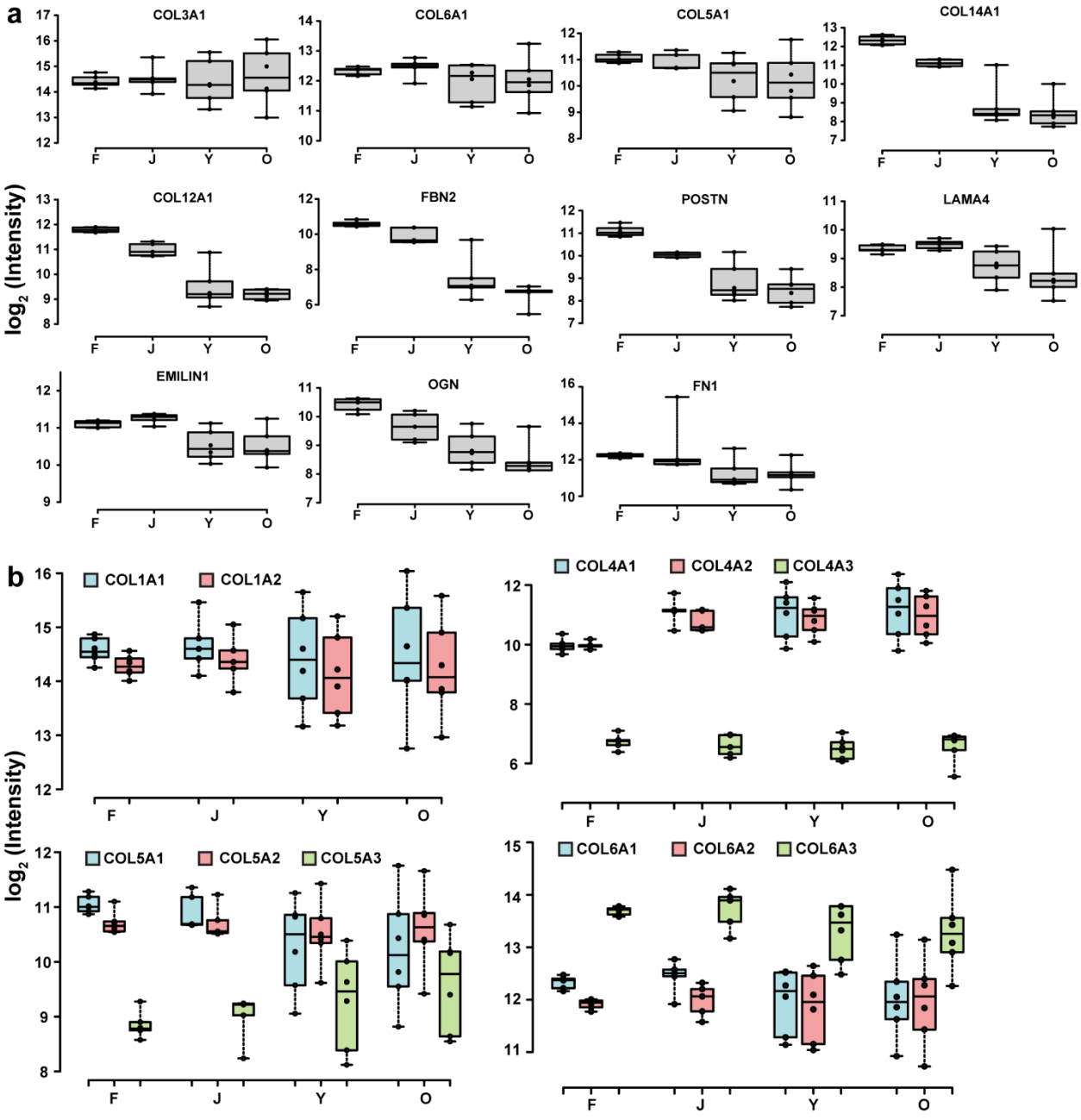

Figure S7. ECM remodeling of human pancreata across developmental stages. Box plots showing expression levels of selected ECM proteins (**a**) and different chains of the same collagen molecule (**b**) at different developmental stages. Dots within boxes indicate replicate data points. All boxplots indicate median (center line), 25th and 75th percentiles (bounds of box), and minimum and maximum (whiskers).

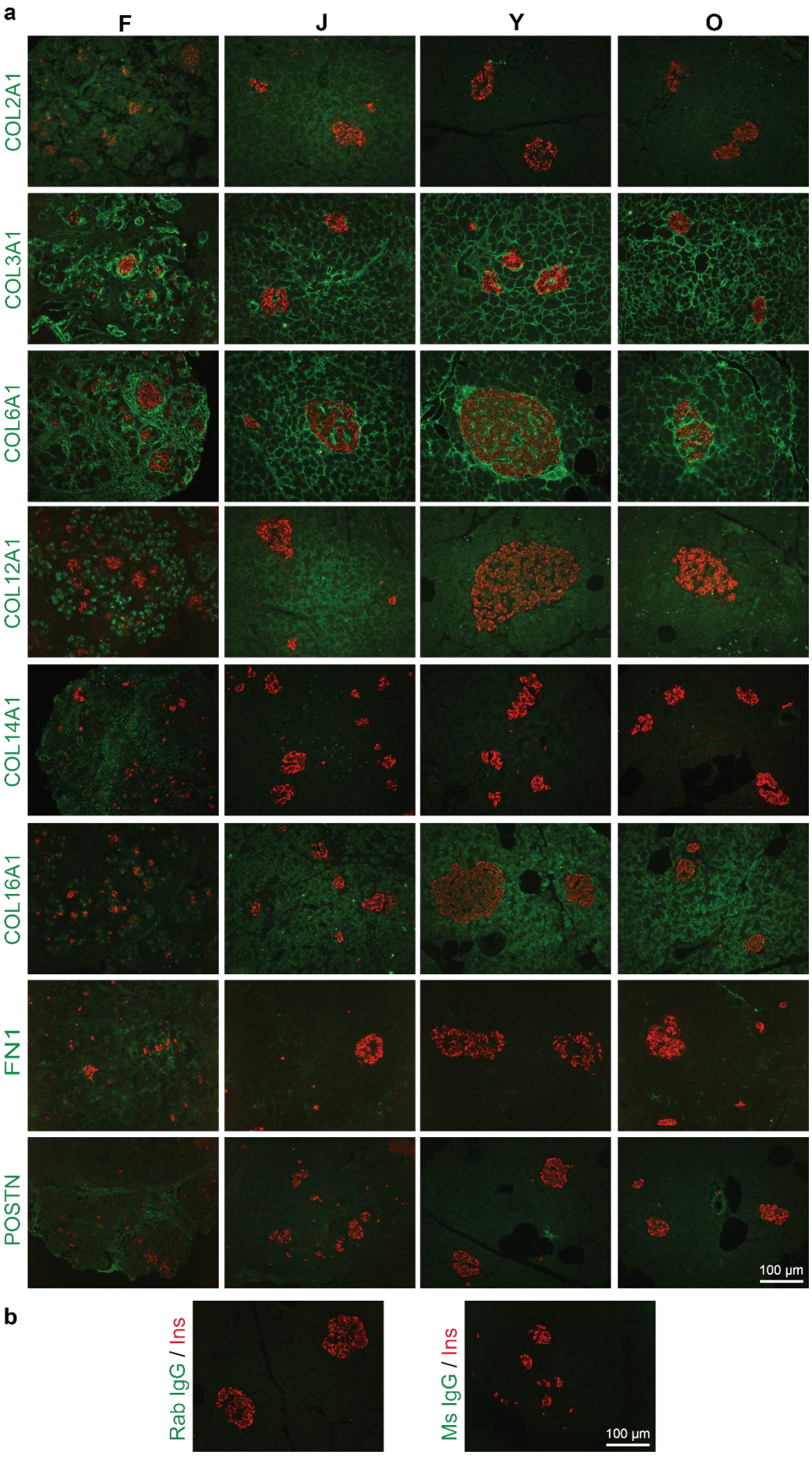
a
F
J
Y
O
COL2A1
COL3A1
COL6A1
COL12A1
COL14A1
COL16A1
FN1
POSTN
100 µm
b
Rab IgG / Ins
Ms IgG / Ins
100 µm

Figure S8. Visualizing ECM proteins in human pancreata across developmental stages. a, Immunofluorescent images of ECM proteins (green) co-stained with insulin (red) in fetal (F), juvenile (J), young adult (Y) and older adult (O) pancreata. Qualitative trends in protein levels corroborate MS data. Representative images are shown, images were taken for N=3 donors per developmental group. **b,** IgG isotype control images with rabbit or mouse IgG (green) indicate low levels of non-specific signal, N=3 donors. Scale bar = 100 microns.

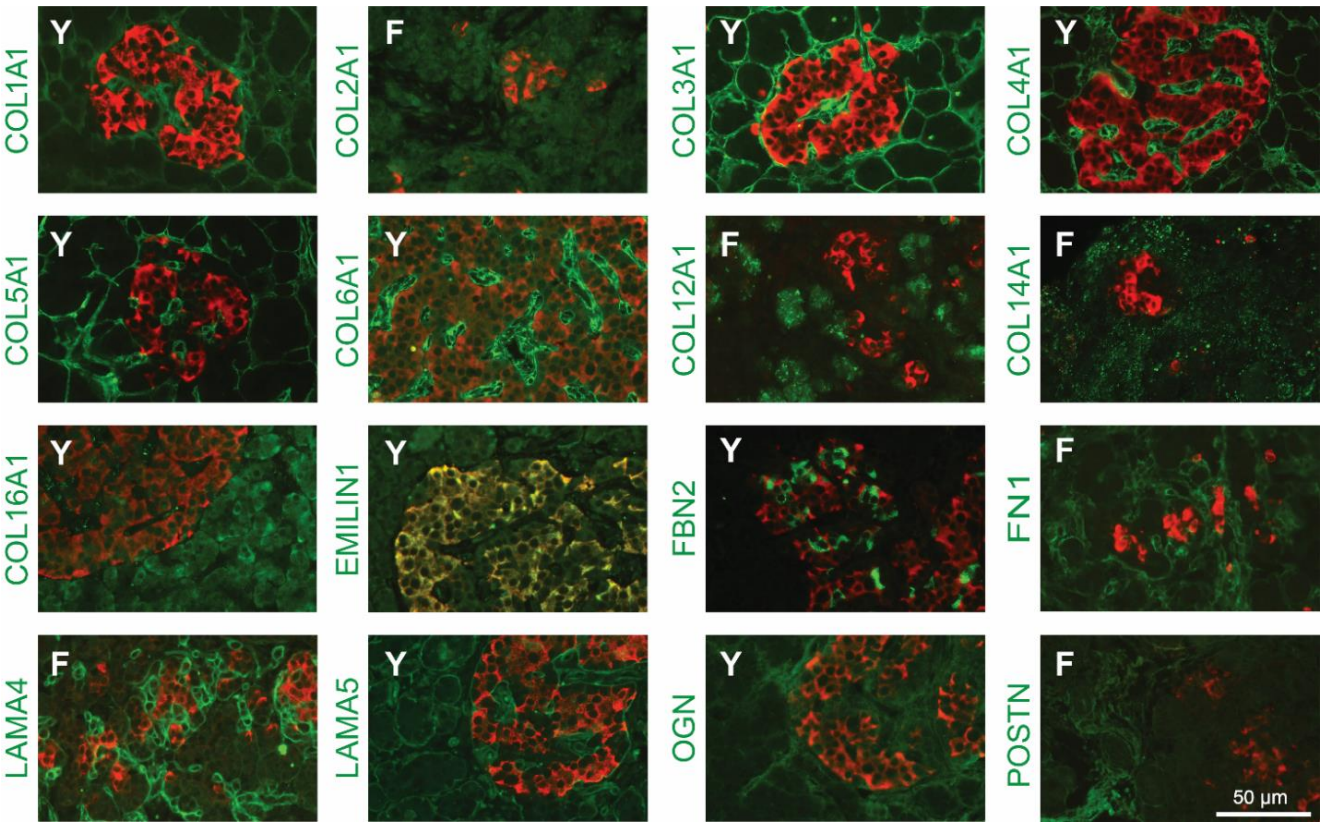

Figure S9. Cellular localization of ECM proteins in pancreatic islets. Enlarged images of immunofluorescent staining for ECM proteins (green) within human pancreatic islets (insulin = red). Images represent either fetal (F) or young adult (Y) donors as indicated, selected based on which age had higher intensity staining for each protein. Images clearly show differences in subcellular localization; most ECM proteins are expressed extracellularly while some proteins (such as EMILIN1 and FBN2) appear to be expressed intracellularly. Scale bar = 50 microns. Representative images shown, from N=3 donors per developmental group.

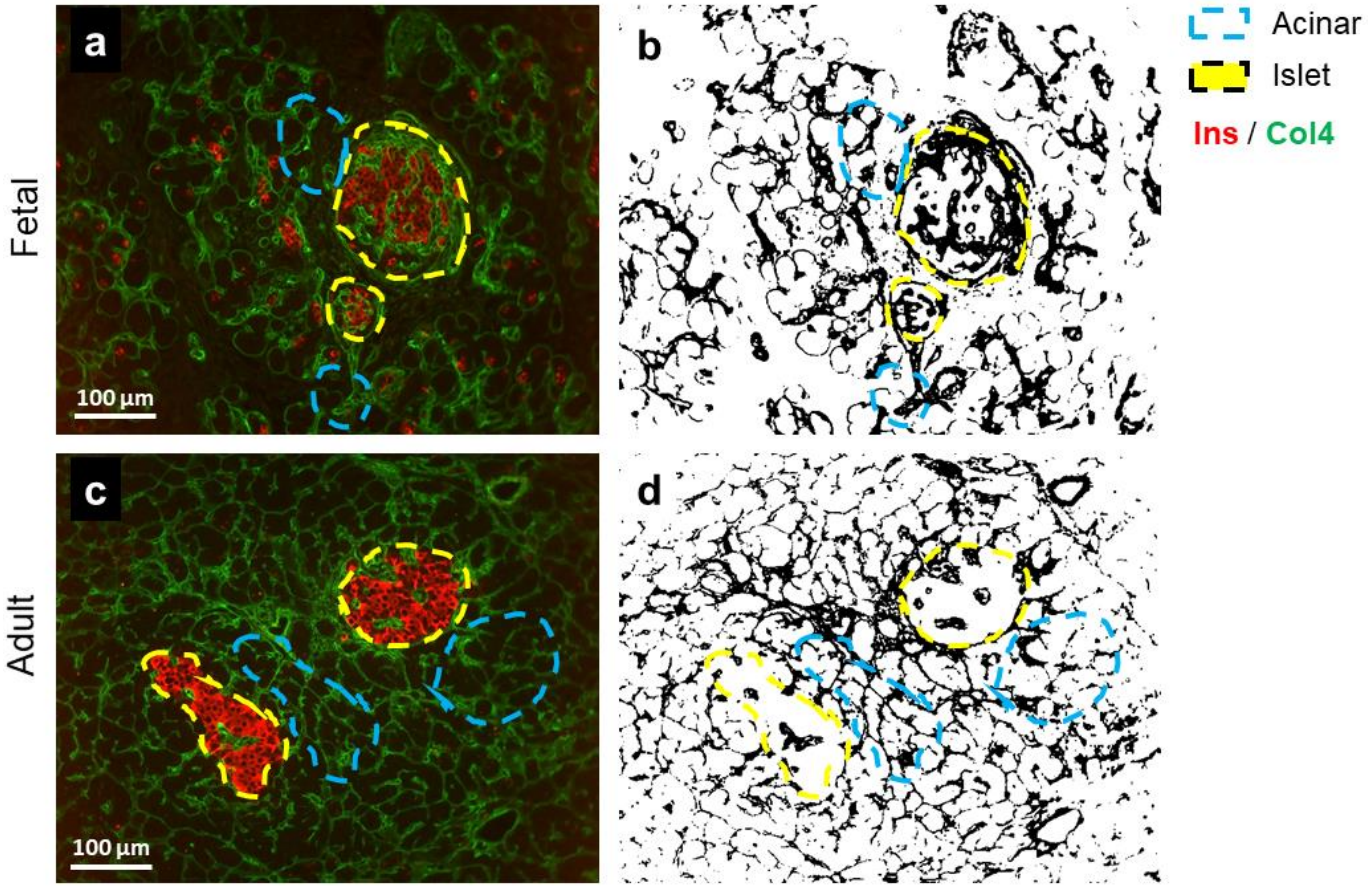

Figure S10. Method for quantification of Islet/Acinar Ratios. Immunofluorescent images of ECM proteins (green) co-stained with insulin (red) were analyzed using ImageJ software by tracing islets (red stain, yellow dashed line) and acinar (not red, morphologically determined, blue dashed line) in the original images **(a**, **c)**, and measuring the intensity of the ECM signal on adjusted binary images **(b**, **d)**. Representative images of fetal and adult tissue are shown for comparison. Scale = 100 microns. Representative images shown, from N=3 donors per developmental group.

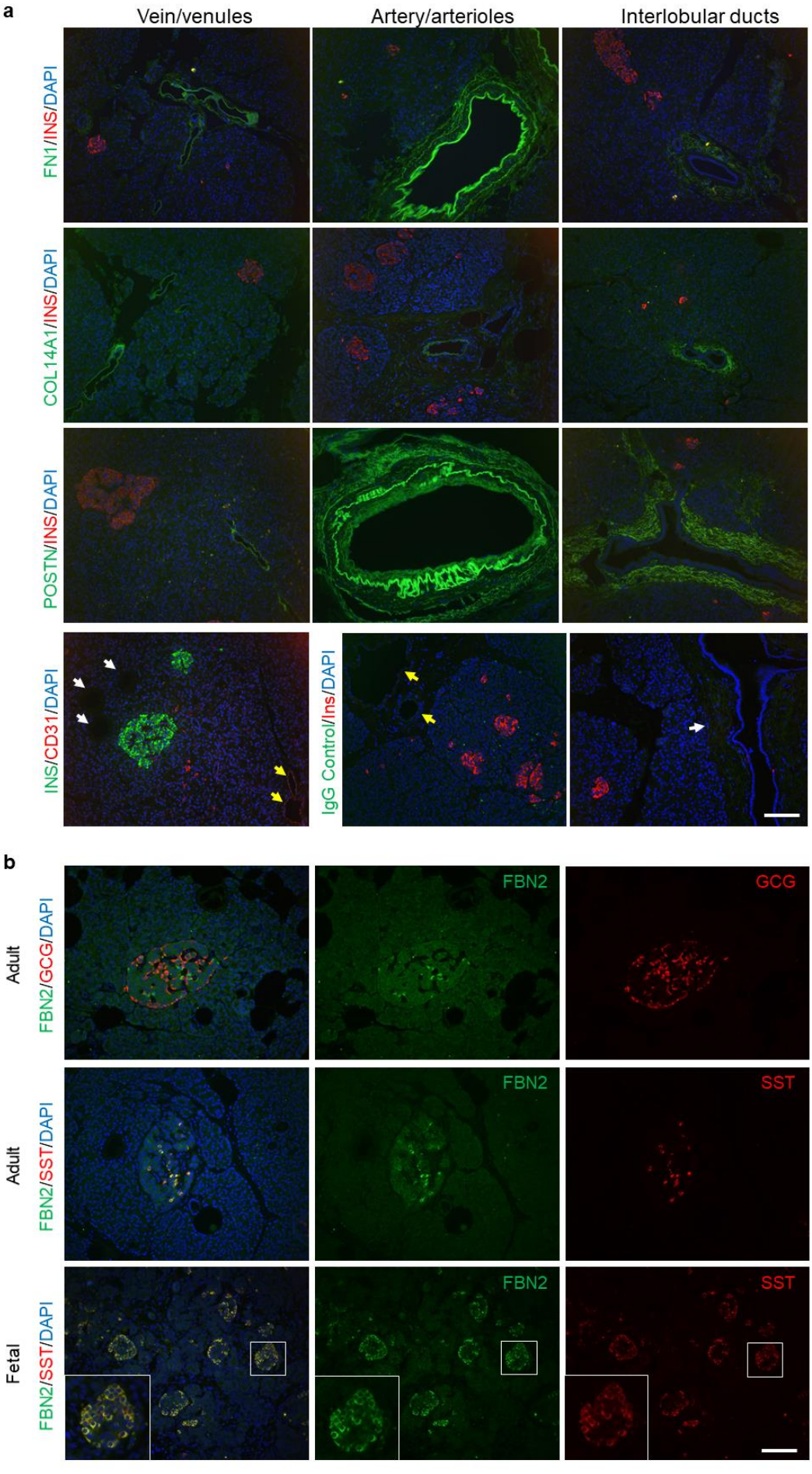
a
Vein/venules
Artery/arterioles
Interlobular ducts
FN1/INS/DAPI
COL14A1/INS/DAPI
POSTN/INS/DAPI
INS/CD31/DAPI
IgG Control/Ins/DAPI
b
Adult
FBN2/GCG/DAPI
FBN2
GCG
Adult
FBN2/SST/DAPI
FBN2
SST
Fetal
FBN2/SST/DAPI
FBN2
SST

Figure S11. Localization of ECM proteins in specific regions of the pancreas. a, Immunofluorescent staining for ECM proteins (FN1, COL14A1, POSTN = green) which had very low levels of signal in islets (Insulin = red) and acinar regions, but relatively high total protein content. These proteins were mainly found to be expressed in vessels and ducts. Control images indicate low levels of autofluorescence in the ducts (white arrows) and vessels (identified in red with positive CD31 staining; yellow arrows) when stained with an insulin antibody in green, or when stained with an IgG isotype control. **b,** Immunofluorescent staining of endocrine markers (GCG, alpha cells = red) (SST, delta cells = red) indicating the at FBN2 co-localizes with delta cells, in both fetal and adult islets. Scale bars = 100 microns. Representative images shown, from N=3 donors per developmental group.

Table S1. Donor information.

Postnatal Donors

	ID	Sex	DCD/DBD	Donor Age (years)	BMI (kg/m^2)	CIT (hours)
Juvenile (J)	*8*	Male	DBD	Range: 5-16 Average: 11	Range: 14.7 - 48.8 Average: 27.4	Average = 11.2
	22	Male	DCD			
	46	Male	DCD			
	59	Female	DCD			
	67	Female	DBD			
Young Adult (Y)	*12*	Male	DCD	Range: 21-29 Average: 25	Range: 20.1 - 37.9 Average: 27.8	Average = 18.9
	31	Male	DBD			
	51	Male	DBD			
	57	Female	DBD			
	61	Female	DCD			
	64	Male	DCD			
Older Adult (O)	*14*	Male	DBD	Range: 50-61 Average: 56.7	Range: 23.8 - 28.6 Average: 26.5	Average = 7.8
	20	Female	DBD			
	48	Male	DBD			
	52	Female	DBD			
	54	Male	DBD			
	68	Female	DBD			

Fetal Donors

ID	Sex	Donor Age (gest. weeks)
f7	Male	20 weeks
f8	Male	19 weeks
f9	Female	19 weeks
f10	Female	19 weeks
f11	Male	18 weeks
f12	Unk	19 weeks

Table S2. List of all identified proteins.

See **Supplementary Data 2** at **https://www.nature.com/articles/s41467-021-21261-w.#Sec18**

Table S3. List of all quantified proteins. Log_2 DiLeu reporter ion intensities of each protein in all samples are shown. Significantly changed proteins among age groups (one-way ANOVA, FDR 0.05) are highlighted in orange. Protein accession, *p* values and log_2(fold change) for all pairwise comparisons are shown. Significantly down-regulated proteins (two-sided *t* test, $p < 0.05$, FC < 0.5) are highlighted in green and up-regulated ones (two-sided *t* test, $p < 0.05$, FC > 2) in red.

See **Supplementary Data 3** at **https://www.nature.com/articles/s41467-021-21261-w.#Sec18**

Table S4. List of all identified ECM proteins. Proteins common to all groups (quantified) are highlighted in yellow.

ECM identified in F/J groups				Common to all	ECM identified in Y/O groups			
Accession	Description	Gene name	ECM category		Accession	Description	Gene name	ECM category
O14672	Disintegrin an	ADAM10	ECM Regulators		O14672	Disintegrin an	ADAM10	ECM Regulators
O00468	Agrin	AGRN	ECM Glycoproteins		O00468	Agrin	AGRN	ECM Glycoproteins
P04083	Annexin A1	ANXA1	ECM-affiliated Proteins		P04083	Annexin A1	ANXA1	ECM-affiliated Proteins
P50995	Annexin A11	ANXA11	ECM-affiliated Proteins		P50995	Annexin A11	ANXA11	ECM-affiliated Proteins
P07355	Annexin A2	ANXA2	ECM-affiliated Proteins		P07355	Annexin A2	ANXA2	ECM-affiliated Proteins
P12429	Annexin A3	ANXA3	ECM-affiliated Proteins		P12429	Annexin A3	ANXA3	ECM-affiliated Proteins
P09525	Annexin A4	ANXA4	ECM-affiliated Proteins		P09525	Annexin A4	ANXA4	ECM-affiliated Proteins
P08758	Annexin A5	ANXA5	ECM-affiliated Proteins		P08758	Annexin A5	ANXA5	ECM-affiliated Proteins
P08133	Annexin A6	ANXA6	ECM-affiliated Proteins		P08133	Annexin A6	ANXA6	ECM-affiliated Proteins
P20073	Annexin A7	ANXA7	ECM-affiliated Proteins		P20073	Annexin A7	ANXA7	ECM-affiliated Proteins
Q9BXN1	Asporin	ASPN	Proteoglycans		Q9BXN1	Asporin	ASPN	Proteoglycans
P21810	Biglycan	BGN	Proteoglycans		P21810	Biglycan	BGN	Proteoglycans
P02746	Complement (	C1QB	ECM-affiliated Proteins		P02746	Complement (	C1QB	ECM-affiliated Proteins
P08217	Chymotrypsin	CELA2A	ECM Regulators		P08217	Chymotrypsin	CELA2A	ECM Regulators
P08218	Chymotrypsin	CELA2B	ECM Regulators		P08218	Chymotrypsin	CELA2B	ECM Regulators
P09093	Chymotrypsin	CELA3A	ECM Regulators		P09093	Chymotrypsin	CELA3A	ECM Regulators
P08861	Chymotrypsin	CELA3B	ECM Regulators		P08861	Chymotrypsin	CELA3B	ECM Regulators
P12107	Collagen alpha	COL11A1	Collagens		P12107	Collagen alpha	COL11A1	Collagens
Q99715	Collagen alpha	COL12A1	Collagens		Q99715	Collagen alpha	COL12A1	Collagens
Q05707	Collagen alpha	COL14A1	Collagens		Q05707	Collagen alpha	COL14A1	Collagens
P39059	Collagen alpha	COL15A1	Collagens		P39059	Collagen alpha	COL15A1	Collagens
Q07092	Collagen alpha	COL16A1	Collagens		Q07092	Collagen alpha	COL16A1	Collagens
P39060	Collagen alpha	COL18A1	Collagens		P39060	Collagen alpha	COL18A1	Collagens
P02452	Collagen alpha	COL1A1	Collagens		P02452	Collagen alpha	COL1A1	Collagens
P08123	Collagen alpha	COL1A2	Collagens		P08123	Collagen alpha	COL1A2	Collagens
P02458	Collagen alpha	COL2A1	Collagens		P02458	Collagen alpha	COL2A1	Collagens
P02461	Collagen alpha	COL3A1	Collagens		P02461	Collagen alpha	COL3A1	Collagens
P02462	Collagen alpha	COL4A1	Collagens		P02462	Collagen alpha	COL4A1	Collagens
P08572	Collagen alpha	COL4A2	Collagens		P08572	Collagen alpha	COL4A2	Collagens
Q01955	Collagen alpha	COL4A3	Collagens		Q01955	Collagen alpha	COL4A3	Collagens
P20908	Collagen alpha	COL5A1	Collagens		P20908	Collagen alpha	COL5A1	Collagens
P05997	Collagen alpha	COL5A2	Collagens		P05997	Collagen alpha	COL5A2	Collagens
P25940	Collagen alpha	COL5A3	Collagens		P25940	Collagen alpha	COL5A3	Collagens
P12109	Collagen alpha	COL6A1	Collagens		P12109	Collagen alpha	COL6A1	Collagens
P12110	Collagen alpha	COL6A2	Collagens		P12110	Collagen alpha	COL6A2	Collagens
P12111	Collagen alpha	COL6A3	Collagens		P12111	Collagen alpha	COL6A3	Collagens
Q6UXH1	Cysteine-rich	CRELD2	ECM Glycoproteins		Q6UXH1	Cysteine-rich	CRELD2	ECM Glycoproteins
P01034	Cystatin-C	CST3	ECM Regulators		P01034	Cystatin-C	CST3	ECM Regulators
P04080	Cystatin-B	CSTB	ECM Regulators		P04080	Cystatin-B	CSTB	ECM Regulators
P10619	Lysosomal pro	CTSA	ECM Regulators		P10619	Lysosomal pro	CTSA	ECM Regulators
P07858	Cathepsin B	CTSB	ECM Regulators		P07858	Cathepsin B	CTSB	ECM Regulators
P53634	Dipeptidyl pep	CTSC	ECM Regulators		P53634	Dipeptidyl pep	CTSC	ECM Regulators
P07339	Cathepsin D	CTSD	ECM Regulators		P07339	Cathepsin D	CTSD	ECM Regulators
P08311	Cathepsin G	CTSG	ECM Regulators		P08311	Cathepsin G	CTSG	ECM Regulators
P07585	Decorin	DCN	Proteoglycans		P07585	Decorin	DCN	Proteoglycans
Q9Y6C2	EMILIN-1	EMILIN1	ECM Glycoproteins		Q9Y6C2	EMILIN-1	EMILIN1	ECM Glycoproteins
P00488	Coagulation fa	F13A1	ECM Regulators		P00488	Coagulation fa	F13A1	ECM Regulators
P00734	Prothrombin	F2	ECM Regulators		P00734	Prothrombin	F2	ECM Regulators
P23142	Fibulin-1	FBLN1	ECM Glycoproteins		P23142	Fibulin-1	FBLN1	ECM Glycoproteins
P98095	Fibulin-2	FBLN2	ECM Glycoproteins		P98095	Fibulin-2	FBLN2	ECM Glycoproteins
P35555	Fibrillin-1	FBN1	ECM Glycoproteins		P35555	Fibrillin-1	FBN1	ECM Glycoproteins
P35556	Fibrillin-2	FBN2	ECM Glycoproteins		P35556	Fibrillin-2	FBN2	ECM Glycoproteins
P02671	Fibrinogen alp	FGA	ECM Glycoproteins		P02671	Fibrinogen alp	FGA	ECM Glycoproteins

P02675	Fibrinogen be	FGB	ECM Glycoproteins		P02675	Fibrinogen be	FGB	ECM Glycoproteins
P02679	Fibrinogen ga	FGG	ECM Glycoproteins		P02679	Fibrinogen ga	FGG	ECM Glycoproteins
Q08830	Fibrinogen-lik	FGL1	ECM Glycoproteins		Q08830	Fibrinogen-lik	FGL1	ECM Glycoproteins
P02751	Fibronectin	FN1	ECM Glycoproteins		P02751	Fibronectin	FN1	ECM Glycoproteins
Q86XX4	Extracellular m	FRAS1	ECM Glycoproteins		Q86XX4	Extracellular m	FRAS1	ECM Glycoproteins
P51610	Host cell facto	HCFC1	Secreted Factors		P51610	Host cell facto	HCFC1	Secreted Factors
P02790	Hemopexin	HPX	ECM-affiliated Proteins		P02790	Hemopexin	HPX	ECM-affiliated Proteins
P98160	Basement mer	HSPG2	Proteoglycans		P98160	Basement me	HSPG2	Proteoglycans
P18065	Insulin-like gr	IGFBP2	ECM Glycoproteins		P18065	Insulin-like gr	IGFBP2	ECM Glycoproteins
Q16270	Insulin-like gr	IGFBP7	ECM Glycoproteins		Q16270	Insulin-like gr	IGFBP7	ECM Glycoproteins
Q14624	Inter-alpha-tr	ITIH4	ECM Regulators		Q14624	Inter-alpha-tr	ITIH4	ECM Regulators
Q86UX2	Inter-alpha-tr	ITIH5	ECM Regulators		Q86UX2	Inter-alpha-tr	ITIH5	ECM Regulators
P24043	Laminin subur	LAMA2	ECM Glycoproteins		P24043	Laminin subur	LAMA2	ECM Glycoproteins
Q16363	Laminin subur	LAMA4	ECM Glycoproteins		Q16363	Laminin subur	LAMA4	ECM Glycoproteins
O15230	Laminin subur	LAMA5	ECM Glycoproteins		O15230	Laminin subur	LAMA5	ECM Glycoproteins
P07942	Laminin subur	LAMB1	ECM Glycoproteins		P07942	Laminin subur	LAMB1	ECM Glycoproteins
P55268	Laminin subur	LAMB2	ECM Glycoproteins		P55268	Laminin subur	LAMB2	ECM Glycoproteins
P11047	Laminin subur	LAMC1	ECM Glycoproteins		P11047	Laminin subur	LAMC1	ECM Glycoproteins
Q9Y6N6	Laminin subur	LAMC3	ECM Glycoproteins		Q9Y6N6	Laminin subur	LAMC3	ECM Glycoproteins
P09382	Galectin-1	LGALS1	ECM-affiliated Proteins		P09382	Galectin-1	LGALS1	ECM-affiliated Proteins
P05162	Galectin-2	LGALS2	ECM-affiliated Proteins		P05162	Galectin-2	LGALS2	ECM-affiliated Proteins
P17931	Galectin-3	LGALS3	ECM-affiliated Proteins		P17931	Galectin-3	LGALS3	ECM-affiliated Proteins
P56470	Galectin-4	LGALS4	ECM-affiliated Proteins		P56470	Galectin-4	LGALS4	ECM-affiliated Proteins
O00214	Galectin-8	LGALS8	ECM-affiliated Proteins		O00214	Galectin-8	LGALS8	ECM-affiliated Proteins
P49257	Protein ERGIC-	LMAN1	ECM-affiliated Proteins		P49257	Protein ERGIC	LMAN1	ECM-affiliated Proteins
P51884	Lumican	LUM	Proteoglycans		P51884	Lumican	LUM	Proteoglycans
P55001	Microfibrillar-	MFAP2	ECM Glycoproteins		P55001	Microfibrillar-	MFAP2	ECM Glycoproteins
Q08431	Lactadherin	MFGE8	ECM Glycoproteins		Q08431	Lactadherin	MFGE8	ECM Glycoproteins
P14543	Nidogen-1	NID1	ECM Glycoproteins		P14543	Nidogen-1	NID1	ECM Glycoproteins
Q14112	Nidogen-2	NID2	ECM Glycoproteins		Q14112	Nidogen-2	NID2	ECM Glycoproteins
P20774	Mimecan	OGN	Proteoglycans		P20774	Mimecan	OGN	Proteoglycans
Q15113	Procollagen C-	PCOLCE	ECM Glycoproteins		Q15113	Procollagen C-	PCOLCE	ECM Glycoproteins
P00747	Plasminogen	PLG	ECM Regulators		P00747	Plasminogen	PLG	ECM Regulators
O15031	Plexin-B2	PLXNB2	ECM-affiliated Proteins		O15031	Plexin-B2	PLXNB2	ECM-affiliated Proteins
Q15063	Periostin	POSTN	ECM Glycoproteins		Q15063	Periostin	POSTN	ECM Glycoproteins
P51888	Prolargin	PRELP	Proteoglycans		P51888	Prolargin	PRELP	Proteoglycans
P07477	Trypsin-1	PRSS1	ECM Regulators		P07477	Trypsin-1	PRSS1	ECM Regulators
P07478	Trypsin-2	PRSS2	ECM Regulators		P07478	Trypsin-2	PRSS2	ECM Regulators
P35030	Trypsin-3	PRSS3	ECM Regulators		P35030	Trypsin-3	PRSS3	ECM Regulators
P05451	Lithostathine-	REG1A	ECM-affiliated Proteins		P05451	Lithostathine-	REG1A	ECM-affiliated Proteins
P48304	Lithostathine-	REG1B	ECM-affiliated Proteins		P48304	Lithostathine-	REG1B	ECM-affiliated Proteins
Q06141	Regenerating i	REG3A	ECM-affiliated Proteins		Q06141	Regenerating	REG3A	ECM-affiliated Proteins
P60903	Protein S100-A	S100A10	Secreted Factors		P60903	Protein S100-A	S100A10	Secreted Factors
P31949	Protein S100-A	S100A11	Secreted Factors		P31949	Protein S100-A	S100A11	Secreted Factors
Q99584	Protein S100-A	S100A13	Secreted Factors		Q99584	Protein S100-A	S100A13	Secreted Factors
Q96FQ6	Protein S100-A	S100A16	Secreted Factors		Q96FQ6	Protein S100-A	S100A16	Secreted Factors
P26447	Protein S100-A	S100A4	Secreted Factors		P26447	Protein S100-A	S100A4	Secreted Factors
P05109	Protein S100-A	S100A8	Secreted Factors		P05109	Protein S100-A	S100A8	Secreted Factors
P06702	Protein S100-A	S100A9	Secreted Factors		P06702	Protein S100-A	S100A9	Secreted Factors
P01009	Alpha-1-antitr	SERPINA1	ECM Regulators		P01009	Alpha-1-antitr	SERPINA1	ECM Regulators
P01011	Alpha-1-antich	SERPINA3	ECM Regulators		P01011	Alpha-1-antich	SERPINA3	ECM Regulators
P30740	Leukocyte elas	SERPINB1	ECM Regulators		P30740	Leukocyte ela	SERPINB1	ECM Regulators
P35237	Serpin B6	SERPINB6	ECM Regulators		P35237	Serpin B6	SERPINB6	ECM Regulators
P05155	Plasma protea	SERPING1	ECM Regulators		P05155	Plasma protea	SERPING1	ECM Regulators
P50454	Serpin H1	SERPINH1	ECM Regulators		P50454	Serpin H1	SERPINH1	ECM Regulators
O75830	Serpin I2	SERPINI2	ECM Regulators		O75830	Serpin I2	SERPINI2	ECM Regulators
Q15582	Transforming g	TGFBI	ECM Glycoproteins		Q15582	Transforming	TGFBI	ECM Glycoproteins
P21980	Protein-glutan	TGM2	ECM Regulators		P21980	Protein-glutan	TGM2	ECM Regulators
P07996	Thrombospon	THBS1	ECM Glycoproteins		P07996	Thrombospon	THBS1	ECM Glycoproteins
Q9GZM7	Tubulointersti	TINAGL1	ECM Glycoproteins		Q9GZM7	Tubulointersti	TINAGL1	ECM Glycoproteins
P22105	Tenascin-X	TNXB	ECM Glycoproteins		P22105	Tenascin-X	TNXB	ECM Glycoproteins

P04004	Vitronectin	VTN	ECM Glycoproteins		P04004	Vitronectin	VTN	ECM Glycoproteins
Q6PCB0	von Willebran	VWA1	ECM Glycoproteins		Q6PCB0	von Willebran	VWA1	ECM Glycoproteins
P04275	von Willebran	VWF	ECM Glycoproteins		P04275	von Willebran	VWF	ECM Glycoproteins
Q13444	Disintegrin an	ADAM15	ECM Regulators		P13942	Collagen alph	COL11A2	Collagens
Q13443	Disintegrin an	ADAM9	ECM Regulators		P15502	Elastin	ELN	ECM Glycoproteins
Q8IUX7	Adipocyte enh	AEBP1	ECM Glycoproteins		Q9NR99	Matrix-remod	MXRA5	ECM Glycoproteins
P02745	Complement (	C1QA	ECM-affiliated Proteins		Q6UXI9	Nephronectin	NPNT	ECM Glycoproteins
Q6YHK3	CD109 antigen	CD109	ECM Regulators		P05154	Plasma serine	SERPINA5	ECM Regulators
P0CG36	Cryptic family	CFC1B	Secreted Factors		P01008	Antithrombin	SERPINC1	ECM Regulators
Q86T13	C-type lectin d	CLEC14A	ECM-affiliated Proteins		P01023	Alpha-2-macr	A2M	ECM Regulators
Q96P44	Collagen alpha	COL21A1	Collagens		P06703	Protein S100-A	S100A6	Secreted Factors
A6NMZ7	Collagen alpha	COL6A6	Collagens		P0CG37	Cryptic protei	CFC1	Secreted Factors
Q5KU26	Collectin-12	COLEC12	ECM-affiliated Proteins					
Q6UVK1	Chondroitin su	CSPG4	ECM-affiliated Proteins					
P09668	Pro-cathepsin	CTSH	ECM Regulators					
Q9UBR2	Cathepsin Z	CTSZ	ECM Regulators					
Q07507	Dermatoponti	DPT	ECM Glycoproteins					
O95967	EGF-containin	EFEMP2	ECM Glycoproteins					
P08246	Neutrophil ela	ELANE	ECM Regulators					
Q9BXX0	EMILIN-2	EMILIN2	ECM Glycoproteins					
Q9NT22	EMILIN-3	EMILIN3	ECM Glycoproteins					
Q9UBX5	Fibulin-5	FBLN5	ECM Glycoproteins					
Q75N90	Fibrillin-3	FBN3	ECM Glycoproteins					
Q06828	Fibromodulin	FMOD	Proteoglycans					
Q5SZK8	FRAS1-related	FREM2	ECM-affiliated Proteins					
P51654	Glypican-3	GPC3	ECM-affiliated Proteins					
O75487	Glypican-4	GPC4	ECM-affiliated Proteins					
Q9Y625	Glypican-6	GPC6	ECM-affiliated Proteins					
P04196	Histidine-rich	HRG	ECM Regulators					
P24593	Insulin-like gr	IGFBP5	ECM Glycoproteins					
P19827	Inter-alpha-tr	ITIH1	ECM Regulators					
P19823	Inter-alpha-tr	ITIH2	ECM Regulators					
P25391	Laminin subur	LAMA1	ECM Glycoproteins					
Q32P28	Prolyl 3-hydro	LEPRE1	ECM Regulators					
Q8IVL6	Prolyl 3-hydro	LEPREL2	ECM Regulators					
Q14766	Latent-transfo	LTBP1	ECM Glycoproteins					
Q8N2S1	Latent-transfo	LTBP4	ECM Glycoproteins					
O00339	Matrilin-2	MATN2	ECM Glycoproteins					
P21741	Midkine	MDK	Secreted Factors					
Q13361	Microfibrillar-	MFAP5	ECM Glycoproteins					
P50281	Matrix metallo	MMP14	ECM Regulators					
P08253	72 kDa type IV	MMP2	ECM Regulators					
Q9H8L6	Multimerin-2	MMRN2	ECM Glycoproteins					
P13674	Prolyl 4-hydro	P4HA1	ECM Regulators					
O15460	Prolyl 4-hydro	P4HA2	ECM Regulators					
Q02809	Procollagen-ly	PLOD1	ECM Regulators					
O60568	Multifunctiona	PLOD3	ECM Regulators					
O75051	Plexin-A2	PLXNA2	ECM-affiliated Proteins					
P13727	Bone marrow	PRG2	Proteoglycans					
Q92626	Peroxidasin ho	PXDN	ECM Glycoproteins					
Q6UW15	Regenerating	REG3G	ECM-affiliated Proteins					
P80511	Protein S100-A	S100A12	Secreted Factors					
P50453	Serpin B9	SERPINB9	ECM Regulators					
P36955	Pigment epith	SERPINF1	ECM Regulators					
Q8N474	Secreted frizzl	SFRP1	Secreted Factors					
P09486	SPARC	SPARC	ECM Glycoproteins					
Q9HCB6	Spondin-1	SPON1	ECM Glycoproteins					
Q6ZMP0	Thrombospon	THSD4	ECM Glycoproteins					
P01033	Metalloprotei	TIMP1	ECM Regulators					
P13611	Versican core	VCAN	Proteoglycans					
Q8N398	von Willebran	VWA5B2	ECM Glycoproteins					
O96014	Protein Wnt-1	WNT11	Secreted Factors					

Table S5. List of all quantified ECM proteins. Log_2 DiLeu reporter ion intensities of each protein in all samples are shown. Significantly changed proteins among age groups (one-way ANOVA, FDR 0.05) are highlighted in orange. Protein accession, *p* values and log_2(fold change) for all pairwise comparisons are shown. Significantly down-regulated proteins (two-sided *t* test, $p < 0.05$, $FC < 0.5$) are highlighted in green and up-regulated ones (two-sided *t* test, $p < 0.05$, $FC > 2$) in red.

Quantified: ECM Proteins — Significantly changed among age groups (one-way ANOVA, FDR 0.05)

Accession	Description	Gene name	ECM category	F_7	F_8	F_9	F_10	F_11	F_12	J_8	J_22	J_46	J_59	J_67	Y_12	Y_31	Y_51	Y_57	Y_61	Y_64	O_14	O_20	O_48	O_52	O_54	O_68
Q05707	Collagen alpl	COL14A1	Collagens	12.28517	12.61627	12.36024	12.06733	12.11625	12.5275	11.30235	11.10276	11.31538	10.94764	10.92184	11.01255	8.335615	8.074736	8.665009	8.386578	8.44311	8.22979	7.897175	8.446309	8.542411	9.999971	7.732981
Q99715	Collagen alpl	COL12A1	Collagens	11.78455	11.70546	11.68216	11.8966	11.79722	11.88463	11.31277	10.75455	11.21632	10.90079	10.73699	10.8779	9.157544	9.247213	9.71862	8.702624	9.067904	9.308889	9.000909	9.401832	9.126645	9.383077	8.951067
P35556	Fibrillin-2	FBN2	ECM Glycoprot	10.44108	10.65526	10.47715	10.52229	10.8396	10.58167	10.3771	9.550747	10.38165	9.661065	9.546701	9.682292	7.052427	6.977955	7.081956	7.503047	6.282041	6.732689	7.036499	6.711808	6.717028	5.466368	6.811267
Q15063	Periostin	POSTN	ECM Glycoprot	10.96239	10.84337	10.91692	11.07948	11.22545	11.4624	10.14453	9.933101	10.13584	10.03796	9.914236	10.15949	8.37999	8.265463	9.418394	8.564928	8.017195	8.729836	7.911787	8.707256	8.347266	9.407013	7.728118
Q16363	Laminin subu	LAMA4	ECM Glycoprot	9.277752	9.293931	9.461888	9.486232	9.302867	9.143893	9.359091	9.706324	9.580635	9.504819	9.277985	9.24103	9.427337	8.699516	8.819983	8.33351	7.898936	8.183179	8.469993	8.25749	8.005812	10.03398	7.523315
Q9Y6C2	EMILIN-1	EMILIN1	ECM Glycoprot	11.1203	11.015	11.17767	11.20163	10.99626	11.13872	11.34413	11.03528	11.37401	11.21183	11.29341	11.1238	10.21831	10.33831	10.88109	10.52794	10.03201	10.34559	10.29712	10.77252	10.39722	11.24683	9.931952
P20774	Mimecan	OGN	Proteoglycans	10.50571	10.60001	10.48985	10.23948	10.08068	10.6279	9.196725	10.19955	9.09566	9.643856	10.06878	9.749367	9.30892	8.802164	8.725276	8.39136	8.152748	8.393835	8.36272	8.197861	8.129558	9.649967	8.129453
P02458	Collagen alpl	COL2A1	Collagens	9.466179	9.209941	9.136478	9.378078	9.877131	9.646198	9.687726	8.861708	9.786106	8.716648	8.715276	9.022923	8.833305	8.283484	8.897985	8.062251	7.589664	9.004616	8.285621	8.588086	8.618514	9.321606	7.557706
P02462	Collagen alpl	COL4A1	Collagens	9.971112	10.03122	9.919608	9.845961	10.3614	9.672956	10.46577	11.73043	11.15006	11.17111	11.11745	11.59012	11.4086	11.05893	12.10103	10.27081	9.860288	11.49733	11.03801	11.89737	10.35335	12.36612	9.790087
P08572	Collagen alpl	COL4A2	Collagens	9.940313	9.978996	9.98014	9.936196	10.18314	9.828772	10.46944	11.17055	11.13994	10.58252	10.48905	11.18704	11.13877	10.79206	11.56432	10.49026	10.09034	11.28777	10.63824	11.61875	10.34601	11.8076	10.05102
P12107	Collagen alpl	COL11A1	Collagens	7.380245	7.412781	7.259273	7.400879	7.651769	7.171927	8.815703	7.234578	10.7115	7.96116	7.578184	7.516488	8.172613	8.283484	9.088993	8.56228	8.077073	8.899323	8.975769	8.438541	8.236785	8.930107	8.085266
Q07092	Collagen alpl	COL16A1	Collagens	7.694184	7.423746	7.579693	7.500643	8.162392	7.208478	9.182891	7.177918	9.090113	7.780047	7.343408	7.807355	8.43162	8.303247	8.916629	8.823015	8.865369	8.777205	8.711671	9.047666	8.953755	9.013132	8.015942
O15230	Laminin subu	LAMA5	ECM Glycoprot	11.39484	11.48397	11.57242	11.62063	11.5426	11.40333	11.83672	12.65253	12.5155	12.45064	12.50343	12.48054	12.48169	12.2994	12.68142	11.90787	11.51023	12.34767	12.07194	12.79296	11.90513	13.17388	11.33051
P02671	Fibrinogen al	FGA	ECM Glycoprot	9.141341	9.542065	9.308339	9.430871	9.910043	9.817943	9.067165	10.51294	9.704077	10.45656	10.54767	9.089848	10.3984	10.24142	10.91943	10.56109	10.74959	11.89182	11.08986	10.7324	10.50827	13.37873	10.19206
P02675	Fibrinogen b	FGB	ECM Glycoprot	9.628992	9.970393	9.733863	9.944712	10.33539	10.08985	10.47938	10.86805	11.51875	11.13064	10.98826	10.1486	10.79783	10.77588	11.40402	10.88717	11.04414	12.45694	11.36029	11.20422	11.368	13.61229	10.5068
P02679	Fibrinogen g	FGG	ECM Glycoprot	9.423326	9.912889	9.646019	9.690697	10.11035	9.99859	10.14644	10.83487	12.36761	11.07294	10.8892	9.978138	10.80457	10.85735	11.53887	10.98287	11.27557	12.52372	11.39527	11.37836	11.17177	13.50466	10.77767
P02751	Fibronectin	FN1	ECM Glycoprot	12.07965	12.18072	12.18146	12.24517	12.31535	12.35339	15.4397	11.73666	11.91838	12.01451	11.75597	11.52028	10.69277	10.77824	12.62127	10.92199	10.8793	11.24	11.04436	11.3096	11.08382	12.25292	10.35847
P04004	Vitronectin	VTN	ECM Glycoprot	8.165912	8.311975	8.146696	8.261154	8.215775	8.553053	8.802194	8.502633	9.387156	8.563005	8.510171	8.715619	9.243246	9.242103	9.619419	9.351725	9.014016	9.566377	9.429051	9.733377	9.334067	10.33647	8.731524
P04275	von Willebra	VWF	ECM Glycoprot	10.28771	10.1334	10.31436	10.4544	10.16792	10.43056	9.917372	9.618569	9.985415	9.615078	9.436087	9.657318	7.280364	7.66123	7.656263	7.430808	6.961342	7.191767	7.521339	7.715686	7.373305	8.601556	6.859321
P07996	Thrombospo	THBS1	ECM Glycoprot	7.510171	7.770168	7.661778	7.306518	7.67948	7.539159	6.860466	7.974988	6.821455	7.123086	7.361944	6.827819	8.817025	8.769384	9.199345	10.07958	8.979386	8.819378	8.566039	8.952696	9.045683	8.363963	8.707021
P18065	Insulin-like g	IGFBP2	ECM Glycoprot	8.006747	7.436295	7.567576	8.307884	7.638798	8.163901	7.412781	6.602884	7.426265	8.467198	8.632996	7.687201	8.448025	8.438072	9.195444	8.425571	9.033773	8.136496	9.279337	8.755669	9.280687	8.16422	7.869955
P22105	Tenascin-X	TNXB	ECM Glycoprot	8.921246	8.797013	8.673486	8.67772	8.589838	8.642052	9.389954	9.009549	9.011786	8.68545	8.924218	8.282162	6.2979	5.162469	4.85097	4.390763	4.234247	4.661411	4.34129	3.834279	4.776002	6.533598	3.638706
P24043	Laminin subu	LAMA2	ECM Glycoprot	11.56672	11.65616	11.76748	11.73974	11.79051	11.51052	12.02607	12.93297	12.78097	12.74878	12.52386	12.61142	13.14005	12.86263	13.52147	12.08102	11.71127	12.97732	12.41543	13.07029	11.93452	13.83998	11.42428
P35555	Fibrillin-1	FBN1	ECM Glycoprot	13.30444	13.34474	13.2216	13.37424	13.35979	13.54636	12.64791	13.0179	12.92264	12.81196	12.79338	12.92533	12.52264	12.16805	12.92476	11.91702	11.6237	12.52056	12.36366	12.36112	12.10623	13.98537	11.32818
P55268	Laminin subu	LAMB2	ECM Glycoprot	10.80816	10.54863	10.44853	10.51175	10.46485	10.439	11.01297	11.58021	11.1298	11.03714	10.96838	10.96087	11.8423	11.28688	11.57385	11.26683	10.51248	11.10246	11.11134	11.33573	11.01174	12.20525	10.72666
P98095	Fibulin-2	FBLN2	ECM Glycoprot	6.817623	6.915282	6.837943	6.787902	6.912889	6.658212	7.78398	6.421223	7.575917	6.874059	6.723832	6.794416	8.312573	8.103575	8.470384	7.426154	8.226027	9.205613	8.41313	8.010666	7.92847	9.339627	7.531153
Q08431	Lactadherin	MFGE8	ECM Glycoprot	7.968091	8.064473	7.982423	7.841722	7.99605	7.875288	8.090642	8.936343	8.609549	8.078418	8.263504	8.116343	7.164547	7.48093	8.250825	6.531737	6.250548	7.449798	6.879793	7.644753	6.5876	8.677519	6.201387
Q08830	Fibrinogen-li	FGL1	ECM Glycoprot	7.371559	7.451211	7.401733	7.338959	7.415319	7.254556	9.033698	8.763544	9.045214	7.818902	8.247928	7.758223	8.590025	8.384037	8.629704	8.771662	8.544102	8.388865	8.22909	8.641634	8.572694	7.584864	8.191529
Q15113	Procollagen	PCOLCE	ECM Glycoprot	9.05474	9.349613	9.108002	8.835103	9.17742	8.878051	9.220137	8.910792	9.417642	8.840778	8.722124	8.702866	7.93207	8.211402	8.629704	8.55537	8.174271	8.45872	8.341925	8.189341	8.264877	8.458621	8.020721
Q15582	Transforming	TGFBI	ECM Glycoprot	10.47371	10.29944	10.3423	10.38219	10.45841	10.32823	10.00408	10.39264	10.38802	10.51037	10.70018	10.15823	9.663149	9.289111	9.559434	9.158334	8.875916	9.305266	8.959727	9.579146	9.1379	10.45327	8.547806
Q6PCB0	von Willebra	VWA1	ECM Glycoprot	9.022645	9.047397	9.060426	9.041112	9.126704	8.896938	9.321703	9.844078	9.666579	9.982851	10.18525	9.611209	8.892662	8.757973	9.250825	8.343418	7.809155	8.640129	8.489033	9.021106	8.2198	9.436463	7.592367
Q6UXH1	Cysteine-rich	CRELD2	ECM Glycoprot	8.228819	8.022923	8.097506	8.043847	8.577051	8.021812	9.073874	8.37243	8.739443	8.441284	8.72928	9.266084	9.31655	8.530958	8.555289	8.941893	8.96464	8.764789	9.255612	8.85148	9.550857	8.186519	8.866861
Q9GZM7	Tubulointers	TINAGL1	ECM Glycoprot	8.171927	8.218684	8.217231	8.422906	8.181897	8.168922	8.455327	9.370252	8.437544	9.046851	9.16088	9.266552	9.748774	9.570365	9.794188	9.517314	8.645694	9.896044	9.267356	9.918308	9.553571	10.84902	8.574599
Q9Y6N6	Laminin subu	LAMC3	ECM Glycoprot	8.864805	8.814423	8.995202	9.010948	9.053111	8.723831	9.501837	9.682468	9.850031	9.349834	9.237688	9.418907	10.10755	10.13001	10.40648	9.551103	8.970802	9.963207	9.579291	10.9242	9.474152	11.28651	8.768462
P21810	Biglycan	BGN	Proteoglycans	9.81298	9.682819	9.707014	9.764374	9.481194	9.717162	9.34607	9.729791	8.976134	9.570805	9.668708	9.564912	8.94497	8.473391	9.278912	8.382379	8.28327	8.875092	8.087875	8.633805	8.55655	9.922155	7.910008
P51888	Prolargin	PRELP	Proteoglycans	8.858292	8.829088	8.728261	8.748864	8.535665	8.622784	8.654278	9.733355	8.782343	8.866661	9.324181	8.897543	10.23574	10.06705	10.35082	10.20679	10.27599	10.11483	10.21322	10.27739	10.26272	10.59351	9.931318
Q9BXN1	Asporin	ASPN	Proteoglycans	8.338068	8.083214	8.137504	7.921246	7.990387	7.804776	8.052568	8.316507	8.955069	8.013462	8.670302	9.166916	10.76402	9.945926	10.67714	9.871533	9.945366	10.19266	9.825673	9.990579	10.08072	10.06521	10.28703
O14672	Disintegrin a	ADAM10	ECM Regulator	8.062316	7.984703	8.090113	7.993221	8.020146	7.968091	8.306518	8.145168	8.521208	8.14873	8.034524	8.178416	7.365989	7.843243	8.117007	8.848891	7.628644	7.718985	7.404647	8.13344	7.455699	7.836851	7.485781
O75830	Serpin I2	SERPINI2	ECM Regulator	9.631722	9.693661	9.134683	9.247215	9.88249	8.968955	10.27764	9.80187	9.782998	9.182394	9.648717	9.619853	9.781985	10.06396	9.70097	9.863811	10.44222	9.767424	10.03566	10.52219	10.41584	9.330645	11.00714
P00488	Coagulation	F13A1	ECM Regulator	9.010948	8.952159	8.85767	8.60696	8.800576	8.633722	8.906891	9.015972	9.467402	8.982138	9.738092	9.137504	8.421194	8.853109	9.674211	8.567573	8.723183	8.892756	8.681391	8.670999	8.758231	8.816978	8.180893
P00734	Prothrombin	F2	ECM Regulator	8.201633	8.345183	7.975562	8.124122	8.76686	8.264442	9.33539	8.405993	8.904785	8.871597	9.073606	9.40216	8.959877	9.185369	9.243631	9.512931	9.099371	9.351668	9.418256	9.24795	9.526476	9.531277	9.188347
P00747	Plasminogen	PLG	ECM Regulator	8.826549	8.849249	8.73098	8.691744	8.929554	8.759555	9.034799	8.882031	9.214561	9.413839	9.052568	8.893605	8.357652	8.730988	8.80377	9.071221	8.817387	8.385125	8.660676	8.679135	8.889476	8.960618	8.588079
P01009	Alpha-1-anti	SERPINA1	ECM Regulator	9.692616	9.581013	9.445015	9.545737	9.260684	9.786269	9.842036	10.3516	9.619853	10.33651	10.81362	9.881267	9.586736	10.17486	9.994201	11.40029	11.39786	10.34511	10.00512	10.52544	11.11263	10.53895	11.93549
P01011	Alpha-1-anti	SERPINA3	ECM Regulator	8.677014	8.751879	8.7303	8.595817	8.455327	8.52943	9.337175	10.92719	9.937963	9.861708	10.13481	9.438376	9.679451	10.75928	10.5637	10.16673	10.6011	10.32133	10.1998	10.40229	9.861056	10.23011	10.16064
P01034	Cystatin-C	CST3	ECM Regulator	9.208966	8.68159	8.970106	8.917372	8.589089	9.522778	8.435462	8.386724	8.171927	8.136992	8.109047	7.908092	8.624112	8.500465	8.57289	8.074943	8.836861	8.116339	7.916057	8.316058	7.909104	8.165717	8.416745
P04080	Cystatin-B	CSTB	ECM Regulator	8.51767	8.858603	8.816664	8.720757	8.404716	8.743825	8.618018	8.908092	8.306062	8.667466	8.936638	8.60807	9.813981	10.02442	9.717384	10.69273	10.8216	10.29351	9.848638	9.935604	10.13772	10.63424	10.47218
P05155	Plasma prote	SERPING1	ECM Regulator	7.714245	7.886306	7.805421	7.811856	7.68159	7.823367	8.310159	8.698358	8.361067	8.752882	9.038096	8.15127	7.206577	7.676465	7.862645	9.347424	8.263994	7.556713	7.505559	7.850581	8.156942	8.010218	7.941437
P07477	Trypsin-1	PRSS1	ECM Regulator	11.48854	11.21584	11.30572	11.63427	11.47467	11.40934	13.49125	12.15013	13.4434	13.02034	13.08187	13.16811	14.64671	13.97737	14.3787	13.1623	14.04673	13.70685	14.32329	14.25228	14.34816	13.77852	13.69567
P07478	Trypsin-2	PRSS2	ECM Regulator	9.43567	9.307428	9.304693	9.634811	9.773469	9.383704	9.834945	11.17005	9.352706	10.8149	11.83044	11.60719	11.46096	12.35395	11.63815	11.61333	12.59406	12.14592	11.2159	12.26811	11.81594	11.70995	13.1957
P07858	Cathepsin B	CTSB	ECM Regulator	8.916775	8.484622	8.570615	8.544965	8.647458	8.637712	9.784962	8.627898	9.181152	8.907492	9.046033	8.720757	7.822628	8.296032	8.437135	8.698288	8.622395	8.586325	8.07036	8.226596	7.801798	8.288314	8.531153
P08217	Chymotrypsi	CELA2A	ECM Regulator	8.180903	8.032872	7.960002	8.100662	8.070389	8.055824	9.454916	9.491252	9.058533	9.592083	10.68141	11.43728	11.60338	11.2829	10.86287	11.01877	11.46379	11.46397	11.79939	12.2546	12.16304	11.40472	11.74085
P08218	Chymotrypsi	CELA2B	ECM Regulator	6.551516	6.574404	6.581954	6.525129	6.656782	6.427941	8.488242	7.558421	7.990955	8.657497	9.443565	9.900565	8.493224	9.320806	8.659355	8.521411	9.788137	9.124585	9.242752	8.996899	9.372076	8.161221	10.28041
P08861	Chymotrypsi	CELA3B	ECM Regulator	8.465158	8.32373	8.262565	8.916476	8.173927	8.421644	9.581201	9.058262	9.392103	9.446257	9.615262	9.83747	9.24707	9.853557	9.863539	9.12928	10.33052	9.626205	9.859984	9.442132	9.855569	9.539963	11.20165
P09093	Chymotrypsi	CELA3A	ECM Regulator	11.44827	11.27961	10.95441	11.28996	10.82106	11.4351	12.12996	11.18258	11.35552	11.79263	11.79944	11.75814	12.3002	12.23642	12.82021	12.13958	13.08749	11.82169	12.85293	12.2692	12.49664	12.12194	13.1019
P10619	Lysosomal pr	CTSA	ECM Regulator	6.719731	6.289097	6.106432	6.0917	5.620586	6.151778	5.311067	6.405993	5.916477	6.821455	6.75889	6.238405	8.445001	8.792408	8.649021	9.136777	9.322674	8.019968	7.59562	8.64631	9.283306	7.752674	9.025088
P30740	Leukocyte el	SERPINB1	ECM Regulator	8.563005	8.898752	8.755889	8.716648	8.750204	8.489045	9.587777	10.17854	9.584023	9.7975	10.26913	10.71896	10.74485	10.72502	10.35987	10.58085	11.22697	10.38544	10.31566	10.17125	10.28069	10.20739	10.90258
P35030	Trypsin-3	PRSS3	ECM Regulator	8.286789	8.209454	8.234099	8.504621	8.846744	8.346957	8.517276	10.15102	8.358432	9.539159	10.67384	10.24115	11.08033	10.893	10.40479	10.29861	11.50521	10.72971	10.1933	11.38437	10.542	10.31099	11.75616
P35237	Serpin B6	SERPINB6	ECM Regulator	8.744161	9.008709	8.897845	9.00422	8.642052	8.409391	9.139551	8.918267	9.127995	8.79831	9.203592	9.049849	9.47999	9.980706	9.429607	10.81625	10.83989	8.8888	8.796559	9.710981	9.916502	8.862468	10.90721
P50454	Serpin H1	SERPINH1	ECM Regulator	11.4922	11.60886	11.43546	11.68168	11.5158	11.93325	10.32351	10.31061	10.21177	10.62699	10.32868	10.2662	9.688173	9.954907	10.46936	9.814582	10.01878	9.71129	9.61045	9.621456	9.791029	9.501052	9.266554
P53634	Dipeptidyl pe	CTSC	ECM Regulator	8.945444	9.058262	8.78725	8.885086	8.81986	8.77742	8.493455	8.840464	8.395748	8.551901	8.782671	8.81025	6.35656	7.05164	7.843187	6.268341	6.83468	6.267266	7.509998	6.336999	7.153221	6.946399	6.471839
Q14624	Inter-alpha-t	ITIH4	ECM Regulator	6.289097	7.10957	6.776104	6.395748	6.498251	6.78136	7.317412	9.008988	7.585714	9.152031	9.936785	6.436295	6.115181	7.207911	7.071301	8.085311	7.678697	7.205966	7.588148	7.117811	7.410906	6.960187	6.82694
Q86UX2	Inter-alpha-t	ITIH5	ECM Regulator	7.451211	7.478163	7.423746	7.795065	7.343408	7.310158	8.427103	8.33673	8.280307	7.844235	8.152285	8.276124	8.435891	8.627137	9.081712	9.25663	8.158519	8.670655	8.389374	9.296357	8.825345	8.918163	8.185157
O15031	Plexin-B2	PLXNB2	ECM-affiliated	9.508191	9.506804	9.395535	9.574215	9.424796	9.451417	9.687026	9.169173	9.833523	9.182891	8.982993	9.241983	7.919053	8.219746	8.250825	8.169167	7.917773	7.755164	7.981892	8.04255	8.122995	7.81316	7.560991
P02746	Complement	C1QB	ECM-affiliated	8.190318	8.611394	8.22062	8.420802	8.56453	8.449148	8.378511	8.856737	8.684398	8.679481	9.295769	7.705978	6.813277	7.056219	7.421433	7.715555	6.869332	6.781952	6.64687	7.231446	7.287225	7.304037	6.469558
P02790	Hemopexin	HPX	ECM-affiliated	8.063396	8.014579	8.004501	7.985273	7.930146	8.006186	8.654994	8.389309	8.925109	8.797013	8.865424	8.514912	7.467229	8.111816	7.789256	7.717461	8.805022	7.418434	7.767082	7.746987	7.859961	8.481647	8.733466
P04083	Annexin A1	ANXA1	ECM-affiliated	9.693835	10.28123	10.18475	10.31945	9.872367	10.33114	9.246503	9.74231	9.610287	10.0588	9.760221	9.365229	9.453753	9.611105	9.719608	9.61307	10.07764	9.469355	9.499028	9.486853	9.752101	9.847619	9.729092
P05162	Galectin-2	LGALS2	ECM-affiliated	8.245553	8.435045	8.390169	8.346514	8.501041	8.110614	9.996473	8.551131	9.258095	10.8873	10.6397	8.650693	10.33365	10.36002	10.49728	10.1483	10.21845	9.921866	10.32629	9.214401	9.652328	9.032957	9.563996
P05451	Lithostathine	REG1A	ECM-affiliated	11.00661	11.03775	10.99067	10.95121	10.84933	10.8279	12.65287	13.22772	13.09795	13.37637	12.96924	11.56425	12.5315	12.6588	13.3991	13.97943	14.07148	12.76305	12.58591	12.46555	12.73538	12.13011	13.06745
P07355	Annexin A2	ANXA2	ECM-affiliated	12.42343	12.44458	12.46638	12.58818	12.53729	12.50787	11.70788	12.5843	11.55272	12.70349	12.29898	12.21474	11.26758	11.91093	12.09781	11.23193	11.46481	11.20013	12.07671	11.7643	11.38919	11.84756	11.37754
P08133	Annexin A6	ANXA6	ECM-affiliated	12.14952	12.39379	12.24457	12.54875	12.11319	12.42608	11.50338	11.51328	11.19881	11.70753	11.52914	11.37792	11.47169	11.9083	11.79992	11.26919	11.4504	11.45412	11.64796	11.8276	11.67151	11.8454	11.51952
P08758	Annexin A5	ANXA5	ECM-affiliated	9.869748	11.19309	10.19266	10.76363	9.676839	10.10852	9.676486	9.611763	9.31379	9.671718	9.749367	9.695229	10.45465	10.92524	10.03803	10.82401	11.39534	10.26836	10.16129	10.45344	10.35553	10.18005	11.76265
P09382	Galectin-1	LGALS1	ECM-affiliated	10.67136	10.92689	10.99138	10.11048	10.20677	10.36457	9.707532	9.748864	9.854712	9.705287	9.82893	9.764042	10.05338	9.992927	10.45845	10.15833	10.37566	10.63279	10.11895	10.42832	9.914392	10.49511	10.45918
P09525	Annexin A4	ANXA4	ECM-affiliated	12.7496	12.73265	12.84755	12.87556	12.61261	12.96355	12.07942	12.59024	11.93077	12.91526	12.56069	12.56808	12.04432	12.55853	12.51122	12.1414	12.4337	12.20901	12.87201	12.7778	12.57867	12.61621	12.83117
P12429	Annexin A3	ANXA3	ECM-affiliated	8.104337	8.708739	8.103812	8.313336	8.238882	8.059074	8.701826	8.584963	8.039468	8.068778	8.686151	8.324631	8.728275	9.264791	9.032471	9.945962	9.590466	9.156375	8.626692	9.097844	9.040662	8.527788	9.621765
P17931	Galectin-3	LGALS3	ECM-affiliated	9.551131	9.620586	9.616365	9.379811	9.41722	9.265849	10.36249	10.29531	10.16428	10.47644	10.27508	9.814102	8.824491	9.20896	8.978603	9.029643	9.483899	8.770533	9.24818	8.883963	9.031736	9.692139	9.641036
P20073	Annexin A7	ANXA7	ECM-affiliated	9.161635	9.889808	9.895878	9.82002	9.355792	9.947491	10.78259	9.209697	10.40312	9.314016	9.248639	9.286558	8.684844	8.752711	8.947739	9.074572	8.870651	8.807787	9.372695	9.163113	9.181011	9.108086	9.15359
P48304	Lithostathine	REG1B	ECM-affiliated	8.928963	8.87621	8.854868	8.835103	8.899357	8.73471	10.79596	11.83846	11.09044	12.08371	11.49215	9.796364	9.774778	10.96487	11.28008	12.29765	12.1891	11.07642	9.951431	10.23921	10.36777	9.929887	10.98733
P50995	Annexin A11	ANXA11	ECM-affiliated	10.65311	10.56919	10.69558	10.70079	10.43369	10.71862	10.09302	10.27298	9.85767	10.22497	10.04015	10.19081	9.363109	9.631052	9.570154	9.383579	9.792783	9.331085	9.625649	9.211248	9.297623	9.450966	9.616249
P56470	Galectin-4	LGALS4	ECM-affiliated	8.039468	8.069853	7.986411	7.95594	7.887525	7.749534	8.554589	9.357992	8.201143	9.214561	10.00211	9.02098	8.916435	8.648411	8.915767	9.783536	9.089965	8.705555	8.682895	9.072201	9.532541	8.524874	9.015542
Q06141	Regenerating	REG3A	ECM-affiliated	9.072535	9.080817	9.075212	9.001409	9.119979	8.884171	10.11009	12.22243	10.17829	11.07327	11.4431	9.565483	8.878835	9.452599	10.04257	9.485793	11.28574	8.924874	9.116727	9.1327	8.821748	8.767049	8.850398
P06702	Protein S100	S100A9	Secreted Facto	10.09606	10.37047	10.34884	9.521796	9.45697	9.892998	9.301497	10.16955	10.17542	9.839833	10.59572	9.479982	7.933797	8.415854	9.416568	8.781715	10.06253	8.669631	9.480986	8.209143	8.768546	9.282804	8.316648
P51610	Host cell fact	HCFC1	Secreted Facto	9.000564	8.714246	9.044941	8.837313	8.751209	8.793441	8.780047	8.585713	8.937815	8.52943	8.623516	8.598425	7.579585	7.778826	7.571249	7.753201	7.282609	7.748388	7.758548	7.663329	7.675935	7.447585	6.910008
P60903	Protein S100	S100A10	Secreted Facto	8.050393	8.207991	8.199673	8.249351	8.093814	8.499447	9.040564	8.377211	9.73081	9.2293	8.422485	8.468827	8.485003	8.873082	9.460371	9.435447	8.408416	8.813826	9.822265	9.302625	9.48329	9.428377	8.394207
Q99584	Protein S100	S100A13	Secreted Facto	8.354469	8.613605	8.474112	8.311522	8.231701	8.201633	8.19525	8.863876	8.416587	8.677367	8.797013	8.314243	9.157914	8.980501	10.27455	9.488028	8.670137	9.321009	9.357392	10.07143	9.011744	10.32931	8.710476

P02461	Collagen alpl	COL3A1	Collagens	14.38002	14.29905	14.27779	14.13487	14.76274	14.56737	13.91992	14.48689	15.35286	14.38522	14.52864	14.30088	15.20432	14.24939	15.55339	13.75865	13.31633	15.5151	14.12315	14.99929	14.05359	16.0542	12.99322
P12109	Collagen alpl	COL6A1	Collagens	12.3761	12.21784	12.36755	12.39962	12.1629	12.47251	11.91495	12.7687	12.44729	12.50064	12.57355	12.27035	12.51638	12.05893	12.5333	11.28116	11.13895	12.3423	11.85538	12.051	11.62743	13.23769	10.92283
P20908	Collagen alpl	COL5A1	Collagens	10.9935	10.92229	10.86821	11.01514	11.28251	11.18488	11.18059	10.67675	11.35651	10.69009	10.668	10.82026	10.85807	10.1827	11.25552	9.573634	9.055586	10.87109	9.81451	10.43073	9.550857	11.75554	8.816898
P02452	Collagen alpl	COL1A1	Collagens	14.60742	14.48455	14.44567	14.2514	14.868	14.7926	14.09926	14.79304	15.46113	14.42156	14.59907	14.60313	15.16801	14.19117	15.64762	13.68161	13.16306	15.35959	14.00991	14.64765	14.02475	16.0376	12.75487
P05997	Collagen alpl	COL5A2	Collagens	10.70295	10.55641	10.60779	10.54206	11.09994	10.73767	10.75939	10.5265	11.22689	10.55574	10.51086	10.50094	10.79511	10.34278	11.42749	10.41298	9.617166	10.88957	10.37145	10.85204	10.40761	11.65936	9.417651
P08123	Collagen alpl	COL1A2	Collagens	14.3495	14.18751	14.16245	14.00776	14.55966	14.42078	13.79568	14.56919	15.04921	14.23418	14.35811	14.21881	14.81156	13.9048	15.20254	13.4132	13.17949	14.90083	13.85485	14.29456	13.79346	15.57932	12.96214
P12110	Collagen alpl	COL6A2	Collagens	11.98253	11.77314	11.96047	11.92552	11.85354	12.00576	11.57351	12.32108	11.77709	12.06215	12.20105	11.81594	12.45695	12.09311	12.64337	11.15289	11.03848	12.39453	11.84153	12.27617	11.42994	13.142	10.72666
P12111	Collagen alpl	COL6A3	Collagens	13.70789	13.62331	13.7422	13.77977	13.58385	13.72866	13.16775	14.11262	13.48637	13.89306	13.95764	13.62037	13.78267	13.32375	13.78066	12.75873	12.48135	13.55933	13.08387	13.42917	12.90436	14.47857	12.26
P25940	Collagen alpl	COL5A3	Collagens	8.767191	8.576674	8.748528	8.796364	9.275891	8.901169	8.238882	9.23721	9.219411	9.02597	9.243174	9.284477	10.00636	9.632615	10.38811	8.385979	8.122067	10.15784	9.399777	10.18756	8.641655	10.67935	8.547806
P39059	Collagen alpl	COL15A1	Collagens	9.741467	9.702692	9.805422	9.793278	9.566815	9.752882	9.371123	10.59861	9.375908	10.16729	10.27915	9.491052	10.38914	9.922264	10.17308	9.549766	9.174983	9.667324	9.863084	9.7927	9.524263	10.90197	9.065742
P39060	Collagen alpl	COL18A1	Collagens	9.841564	9.833838	9.937963	9.861552	10.04111	9.696794	9.838416	10.62753	10.10119	10.06083	10.17592	10.45368	10.12598	10.29537	10.85929	9.864885	9.571936	10.19729	10.18395	11.34415	9.855349	11.02833	9.493252
Q01955	Collagen alpl	COL4A3	Collagens	6.786761	6.764717	7.099865	6.753095	6.384303	6.617492	6.553619	6.973998	6.192756	6.31951	6.953565	7.044463	6.434567	6.712988	6.158667	6.536853	6.071407	6.860817	6.901375	6.940444	6.774246	6.451614	5.561034
O00468	Agrin	AGRN	ECM Glycoprot	10.10931	10.14134	10.23554	10.25727	10.09262	10.02528	10.19426	11.01862	10.25302	10.11322	10.3661	10.10079	10.28612	10.10733	10.28693	9.192648	9.156358	9.648472	9.629555	9.924097	9.284286	11.05312	8.990991
P07942	Laminin subu	LAMB1	ECM Glycoprot	11.30469	11.42826	11.5813	11.58675	11.52997	11.3374	11.55679	12.21605	12.0191	12.08607	11.92885	11.93553	12.11399	11.89501	12.63433	11.37545	11.09862	12.18761	11.74475	12.3725	11.33249	12.84596	10.92102
P11047	Laminin subu	LAMC1	ECM Glycoprot	11.45805	11.57237	11.67671	11.67996	11.5939	11.42107	11.2999	12.52202	12.31557	12.23649	12.12121	12.17177	12.51336	12.18342	12.87951	11.46112	10.96937	12.52744	11.82936	12.75759	11.28877	13.33152	10.70517
P14543	Nidogen-1	NID1	ECM Glycoprot	10.74054	10.86905	11.00485	11.0045	10.9348	10.76064	10.58787	11.59675	11.01465	11.14536	11.09329	11.16754	11.14856	10.86468	11.30383	10.58594	9.996201	10.59552	10.43025	10.9272	10.21329	11.73721	9.684607
P23142	Fibulin-1	FBLN1	ECM Glycoprot	8.447909	8.656782	8.300581	8.290019	7.796364	8.316507	8.012904	9.050665	7.510962	8.15583	8.337175	7.93428	7.882863	7.945716	8.29325	9.32009	8.230143	8.336259	7.805332	8.295696	8.253572	8.293802	8.133136
P55001	Microfibrillar	MFAP2	ECM Glycoprot	9.247453	9.37417	9.217957	9.292552	9.260449	9.484017	9.161888	8.885696	9.355572	9.02098	9.373953	8.982423	10.30142	9.97394	10.186	8.933309	10.03632	9.769815	9.56223	9.506132	8.990271	11.12589	8.821014
Q14112	Nidogen-2	NID2	ECM Glycoprot	10.69139	10.73352	10.83558	10.77149	10.71063	10.59479	10.88104	11.34085	10.88134	10.96051	10.94464	10.96744	10.96199	10.63405	10.96085	10.36333	10.00123	10.57118	10.26267	10.83189	10.21243	11.42252	9.961463
Q16270	Insulin-like g	IGFBP7	ECM Glycoprot	8.435462	8.332708	8.255029	8.263973	8.15583	8.553822	7.856737	8.459432	8.34074	8.265849	7.783326	7.889352	8.471396	8.451122	9.475071	7.477664	8.210833	8.585782	7.647897	8.874693	7.697689	9.415219	7.987739
Q86XX4	Extracellular	FRAS1	ECM Glycoprot	8.937815	8.912291	8.957393	9.258566	9.044394	8.755221	8.673839	9.483212	8.695924	9.142363	9.214319	8.940754	9.007387	9.013268	8.925219	8.187845	8.800416	8.770533	8.422764	9.143031	8.409709	9.693697	8.228497
P07585	Decorin	DCN	Proteoglycans	11.39687	11.33209	11.3199	11.27851	11.22442	11.69936	10.2724	12.04286	10.93568	10.98499	11.41659	10.90351	11.71654	10.90232	11.80464	10.2275	10.46467	11.60285	10.89112	11.04018	11.05433	12.57165	10.29659
P51884	Lumican	LUM	Proteoglycans	10.25006	10.27857	10.19869	9.992655	10.18871	10.21832	10.18103	10.06864	9.991522	9.304693	9.971256	9.921692	10.77875	10.11387	10.85424	10.85377	9.885734	10.44562	10.19927	10.28807	9.986861	11.1204	9.914937
P98160	Basement m	HSPG2	Proteoglycans	12.92192	13.05168	13.02986	13.12027	13.03571	12.9608	13.06639	13.88452	13.3944	13.53723	13.5046	13.20582	13.52999	13.28104	13.52954	12.60974	12.32956	13.16601	13.03468	13.39519	12.65577	14.33292	12.22423
P07339	Cathepsin D	CTSD	ECM Regulator	9.99407	9.708912	9.562433	9.519047	9.009549	9.855647	9.640787	10.2588	9.212132	10.69818	10.60363	9.571373	9.666528	10.03549	10.17001	9.998632	10.4437	9.925089	10.16183	10.08201	9.765038	10.40067	10.68197
P08311	Cathepsin G	CTSG	ECM Regulator	9.83289	9.952742	9.525129	8.801547	8.946321	9.332037	9.622418	11.16333	9.89224	9.50124	9.807999	8.731659	10.20479	9.586047	9.143016	9.217101	10.34372	9.235339	9.984536	9.167457	9.36653	10.28522	10.20244
P21980	Protein-gluta	TGM2	ECM Regulator	9.918715	9.744666	9.920799	10.10656	10.14249	10.12696	9.164656	10.73649	9.033148	10.56043	10.29393	9.473503	9.606898	10.02085	11.5008	9.202888	9.10087	9.500218	9.502439	9.836327	9.264545	11.17801	8.737343
O00214	Galectin-8	LGALS8	ECM-affiliated	8.817304	8.784307	8.707705	8.812178	8.828772	8.333603	9.285865	8.897543	9.024586	9.080551	8.904785	9.183139	8.395737	8.687535	8.776495	10.15956	8.552892	9.056082	8.572545	9.749404	8.959906	8.739663	8.426984
P49257	Protein ERGI	LMAN1	ECM-affiliated	10.36205	10.19808	10.31288	10.30401	10.62214	10.46087	10.22497	10.71751	10.20701	10.54438	10.66302	10.84815	10.55833	10.47252	10.35892	10.54454	10.73373	10.35804	10.51741	10.38156	10.79608	10.20502	11.15902
P05109	Protein S100	S100A8	Secreted Facto	10.1407	10.94325	10.53732	9.27333	9.073339	9.908843	8.929851	10.32598	11.64282	9.161888	10.61738	8.883559	8.386309	9.097551	9.956774	9.224826	11.17756	9.073224	9.464172	8.61272	9.129922	10.08493	8.848159
P26447	Protein S100	S100A4	Secreted Facto	8.543805	8.591335	8.687201	8.058533	8.167418	8.442944	7.4733	8.972692	7.714245	9.318316	9.243888	7.791163	9.079357	8.809093	10.0473	8.537143	11.23365	9.555605	9.900196	8.882201	8.84586	9.55089	9.707268
P31949	Protein S100	S100A11	Secreted Facto	8.622052	8.941048	9.131343	8.508191	8.846117	8.988116	8.891784	9.044941	8.84549	9.902677	9.002253	8.541871	8.250192	8.576063	9.850294	9.9577	9.031417	9.025183	9.28	9.221382	8.967252	8.834499	8.867304
Q96FQ6	Protein S100	S100A16	Secreted Facto	8.442944	8.576296	8.546509	8.263973	8.310613	8.34474	9.520226	8.559186	9.18933	8.731319	8.321477	8.226893	8.512418	8.233545	9.077865	8.024263	7.529703	8.762388	8.418557	8.986691	8.023147	8.997662	8.140474

Pairwise comparisons between groups

Significantly down-regulated proteins (FC < 0.5, *p < 0.05) shown in green

Significantly up-regulated proteins (FC > 2, *p < 0.05) shown in red

J vs F			Y vs F			O vs F			Y vs J			O vs J			O vs Y		
Gene	log2FoldChange	pvalue	Gene	log2FoldChange	pvalue	Gene	log2FoldChange	pvalue	Gene	log2FoldChange	pvalue	Gene	log2FoldChange	pvalue	Gene	log2FoldChange	pvalue
SERPINH1	-1.250896333	6.41072E-07	COL14A1	-3.509193667	0.000407721	TNXB	-4.0860115	0.000177733	TNXB	-3.4677729	0.00231002	TNXB	-4.3733104	7.82586E-05	SERPINH1	-0.451695167	0.011159446
COL14A1	-1.210799333	3.99426E-06	TNXB	-3.180474	0.00373684	FBN2	-4.0068985	4.38202E-06	FBN2	-2.4734996	0.002366774	FBN2	-3.3241761	1.55327E-06	CELA2A	0.526263333	0.01764417
POSTN	-1.048534933	1.57318E-05	FBN2	-3.156222	0.001107321	COL14A1	-3.8540205	4.03209E-05	COL14A1	-2.298394333	0.003182584	COL14A1	-2.643221167	0.000327328	REG3A	-0.849587333	0.052327024
LGALS2	1.5283838	0.023626069	VWF	-2.523504167	0.001212983	VWF	-2.754229333	5.83896E-05	VWF	-1.939950033	0.003500504	VWF	-2.1706752	9.81539E-05	SERPINB1	-0.348918333	0.054134602
LGALS4	1.117957167	0.023116974	COL12A1	-2.3298025	0.000636053	POSTN	-2.609789	4.5497E-05	COL12A1	-1.5223165	0.003353248	REG3A	-2.069853333	0.005933776	ANXA4	0.27127	0.074456103
FGG	1.2652335	0.022448402	POSTN	-2.280758333	0.000698056	COL12A1	-2.596366833	1.46071E-08	C1QB	-1.515346867	8.21762E-05	CTSC	-1.831894133	9.8431E-05	FGG	0.885680333	0.098543208
ITIH4	1.958497667	0.01499303	CTSC	-1.684777333	0.007954128	CTSC	-2.097933333	0.000100265	ITIH4	-1.50107	0.037405966	C1QB	-1.825464533	1.33626E-05	PLXNB2	-0.406965833	0.105249219
REG3A	1.9664155	0.007593966	SERPINH1	-1.575874667	2.26933E-06	SERPINH1	-2.027569833	3.07809E-09	CTSC	-1.418738133	0.01516056	COL12A1	-1.788880833	3.61149E-06	FGB	0.908483333	0.10537637
FGL1	1.209634367	0.006870246	OGN	-1.568965833	0.000705862	OGN	-1.946706	0.00020043	POSTN	-1.2322234	0.013819496	POSTN	-1.561254067	0.001300263	FGA	0.972227	0.115013347
SERPINA3	1.4164747	0.004708503	PLXNB2	-1.190401833	0.001778758	PLXNB2	-1.597367667	2.05016E-06	REG3A	-1.220266	0.045868008	PLXNB2	-1.491662533	0.000182052	ANXA7	0.194867833	0.127794938
CELA2A	1.588863467	0.004065695	C1QB	-1.145836333	0.000443479	C1QB	-1.455954	4.36497E-05	MFGE8	-1.0965362	0.02052757	ITIH4	-1.415193	0.044631813	CST3	-0.279557167	0.132879196
CELA2B	1.874781667	0.00401879	HCFC1	-1.096303167	0.001269966	HCFC1	-1.322986833	4.94992E-05	LGALS3	-1.091437	0.000417163	VWA1	-1.2336092	0.003634019	ANXA11	-0.233111167	0.144516208
COL4A1	1.159921167	0.002241611	FBN1	-1.011611667	0.005167205	ANXA11	-1.20636	2.84139E-07	PLXNB2	-1.0846967	0.002391198	S100A9	-1.228444333	0.002779617	FBN2	-0.8506765	0.151512926
PRSS1	1.616001333	0.001890536	COL2A1	-1.0040655	0.005395314	S100A9	-1.159896	0.001532446	VWA1	-1.022551867	0.009861143	OGN	-1.163681867	0.006104735	F2	0.143439167	0.186972377
FGB	1.046482667	0.001287729	THBS1	1.2008775	0.039866697	TGFBI	-1.050194167	0.010496099	S100A9	-1.001329667	0.03031578	HCFC1	-1.1573387	7.83987E-05	C1QB	-0.310117667	0.195286635
SERPINB1	1.1878035	0.000464119	COL4A1	1.0812535	0.024311506	BGN	-1.030009667	0.016011814	CSTB	1.259108967	0.010221024	MFGE8	-1.155549533	0.022879327	LGALS1	0.207658833	0.200747267
REG1B	2.6052755	0.000337851	SERPINB6	1.148319	0.012047114	FN1	-1.011095	0.009286154	PRSS2	1.2773318	0.04347187	LGALS3	-1.1034555	0.000554892	ADAM10	-0.3244645	0.214354616
REG1A	2.120918333	2.56334E-05	CSTB	1.269325	0.01058656	LAMC3	1.088917	0.036859176	PRSS3	1.289236267	0.042840329	LAMA4	-1.073475967	0.026869667	FBLN2	0.515921667	0.224313907
ANXA6	-0.822222	4.74097E-05	FBLN2	1.067213	0.009465003	F2	1.097695333	3.05453E-05	THBS1	1.5503687	0.014761692	TGFBI	-1.068538833	0.008195383	VTN	0.324123	0.22789703
COL12A1	-0.807486	0.001571785	REG1B	2.195605167	0.004561291	S100A13	1.1023865	0.006384109	CELA2A	1.6225262	0.00209699	REG1B	-1.034802667	0.009761937	REG1B	-0.625132333	0.247557416
OGN	-0.783024133	0.020592975	CTSA	2.267493667	0.003252542	COL16A1	1.1583725	0.000236938	ASPN	1.660235233	0.000412156	PRELP	1.160001067	0.002319477	S100A10	0.352240667	0.252005876
LGALS1	-0.776176667	0.003321042	REG1A	2.090181667	0.003084585	THBS1	1.164584333	7.72504E-06	CTSA	2.187937933	0.003523615	FBLN2	1.3289743	0.008323868	LGALS2	-0.4160925	0.253113786
CST3	-0.733619767	0.002113433	FGG	1.1092495	0.002314433	COL4A1	1.190001667	0.028400403	FN1	-1.337402333	0.142420968	PRSS3	1.371476267	0.035530116	TNXB	-0.9055375	0.261527274
ANXA5	-0.696123033	0.029550272	PRELP	1.285163167	0.001883189	SERPINA1	1.191901	0.006939054	TGFBI	-0.948362333	0.002185984	PRSS2	1.457978467	0.028404256	COL11A1	0.31081	0.272864434
FBN2	-0.6827224	0.022197397	LGALS2	1.696576167	0.001496277	COL11A1	1.2148195	0.000252788	HCFC1	-0.930655033	0.002551387	THBS1	1.514075533	0.000712661	OGN	-0.377740167	0.291310492
VWF	-0.583554133	0.002242237	CELA3A	1.185605	0.000839078	CELA3A	1.239315	0.000562313	ANXA7	-0.855433233	0.060266723	CSTB	1.533060133	8.09572E-06	F13A1	-0.229254333	0.31154883
ANXA11	-0.530738667	0.00052156	TINAGL1	1.193553333	0.00075321	VTN	1.246050167	0.001559218	OGN	-0.7859417	0.039954756	ASPN	1.672063733	0.000324609	S100A16	0.287372	0.316686414
FBN1	-0.519770333	0.000190785	ITIH5	1.172377	0.000696093	ITIH5	1.247216667	0.000217113	LAMA4	-0.749052133	0.021496391	CTSA	2.1443846	0.000488178	CTSD	0.188479	0.341158274
ANXA1	-0.430171133	0.035152798	SERPINA3	1.578145	0.000632866	LGALS2	1.280483667	0.000739558	EMILIN1	-0.731488667	0.00602858	CELA2A	2.148789533	0.000386133	HCFC1	-0.226683667	0.341251054
ANXA4	-0.381644	0.10143636	FGL1	1.074069333	0.000601357	TINAGL1	1.446221833	0.005236953	COL2A1	-0.705224133	0.063902308	FN1	-1.358182333	0.134623436	CTSB	-0.182114833	0.344349694
THBS1	-0.3494912	0.176920956	LGALS4	1.114393833	0.000525034	CELA3B	1.493667167	0.001478238	S100A8	-0.681153967	0.31753782	S100A8	-0.933395967	0.139876719	COL16A1	0.228689333	0.355732701
ANXA2	-0.325147667	0.232782174	CELA3B	1.282989333	0.000287966	PRELP	1.511672333	5.02844E-07	CTSB	-0.6766349	0.022492994	CTSB	-0.858749733	0.006991017	SERPINI2	0.267683333	0.356431639
COL2A1	-0.298841367	0.304994007	ASPN	2.015951	0.000207579	CSTB	1.543276167	4.90532E-06	HPX	-0.6587538	0.024223981	SERPING1	-0.795204067	0.001620597	REG1A	-0.409518333	0.361023623
LUM	-0.284405967	0.140217577	CELA2B	2.560962	0.000169255	REG1B	1.570472833	0.000523244	PCOLCE	-0.654680767	0.005483292	BGN	-0.7940541	0.042769054	ANXA6	0.114781667	0.369702959
CTSC	-0.2660392	0.03208065	PRSS1	2.475256667	0.000114355	SERPINA3	1.5725765	6.30071E-08	BGN	-0.636995933	0.036121972	SERPINH1	-0.7766735	3.46751E-05	PRELP	0.226509167	0.382397034
DCN	-0.244687667	0.456719249	PRSS3	2.332404333	1.62546E-05	FBLN2	1.583134667	0.002958877	COL5A1	-0.623426333	0.14292448	EMILIN1	-0.753193333	0.008223091	SERPINB6	-0.418962	0.382865511
BGN	-0.235955567	0.16494013	PRSS2	2.404644167	1.2798E-05	REG1A	1.680663333	2.13727E-05	S100A16	-0.596859767	0.080138292	PCOLCE	-0.733260433	0.002156878	CTSC	-0.413156	0.38425106
IGFBP7	-0.1915872	0.245178369	SERPINB1	2.030496167	5.80561E-07	SERPINB1	1.681577833	1.18229E-06	SERPING1	-0.547383233	0.136062734	HPX	-0.725106967	0.014578388	S100A13	0.319331333	0.416672328
HCFC1	-0.165648133	0.112950235	CELA2A	3.211389667	3.54432E-07	FGA	1.773756333	0.012837275	AGRN	-0.534014667	0.0820614	COL5A1	-0.707783833	0.164959997	FGL1	-0.1781795	0.423616396
MFAP2	-0.153148533	0.191796401	FN1	-0.990315	0.021471831	FGB	1.80089	0.00949974	FBN1	-0.491841333	0.075718907	ANXA11	-0.675621333	0.000105218	COL12A1	-0.266564333	0.442948441
IGFBP2	-0.1451087	0.732624063	ANXA11	-0.973248833	0.000278626	FGG	1.994929833	0.00416739	COL6A1	-0.474514333	0.139192242	ANXA7	-0.6605654	0.118457064	SERPINA1	0.337921	0.448359669
COL4A3	-0.135682567	0.491132296	S100A9	-0.932781333	0.033417731	ASPN	2.0277795	2.53014E-09	ANXA2	-0.471507333	0.143756778	AGRN	-0.633957167	0.095546562	POSTN	-0.329030667	0.450781668
COL5A1	-0.130033667	0.452315717	TGFBI	-0.930017667	0.00347487	CTSA	2.223940333	0.000169127	CTSG	-0.459723233	0.27701614	COL2A1	-0.5908013	0.122127363	SERPING1	-0.247820833	0.458296291
PLXNB2	-0.105705133	0.558462775	BGN	-0.8729515	0.008817676	PRSS3	2.414644333	7.45428E-05	ANXA11	-0.442510167	0.016339661	ANXA2	-0.560235667	0.077696041	LAMA4	-0.324423833	0.459448626
FBLN1	-0.087853133	0.76149678	ANXA2	-0.796655	0.005912112	PRSS2	2.585290833	0.000118424	COL6A3	-0.432233	0.16389462	ADAM10	-0.558662433	0.0037535	CSTB	0.273951167	0.460545369
PCOLCE	-0.044860233	0.774635017	ANXA6	-0.766413333	0.000177497	PRSS1	2.596065	3.93262E-07	REG1B	-0.409670333	0.446125535	F13A1	-0.555224333	0.023411264	NID1	-0.246445667	0.495553906
TGM2	-0.0356342	0.926453236	COL5A1	-0.75346	0.081448125	CELA2B	2.6433695	0.00019764	HSPG2	-0.396479667	0.139546833	NID1	-0.489571167	0.173137638	TINAGL1	0.2526685	0.497244073
COL5A2	0.008072667	0.961112797	ANXA7	-0.742258167	0.001526038	CELA2A	3.737653	6.9816E-07	NID2	-0.353527667	0.094346868	NID2	-0.457983833	0.0913067	PLG	-0.085252333	0.523241328
CSTB	0.010216033	0.942005325	PCOLCE	-0.699541	0.001267591	LAMA4	-0.915465667	0.046838417	PLG	-0.340455767	0.030181439	REG1A	-0.440255	0.035759977	CELA3B	0.210677833	0.529280421
TGFBI	0.018344667	0.882681147	MFGE8	-0.6555195	0.105789882	FBN1	-0.914341667	0.048937182	F13A1	-0.32597	0.218072034	COL6A3	-0.437604667	0.24648123	LGALS4	-0.140247667	0.531219881
COL6A3	0.029208	0.875725327	LAMA4	-0.591041833	0.050489981	COL2A1	-0.889642667	0.014086073	SERPINH1	-0.324978333	0.045667812	COL6A1	-0.434921	0.248939383	COL14A1	-0.344826833	0.549143489
FRAS1	0.064318067	0.720368511	EMILIN1	-0.58802	0.016828745	ANXA2	-0.885383333	0.001144506	COL1A2	-0.279540667	0.483656044	PLG	-0.4257081	0.007877139	S100A9	-0.227114667	0.56643106
S100A9	0.068548333	0.805699453	CST3	-0.562089667	0.019194403	CST3	-0.841646833	0.000830338	COL1A1	-0.265712	0.563324807	FBN1	-0.394571333	0.319331317	DCN	0.23959	0.572761623
COL6A2	0.070149333	0.644506992	S100A8	-0.525033833	0.319387424	COL5A1	-0.8378175	0.10596185	NID1	-0.2431255	0.367900716	HSPG2	-0.342628	0.321900368	COL5A3	0.299063333	0.573034295
CTSA	0.079555733	0.810701925	ANXA1	-0.473727833	0.009924394	PCOLCE	-0.778120667	2.34704E-05	ADAM10	-0.234197933	0.338720932	FGL1	-0.313744533	0.307274172	VWA1	-0.211057333	0.580794782
LMAN1	0.094706333	0.471092033	ANXA4	-0.420711667	0.004283106	S100A8	-0.777275833	0.060087057	FRAS1	-0.2294499	0.289497967	S100A16	-0.309487767	0.296805676	S100A8	-0.252242	0.59533264
COL1A1	0.099872	0.700360936	LGALS1	-0.4114435	0.05347612	MFGE8	-0.714532833	0.104797615	COL5A2	-0.199798333	0.492598122	CTSG	-0.290525067	0.449734047	PCOLCE	-0.078579667	0.597227307
COL6A1	0.108272667	0.503921622	COL6A3	-0.403025	0.135135884	ANXA6	-0.651631667	4.56905E-05	COL11A1	-0.1767365	0.79883093	TGM2	-0.287917133	0.572727107	PRSS2	0.180646667	0.601878401
ANXA7	0.113175067	0.764996305	DCN	-0.371995	0.221726378	EMILIN1	-0.609724667	0.020974616	COL15A1	-0.175035367	0.589845263	FRAS1	-0.2638929	0.363903995	LAMC3	0.235161667	0.611580012
COL1A2	0.119997333	0.609494429	COL6A1	-0.366241667	0.206299096	ANXA7	-0.547390333	0.008161792	LAMA5	-0.164905667	0.489175541	LGALS2	-0.247900133	0.621153646	CTSG	0.169198167	0.621094429
S100A4	0.1292822	0.764931106	AGRN	-0.288530667	0.245352917	ANXA1	-0.483120667	0.005096913	FGG	-0.155984	0.723322541	COL1A1	-0.202417	0.709636115	VWF	-0.230725167	0.629385763
COL3A1	0.131066	0.620351666	VWA1	-0.254996667	0.3769512	VWA1	-0.466054	0.134802913	FGB	-0.154076	0.534796334	COL1A2	-0.170414	0.70107295	ANXA3	-0.135911667	0.649079478
EMILIN1	0.143468667	0.082338096	LGALS3	-0.251877333	0.164771172	COL6A3	-0.408396667	0.236692777	TGM2	-0.140095967	0.789493532	COL15A1	-0.155902367	0.66820906	COL18A1	0.155196667	0.660838441
COL5A3	0.148626567	0.516405204	COL4A3	-0.241214667	0.203471815	AGRN	-0.388473167	0.2413887	COL3A1	-0.137546	0.74930592	LGALS4	-0.143811	0.69526423	BGN	-0.157058167	0.673998565
S100A8	0.156120133	0.795099746	CTSB	-0.200818667	0.2290858	CTSB	-0.3829335	0.022976753	FGL1	-0.135565033	0.644374617	LAMA5	-0.121415667	0.700242097	PRSS1	0.120808333	0.681804735
LAMA4	0.1580103	0.129791973	COL5A2	-0.191725667	0.482003412	ADAM10	-0.347197833	0.02866703	DCN	-0.127307333	0.753589752	LGALS8	-0.121235333	0.574043626	ANXA2	-0.088728333	0.700123879
ANXA3	0.161540533	0.396228965	TGM2	-0.175730167	0.651641782	COL6A1	-0.326648333	0.34994734	COL6A2	-0.120186	0.70365956	COL5A2	-0.1162625	0.736711592	COL3A1	0.225931667	0.701669196
NID1	0.201854	0.285948323	COL1A1	-0.16584	0.685765217	TGM2	-0.323551333	0.385320737	LAMB1	-0.119217	0.645863388	CST3	-0.108027067	0.343624363	NID2	-0.104456167	0.707204488
ADAM10	0.2114646	0.06514268	FRAS1	-0.165131833	0.291591648	NID1	-0.287717167	0.360988833	COL4A3	-0.1055321	0.639038537	S100A11	-0.104825667	0.63585982	TGFBI	-0.1201765	0.717717806
S100A13	0.2255116	0.158255367	COL1A2	-0.159543333	0.648064668	LGALS3	-0.263895833	0.166136607	S100A10	-0.104883	0.754188151	LAMB1	-0.060650333	0.850808521	FBLN1	-0.081424167	0.741279598
COL15A1	0.231323533	0.407950787	S100A16	-0.146731333	0.524759141	LGALS1	-0.203784667	0.305926938	S100A11	-0.1028395	0.777462128	ANXA1	-0.052949533	0.732357659	COL2A1	0.114422833	0.743030335
AGRN	0.245484	0.208505134	NID2	-0.074751667	0.670717564	FRAS1	-0.199574833	0.431203968	CTSD	-0.1017503	0.75804517	FBLN1	-0.0272077	0.921811298	COL4A3	0.088430833	0.743189962
SERPINB6	0.253146667	0.06916732	COL6A2	-0.050036667	0.86165793	NID2	-0.179207833	0.445470514	LGALS8	-0.079439667	0.781151303	LAMC1	-0.025729667	0.956109476	ITIH5	0.074839667	0.763856881
FBLN2	0.254160367	0.385042125	NID1	-0.0412715	0.847696424	COL4A3	-0.152783833	0.539259796	COL4A1	-0.078667667	0.84836039	COL6A2	-0.018504333	0.961426857	TGM2	-0.147821167	0.770931168
NID2	0.278776	0.028327997	FBLN1	-0.033636667	0.896508325	ANXA4	-0.149441667	0.22145015	LAMC1	-0.069279667	0.849608331	COL4A3	-0.017101267	0.950757131	S100A4	0.157385667	0.775694066
TNXB	0.2872989	0.062988045	ADAM10	-0.022733333	0.918513299	F13A1	-0.1434645	0.274633801	ANXA1	-0.0435567	0.800025989	LAMA2	0.007773333	0.984577569	LUM	-0.075483	0.780144175
COL18A1	0.2919737	0.087036211	PLG	-0.018834667	0.860665106	DCN	-0.132405	0.697581729	ANXA4	-0.039067667	0.853364829	COL4A1	0.0300805	0.947562295	AGRN	-0.0999425	0.789535481
S100A11	0.297951167	0.219563585	COL3A1	-0.00648	0.986185155	FBLN1	-0.115060833	0.437571561	REG1A	-0.030736667	0.943084351	CTSD	0.0867287	0.79341674	ITIH4	0.085877	0.798536914
SERPINI2	0.312269467	0.207904141	IGFBP7	-0.003564333	0.99035856	COL5A2	-0.108189833	0.74320866	LGALS4	-0.003563333	0.992264908	COL3A1	0.088385667	0.867290073	MFAP2	-0.106343333	0.804371457
PLG	0.3216211	0.01988724	COL15A1	0.056288167	0.778468454	PLG	-0.104087	0.297939192	COL18A1	0.034412967	0.882172982	LMAN1	0.098143667	0.597762475	PRSS3	0.08224	0.806877097
LGALS8	0.324687833	0.01325476	HSPG2	0.060908333	0.778644107	REG3A	-0.103437833	0.191200791	LAMA2	0.052113333	0.873342819	LAMB2	0.103163333	0.673864937	COL6A2	0.101681667	0.820534088
FN1	0.347087333	0.654656287	HPX	0.066935833	0.762002738	COL1A1	-0.102545	0.838483013	FBLN1	0.054216467	0.87493518	DCN	0.112282667	0.800033321	FBN1	0.09727	0.820787406
PRELP	0.351671267	0.155527313	F13A1	0.085789833	0.677747282	COL1A2	-0.050416667	0.900115258	ANXA6	0.055808667	0.680793785	COL11A1	0.1340735	0.843536723	HPX	-0.066353167	0.824634424
ASPN	0.355715767	0.125972602	CTSG	0.139277833	0.67396735	HPX	0.000582667	0.997835964	LAMB2	0.094835	0.678711725	SERPINA3	0.1561018	0.588480959	COL4A2	0.0811	0.828636861
C1QB	0.369510533	0.06989278	S100A11	0.195111667	0.55058852	IGFBP7	0.035365833	0.908976516	COL4A2	0.106831667	0.700495989	ANXA6	0.170590333	0.140465561	COL1A2	0.109126667	0.829519469
F13A1	0.411759833	0.06239459	LMAN1	0.20936	0.051363059	SERPING1	0.0497775	0.66607867	LMAN1	0.114653667	0.40628748	COL4A2	0.187931667	0.588460563	CELA2B	0.0824075	0.833666372
MFGE8	0.4410167	0.055006249	LUM	0.2135085	0.32737464	COL6A2	0.051645	0.886254017	SERPINA3	0.1616703	0.646019925	COL18A1	0.189609633	0.566802817	COL5A2	0.083535833	0.834033814

CELA3A	0.447291	0.061317222		LGALS8	0.245248167	0.407073718		COL15A1	0.075421167	0.775373916		LGALS2	0.168192367	0.754176887		IGFBP7	0.226953033	0.500411516		COL4A1	0.108748167	0.839493721
S100A16	0.450128433	0.106520522		ANXA5	0.254668667	0.477227643		HSPG2	0.11476	0.712298284		CST3	0.1715301	0.318033536		ANXA4	0.232202333	0.300407868		CELA3A	0.05371	0.849741362
HSPG2	0.457388	0.023714553		SERPING1	0.297598333	0.361437223		LUM	0.1380255	0.480154295		SERPINI2	0.173409367	0.442898206		CRELD2	0.241424133	0.317787139		LAMB1	0.058566667	0.875803035
CTSD	0.474429467	0.187712611		COL18A1	0.326386667	0.140236084		S100A16	0.140640667	0.466766884		IGFBP7	0.188022867	0.559253021		S100A10	0.247357667	0.475268499		COL5A1	-0.0843575	0.880723513
CTSB	0.475816233	0.066746941		CTSD	0.372679167	0.08135498		LMAN1	0.19285	0.255838219		SERPINA1	0.2131452	0.595264435		LUM	0.422431467	0.106090444		HSPG2	0.053851667	0.883371306
VTN	0.477270967	0.043587382		LAMB1	0.38076	0.150339188		S100A11	0.1931255	0.144619468		LAMC3	0.2397704	0.36223167		SERPINI2	0.4410927	0.176251859		LAMA5	0.04349	0.895589742
LAMB1	0.499977	0.007430144		MFAP2	0.422802333	0.153533449		LGALS8	0.2034525	0.360598186		CRELD2	0.257973467	0.197654796		F2	0.4590548	0.035850308		FRAS1	-0.034443	0.898042448
CRELD2	0.505935867	0.011785878		ITIH4	0.457427667	0.205662244		COL3A1	0.219451667	0.655021524		FGA	0.268947267	0.514472473		MFAP2	0.469607533	0.226216099		IGFBP2	0.043046333	0.899613527
COL16A1	0.5200194	0.301166901		COL5A3	0.458965167	0.273085074		ANXA5	0.229472167	0.517969		CELA3B	0.291622333	0.199601914		LAMC3	0.474932067	0.285336638		LGALS8	-0.041795667	0.90071142
LAMC1	0.532028	0.063369865		LAMC1	0.462748333	0.167602992		CTSG	0.308476	0.297708202		COL5A3	0.3103386	0.476082045		SERPINB6	0.476210333	0.225250039		MFGE8	-0.059013333	0.906659021
FGA	0.532582067	0.149432284		SERPINI2	0.485678833	0.030286052		MFAP2	0.316459	0.388564267		F2	0.315615633	0.115719821		SERPINB1	0.493774333	0.026891478		COL1A1	0.063295	0.918574588
CTSG	0.599001067	0.135075879		S100A10	0.638412833	0.022322663		LAMB1	0.439326667	0.19119639		LGALS1	0.364733167	0.016038762		CELA3B	0.502300167	0.123704997		LMAN1	-0.01651	0.920755863
LAMB2	0.60888	0.00281726		IGFBP2	0.684480833	0.029037939		COL18A1	0.481583333	0.156125358		COL16A1	0.409663767	0.414822147		ITIH5	0.506043167	0.028494801		LAMA2	-0.04434	0.92317117
LAMC3	0.613984933	0.002674724		LAMB2	0.703715	0.012516683		LAMC1	0.506298333	0.262856598		ITIH5	0.4312035	0.07253883		SERPINA1	0.5510662	0.150759715		COL6A1	0.039593333	0.923604333
F2	0.638640533	0.010073427		LAMA5	0.723893333	0.009000847		ITIH4	0.543304667	0.009389634		VTN	0.4446562	0.067417267		LGALS1	0.572392	0.00268439		IGFBP7	0.038930167	0.924348286
SERPINA1	0.6408348	0.034690589		REG3A	0.7461495	0.077115162		CTSD	0.561158167	0.016294765		LUM	0.497914467	0.075932628		ANXA3	0.595617133	0.024127498		LAMC1	0.04355	0.931348022
TINAGL1	0.663909633	0.023477653		CRELD2	0.763909333	0.00131557		LAMB2	0.712043333	0.017275023		TINAGL1	0.5296437	0.070953905		COL5A3	0.609401933	0.176363594		EMILIN1	-0.021704667	0.93276572
HPX	0.725689633	0.001315195		S100A13	0.783055167	0.036972902		IGFBP2	0.727527167	0.036156984		S100A13	0.557543567	0.11182752		COL16A1	0.6383531	0.220190586		CTSA	-0.043553333	0.937511283
ITIH5	0.7411735	0.000430543		FGA	0.801529333	0.028973459		SERPINB6	0.729357	0.086327263		MFAP2	0.575950867	0.072307971		FGG	0.729696333	0.218976731		THBS1	-0.036293167	0.938341623
S100A10	0.743295833	0.041456422		S100A4	0.834411667	0.156866262		CRELD2	0.74736	0.008934083		CELA2B	0.686180333	0.133984462		FGB	0.754407333	0.165888253		ANXA1	-0.009392833	0.940286169
VWA1	0.7675552	0.005465041		LAMC3	0.853755333	0.010706092		SERPINI2	0.753362167	0.028374462		S100A4	0.705129467	0.295908186		CELA2B	0.768587833	0.104884178		ANXA5	-0.0251965	0.945062406
COL4A2	0.795707167	0.005619008		SERPINA1	0.85398	0.045491342		ANXA3	0.757157667	0.003701947		ANXA3	0.7315288	0.031429218		VTN	0.7687792	0.020223194		CRELD2	-0.016549333	0.945157023
LGALS3	0.839559667	1.94381E-06		FGB	0.892406667	0.001754108		COL5A3	0.7580285	0.090512456		CELA3A	0.7383314	0.019537723		TINAGL1	0.7823122	0.063295514		COL15A1	0.019133	0.952229462
SERPING1	0.844981567	0.002565513		ANXA3	0.893069333	0.011507778		LAMA5	0.767383333	0.034793892		FBLN2	0.813052633	0.05581608		CELA3A	0.792024	0.013275545		LGALS3	-0.0120185	0.957105049
LAMA5	0.888799	0.002496794		COL4A2	0.902538833	0.008023507		FGL1	0.895889833	0.001901208		IGFBP2	0.829589533	0.100585372		S100A4	0.862515133	0.096629554		FN1	-0.02078	0.958746654
LAMA2	0.930675	0.002967109		COL11A1	0.9040095	0.00689854		LAMA2	0.938448333	0.045107682		SERPINB1	0.842692667	0.001846082		IGFBP2	0.872635867	0.093284192		ASPN	0.0118285	0.963760722
CELA3B	0.991367	7.9455E-05		VTN	0.921927167	0.000261356		LGALS4	0.974146167	0.00078741		PRSS1	0.859255333	0.035769195		S100A13	0.8768749	0.016557641		LAMB2	0.008328333	0.976977105
COL11A1	1.080746	0.157222404		COL16A1	0.929683167	0.002078809		COL4A2	0.983638833	0.019465105		SERPINB6	0.895172333	0.032582803		ANXA5	0.9255952	0.012630063		SERPINA3	-0.0055685	0.981700312
PRSS3	1.043168067	0.080628887		F2	0.954256167	5.91495E-05		S100A10	0.9906535	0.004219982		PRELP	0.9334919	0.012714561		PRSS1	0.980063667	0.010860338		COL6A3	-0.005371667	0.98896386
PRSS2	1.127312367	0.064982548		LAMA2	0.982788333	0.015129335		S100A4	0.991797333	0.001385248		ANXA5	0.9507917	0.011930068		FGA	1.241174267	0.057951719		S100A11	-0.001986167	0.994994669

Table S6. List of significantly changed (two-sided *t* test, $p < 0.05$) cellular components, molecular functions, biological processes and transcription factor targets in J vs F and Y vs J in GSVA analysis.

See **Supplementary Data 6** at **https://www.nature.com/articles/s41467-021-21261-w.#Sec18**

Table S7. List of core ECM proteins sorted by abundance in each developmental group. Matrisome categories are highlighted in different colors. Significantly changed proteins among age groups (one-way ANOVA, FDR 0.05) are highlighted in orange.

Quantified: Core ECM Proteins sorted by abundance in each developmental group																		
	Significantly changed among age groups (one-way ANOVA, FDR 0.05)																	
Rank	Type	Category	Gene name	Avg F	Type	Category	Gene name	Avg J	Type	Category	Gene name	Avg Y	Type	Category	Gene name	Avg O		
1	Core matrisome	Collagens	COL1A1	14.57494	Core matrisome	Collagens	COL1A1	14.67481	Core matrisome	Collagens	COL1A1	14.4091	Core matrisome	Collagens	COL3A1	14.62309		
2	Core matrisome	Collagens	COL3A1	14.40364	Core matrisome	Collagens	COL3A1	14.53471	Core matrisome	Collagens	COL3A1	14.39716	Core matrisome	Collagens	COL1A1	14.4724		
3	Core matrisome	Collagens	COL1A2	14.28128	Core matrisome	Collagens	COL1A2	14.40127	Core matrisome	Collagens	COL1A2	14.12173	Core matrisome	Collagens	COL1A2	14.23086		
4	Core matrisome	Collagens	COL6A3	13.69428	Core matrisome	Collagens	COL6A3	13.72349	Core matrisome	Collagens	COL6A3	13.29126	Core matrisome	Collagens	COL6A3	13.28588		
5	Core matrisome	ECM Glycoproteins	FBN1	13.35853	Core matrisome	Proteoglycans	HSPG2	13.47743	Core matrisome	Proteoglycans	HSPG2	13.08095	Core matrisome	Proteoglycans	HSPG2	13.1348		
6	Core matrisome	Proteoglycans	HSPG2	13.02004	Core matrisome	ECM Glycoproteins	FBN1	12.83876	Core matrisome	ECM Glycoproteins	LAMA2	12.65464	Core matrisome	ECM Glycoproteins	LAMA2	12.6103		
7	Core matrisome	Collagens	COL6A1	12.33275	Core matrisome	ECM Glycoproteins	LAMA2	12.60253	Core matrisome	ECM Glycoproteins	FBN1	12.34692	Core matrisome	ECM Glycoproteins	FBN1	12.44419		
8	Core matrisome	Collagens	COL14A1	12.32879	Core matrisome	ECM Glycoproteins	FN1	12.57304	Core matrisome	ECM Glycoproteins	LAMA5	12.22686	Core matrisome	ECM Glycoproteins	LAMA5	12.27035		
9	Core matrisome	ECM Glycoproteins	FN1	12.22596	Core matrisome	Collagens	COL6A1	12.44103	Core matrisome	ECM Glycoproteins	LAMC1	12.02976	Core matrisome	ECM Glycoproteins	LAMC1	12.07331		
10	Core matrisome	Collagens	COL6A2	11.91683	Core matrisome	ECM Glycoproteins	LAMA5	12.39176	Core matrisome	Collagens	COL6A1	11.96651	Core matrisome	Collagens	COL6A1	12.00611		
11	Core matrisome	Collagens	COL12A1	11.79177	Core matrisome	ECM Glycoproteins	LAMC1	12.09904	Core matrisome	Collagens	COL6A2	11.86679	Core matrisome	Collagens	COL6A2	11.96847		
12	Core matrisome	ECM Glycoproteins	LAMA2	11.67186	Core matrisome	Collagens	COL6A2	11.98698	Core matrisome	ECM Glycoproteins	LAMB1	11.84216	Core matrisome	ECM Glycoproteins	LAMB1	11.90072		
13	Core matrisome	ECM Glycoproteins	LAMC1	11.56701	Core matrisome	ECM Glycoproteins	LAMB1	11.96137	Core matrisome	ECM Glycoproteins	LAMB2	11.24054	Core matrisome	ECM Glycoproteins	FGG	11.79191		
14	Core matrisome	ECM Glycoproteins	LAMA5	11.50297	Core matrisome	ECM Glycoproteins	EMILIN1	11.25173	Core matrisome	ECM Glycoproteins	FN1	11.23564	Core matrisome	ECM Glycoproteins	FGB	11.75142		
15	Core matrisome	ECM Glycoproteins	LAMB1	11.4614	Core matrisome	ECM Glycoproteins	LAMB2	11.1457	Core matrisome	Collagens	COL4A1	11.0483	Core matrisome	ECM Glycoproteins	FGA	11.29886		
16	Core matrisome	Proteoglycans	DCN	11.37519	Core matrisome	Proteoglycans	DCN	11.1305	Core matrisome	Proteoglycans	DCN	11.0032	Core matrisome	ECM Glycoproteins	LAMB2	11.24886		
17	Core matrisome	ECM Glycoproteins	EMILIN1	11.10826	Core matrisome	Collagens	COL4A1	11.12696	Core matrisome	ECM Glycoproteins	FGG	10.90623	Core matrisome	Proteoglycans	DCN	11.24279		
18	Core matrisome	ECM Glycoproteins	POSTN	11.08167	Core matrisome	Collagens	COL14A1	11.11799	Core matrisome	Collagens	COL4A2	10.87713	Core matrisome	ECM Glycoproteins	FN1	11.21486		
19	Core matrisome	Collagens	COL5A1	11.04442	Core matrisome	ECM Glycoproteins	NID1	11.08758	Core matrisome	ECM Glycoproteins	NID1	10.84446	Core matrisome	Collagens	COL4A1	11.15704		
20	Core matrisome	ECM Glycoproteins	NID1	10.88573	Core matrisome	ECM Glycoproteins	FGG	11.06221	Core matrisome	ECM Glycoproteins	FGB	10.84294	Core matrisome	Collagens	COL4A2	10.95823		
21	Core matrisome	ECM Glycoproteins	NID2	10.7229	Core matrisome	ECM Glycoproteins	NID2	11.00168	Core matrisome	ECM Glycoproteins	NID2	10.64815	Core matrisome	Collagens	COL5A2	10.59961		
22	Core matrisome	Collagens	COL5A2	10.7078	Core matrisome	ECM Glycoproteins	FGB	10.99702	Core matrisome	ECM Glycoproteins	EMILIN1	10.52024	Core matrisome	ECM Glycoproteins	NID1	10.59801		
23	Core matrisome	ECM Glycoproteins	FBN2	10.58618	Core matrisome	Collagens	COL12A1	10.98428	Core matrisome	Collagens	COL5A2	10.51608	Core matrisome	ECM Glycoproteins	NID2	10.54369		
24	Core matrisome	ECM Glycoproteins	LAMB2	10.53682	Core matrisome	Collagens	COL5A1	10.91439	Core matrisome	Proteoglycans	LUM	10.40134	Core matrisome	ECM Glycoproteins	EMILIN1	10.49854		
25	Core matrisome	Proteoglycans	OGN	10.42394	Core matrisome	Collagens	COL4A2	10.7703	Core matrisome	ECM Glycoproteins	FGA	10.32663	Core matrisome	Collagens	COL18A1	10.35039		
26	Core matrisome	ECM Glycoproteins	TGFBI	10.38071	Core matrisome	Collagens	COL5A2	10.71588	Core matrisome	Collagens	COL5A1	10.29096	Core matrisome	Proteoglycans	LUM	10.32586		
27	Core matrisome	ECM Glycoproteins	VWF	10.29806	Core matrisome	ECM Glycoproteins	TGFBI	10.39906	Core matrisome	Collagens	COL18A1	10.19519	Core matrisome	Proteoglycans	PRELP	10.23216		
28	Core matrisome	Proteoglycans	LUM	10.18783	Core matrisome	ECM Glycoproteins	AGRN	10.38904	Core matrisome	Proteoglycans	ASPN	10.06182	Core matrisome	Collagens	COL5A1	10.2066		
29	Core matrisome	ECM Glycoproteins	AGRN	10.14356	Core matrisome	Collagens	COL18A1	10.16078	Core matrisome	Proteoglycans	PRELP	10.00566	Core matrisome	Proteoglycans	ASPN	10.07365		
30	Core matrisome	Collagens	COL4A2	9.974593	Core matrisome	ECM Glycoproteins	FGA	10.05768	Core matrisome	ECM Glycoproteins	AGRN	9.855029	Core matrisome	ECM Glycoproteins	LAMC3	9.999304		
31	Core matrisome	Collagens	COL4A1	9.967043	Core matrisome	ECM Glycoproteins	POSTN	10.03313	Core matrisome	Collagens	COL15A1	9.783381	Core matrisome	Collagens	COL15A1	9.802514		
32	Core matrisome	ECM Glycoproteins	FGB	9.950533	Core matrisome	Collagens	COL15A1	9.958416	Core matrisome	ECM Glycoproteins	LAMC3	9.764142	Core matrisome	ECM Glycoproteins	AGRN	9.755087		
33	Core matrisome	Collagens	COL18A1	9.868804	Core matrisome	ECM Glycoproteins	FBN2	9.903453	Core matrisome	ECM Glycoproteins	MFAP2	9.735569	Core matrisome	ECM Glycoproteins	TINAGL1	9.676483		
34	Core matrisome	ECM Glycoproteins	FGG	9.796979	Core matrisome	Proteoglycans	LUM	9.903428	Core matrisome	Collagens	COL12A1	9.461968	Core matrisome	ECM Glycoproteins	MFAP2	9.629225		
35	Core matrisome	Collagens	COL15A1	9.727093	Core matrisome	ECM Glycoproteins	VWA1	9.800092	Core matrisome	ECM Glycoproteins	TGFBI	9.450696	Core matrisome	Collagens	COL5A3	9.602331		
36	Core matrisome	Proteoglycans	BGN	9.694257	Core matrisome	ECM Glycoproteins	VWF	9.714504	Core matrisome	ECM Glycoproteins	TINAGL1	9.423815	Core matrisome	ECM Glycoproteins	VTN	9.521811		
37	Core matrisome	ECM Glycoproteins	FGA	9.5251	Core matrisome	Proteoglycans	OGN	9.640914	Core matrisome	Collagens	COL5A3	9.303268	Core matrisome	ECM Glycoproteins	TGFBI	9.330519		
38	Core matrisome	Collagens	COL2A1	9.452334	Core matrisome	ECM Glycoproteins	LAMC3	9.524372	Core matrisome	ECM Glycoproteins	VTN	9.197688	Core matrisome	Collagens	COL12A1	9.195403		
39	Core matrisome	ECM Glycoproteins	LAMA4	9.327761	Core matrisome	ECM Glycoproteins	LAMA4	9.485771	Core matrisome	ECM Glycoproteins	CRELD2	8.929236	Core matrisome	ECM Glycoproteins	CRELD2	8.912686		
40	Core matrisome	ECM Glycoproteins	MFAP2	9.312766	Core matrisome	Proteoglycans	BGN	9.458302	Core matrisome	Proteoglycans	OGN	8.854973	Core matrisome	ECM Glycoproteins	FRAS1	8.778039		
41	Core matrisome	ECM Glycoproteins	PCOLCE	9.067155	Core matrisome	ECM Glycoproteins	MFAP2	9.159618	Core matrisome	Proteoglycans	BGN	8.821306	Core matrisome	Collagens	COL16A1	8.753229		
42	Core matrisome	ECM Glycoproteins	VWA1	9.032537	Core matrisome	Collagens	COL2A1	9.153493	Core matrisome	Collagens	COL14A1	8.8196	Core matrisome	ECM Glycoproteins	THBS1	8.742463		
43	Core matrisome	ECM Glycoproteins	FRAS1	8.977613	Core matrisome	Proteoglycans	PRELP	9.072164	Core matrisome	ECM Glycoproteins	FRAS1	8.812482	Core matrisome	Proteoglycans	BGN	8.664248		
44	Core matrisome	ECM Glycoproteins	LAMC3	8.910387	Core matrisome	ECM Glycoproteins	FRAS1	9.041931	Core matrisome	ECM Glycoproteins	POSTN	8.80091	Core matrisome	Collagens	COL11A1	8.594299		
45	Core matrisome	Collagens	COL5A3	8.844303	Core matrisome	ECM Glycoproteins	PCOLCE	9.022295	Core matrisome	ECM Glycoproteins	THBS1	8.778757	Core matrisome	ECM Glycoproteins	IGFBP2	8.581061		
46	Core matrisome	Proteoglycans	PRELP	8.720492	Core matrisome	ECM Glycoproteins	TNXB	9.004191	Core matrisome	ECM Glycoproteins	VWA1	8.77754	Core matrisome	ECM Glycoproteins	VWA1	8.566483		
47	Core matrisome	ECM Glycoproteins	TNXB	8.716893	Core matrisome	Collagens	COL5A3	8.992929	Core matrisome	ECM Glycoproteins	LAMA4	8.736719	Core matrisome	Collagens	COL2A1	8.562692		
48	Core matrisome	ECM Glycoproteins	IGFBP7	8.332804	Core matrisome	ECM Glycoproteins	TINAGL1	8.894171	Core matrisome	ECM Glycoproteins	IGFBP2	8.538014	Core matrisome	Proteoglycans	OGN	8.477232		
49	Core matrisome	ECM Glycoproteins	FBLN1	8.30136	Core matrisome	ECM Glycoproteins	VTN	8.753032	Core matrisome	Collagens	COL16A1	8.524539	Core matrisome	Collagens	COL14A1	8.474773		
50	Core matrisome	ECM Glycoproteins	VTN	8.275761	Core matrisome	ECM Glycoproteins	CRELD2	8.671262	Core matrisome	Collagens	COL2A1	8.448269	Core matrisome	ECM Glycoproteins	POSTN	8.471879		
51	Core matrisome	ECM Glycoproteins	TINAGL1	8.230261	Core matrisome	ECM Glycoproteins	FGL1	8.581857	Core matrisome	ECM Glycoproteins	FGL1	8.446292	Core matrisome	ECM Glycoproteins	LAMA4	8.412295		
52	Core matrisome	ECM Glycoproteins	CRELD2	8.165326	Core matrisome	Collagens	COL11A1	8.460225	Core matrisome	ECM Glycoproteins	PCOLCE	8.367614	Core matrisome	ECM Glycoproteins	FBLN2	8.404777		
53	Core matrisome	Proteoglycans	ASPN	8.045866	Core matrisome	Proteoglycans	ASPN	8.401582	Core matrisome	ECM Glycoproteins	IGFBP7	8.32924	Core matrisome	ECM Glycoproteins	IGFBP7	8.36817		
54	Core matrisome	ECM Glycoproteins	MFGE8	7.954675	Core matrisome	ECM Glycoproteins	MFGE8	8.395691	Core matrisome	Collagens	COL11A1	8.283489	Core matrisome	ECM Glycoproteins	PCOLCE	8.289034		
55	Core matrisome	ECM Glycoproteins	IGFBP2	7.853534	Core matrisome	ECM Glycoproteins	FBLN1	8.213507	Core matrisome	ECM Glycoproteins	FBLN1	8.267724	Core matrisome	ECM Glycoproteins	FGL1	8.268113		
56	Core matrisome	Collagens	COL16A1	7.594856	Core matrisome	ECM Glycoproteins	IGFBP7	8.141217	Core matrisome	ECM Glycoproteins	FBLN2	7.888855	Core matrisome	ECM Glycoproteins	FBLN1	8.1863		
57	Core matrisome	ECM Glycoproteins	THBS1	7.577879	Core matrisome	Collagens	COL16A1	8.114875	Core matrisome	ECM Glycoproteins	VWF	7.774554	Core matrisome	ECM Glycoproteins	VWF	7.543829		
58	Core matrisome	Collagens	COL11A1	7.379479	Core matrisome	ECM Glycoproteins	IGFBP2	7.708425	Core matrisome	ECM Glycoproteins	FBN2	7.429953	Core matrisome	ECM Glycoproteins	MFGE8	7.240142		
59	Core matrisome	ECM Glycoproteins	FGL1	7.372223	Core matrisome	ECM Glycoproteins	THBS1	7.228388	Core matrisome	ECM Glycoproteins	MFGE8	7.299155	Core matrisome	Collagens	COL4A3	6.581588		
60	Core matrisome	ECM Glycoproteins	FBLN2	6.821642	Core matrisome	ECM Glycoproteins	FBLN2	7.075802	Core matrisome	Collagens	COL4A3	6.493158	Core matrisome	ECM Glycoproteins	FBN2	6.579277		
61	Core matrisome	Collagens	COL4A3	6.734372	Core matrisome	Collagens	COL4A3	6.59869	Core matrisome	ECM Glycoproteins	TNXB	5.536419	Core matrisome	ECM Glycoproteins	TNXB	4.630881		

Table S8. Antibody information used for IF staining.

Protein	Blocking Blocking buffer; dilution; supplier (catalogue #)	ECM 1° antibody; host, dilution, supplier (catalogue #)	ECM 2° antibody; host, dilution, supplier (catalogue #)	Insulin 1° antibody; host, dilution, supplier (catalogue #)	Insulin 2° antibody; host, dilution, supplier (catalogue #)	Antibody validation information
CD31	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	CD31 monoclonal; Raised in mouse; 1:50; Santa Cruz (sc-376764)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	Insulin polyclonal; Raised in guinea pig; 1:2000; Sigma-Aldrich (I8510)	Goat anti-guinea pig A488; 1:800; Life Technologies (A11073)	[#sc-376764] Specific for an epitope mapping between amino acids 699-727 at the C-terminus of PECAM-1 of mouse origin. Selected Validation Data available on supplier website and in previous publications.
Collagen I	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen I polyclonal; Raised in rabbit; 1:500; Abcam (ab34710)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#ab34710] Specificity Statement: This antibody reacts with most mammalian Type I collagens and has negligible cross-reactivity with Type II, III, IV, V or VI collagens. Non-specific cross-reaction of anti-collagen antibodies with other human serum proteins or non-collagen extracellular matrix proteins is negligible.
Collagen II	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen II polyclonal; Raised in rabbit; 1:200; Abcam (ab34712)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#ab34712] Specificity Statement: This antibody reacts with Type II collagens and shows negligible (less than 1%) cross reactivity with Type I, III, IV, V or VI collagens. Nonspecific cross reaction with other human serum proteins or non collagen extracellular matrix proteins is negligible.
Collagen III	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen III polyclonal; Raised in rabbit; 1:500; Abcam (ab7778)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#ab7778] Specificity Statement: This type specific collagen antibody only recognizes 3D epitopes. Negligible cross-reactivity with Type I, II, IV, V or VI collagens. Non-specific cross reaction of anti-collagen antibodies with other human serum proteins or non-collagen extracellular matrix proteins is negligible
Collagen IV	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen IV polyclonal; Raised in rabbit; 1:300; Abcam (ab6586)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#ab6586] Specificity Statement: Negligible cross-reactivity with Type I, II, III, V or VI collagens. Non-specific cross reaction of anticollagen antibodies with other human serum proteins or non-collagen extracellular matrix proteins is negligible.
Collagen V	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen V polyclonal; Raised in rabbit; 1:100; Abcam (ab7046)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#ab7046] Specificity Statement: Negligible cross-reactivity with Type I, II, III, IV or VI collagens. Non-specific cross reaction of anticollagen antibodies with other human serum proteins or non-collagen extracellular matrix proteins is negligible.
Collagen VI	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen VI polyclonal; Raised in rabbit; 1:200; Proteintech (17023-1-AP)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#17023-1-AP] Selected Validation Data available on supplier website
Collagen XII	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen XII polyclonal; Raised in rabbit; 1:100; Invitrogen (PA5-38890)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#PA5-38890] Immunogen: A synthetic peptide derived from the internal region of human Collagen XII alpha1 Selected Validation Data available on supplier website
Collagen XIV	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen XIV polyclonal; Raised in rabbit; 1:150; Invitrogen (PA5-54886)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#PA5-54886] Immunogen sequence: LLVTPTSGGKTNQLNLQNTATKAIIQGLMPDQNYTVQIIAYNKDKESKPAQGQFRIKDLEKRKDPKPR Selected Validation Data available on supplier website
Collagen XVI	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Collagen XVI polyclonal; Raised in rabbit; 1:100; Sigma-Aldrich (SAB4500398)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#SAB4500398] The antiserum was produced against synthesized peptide derived from human Collagen XVI alpha1. Selected Validation Data available on supplier website
EMILIN1	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	EMILIN1 monoclonal; Raised in mouse; 1:100; Proteintech (60047-1-Ig)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	Insulin polyclonal; Raised in guinea pig; 1:2000; Sigma-Aldrich (I8510)	Goat anti-guinea pig A488; 1:800; Life Technologies (A11073)	[#60047-1-Ig] Selected Validation Data available on supplier website
Fibrillin 2 (FBN2)	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	FBN2 polyclonal; Raised in rabbit; 1:200; Proteintech (20252-1-AP)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#20252-1-AP] Selected Validation Data available on supplier website
Fibrillin 2 (FBN2)	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	FBN2 polyclonal; Raised in rabbit; 1:200; Proteintech (20252-1-AP)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Glucagon monoclonal; Raised in mouse; 1:1000; Sigma-Aldrich (G2654)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#G2654] Monoclonal Anti-Glucagon reacts with pancreatic glucagon in RIA and immunocytochemistry. The affinity constant of 6.1 x 10(8) L/M in RIA. The antibody weakly cross-reacts with gut glucagon (enteroglucagon) in an immunohistological assay. Cross-reactivity has been observed with glucagon-containing cells in fixed sections of pancreas from human, porcine, dog, rabbit, mouse, rat, guinea pig, and cat.
Fibrillin 2 (FBN2)	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	FBN2 polyclonal; Raised in rabbit; 1:200; Proteintech (20252-1-AP)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Somatostatin monoclonal; Raised in mouse; 1:50; Santa Cruz Biotechnology (sc-74556)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#sc-74556] Selected Data and Citations available on supplier website
Fibronectin	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Fibronectin polyclonal; Raised in rabbit; 1:200; Abcam (ab2413)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#ab2413] Selected Validation Data available on supplier website
IgG Isotype Control (Mouse)	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Mouse IgG1, Kappa Monoclonal [B11/6] - Isotype Control; 1:100; Abcam (ab91353)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	Insulin polyclonal; Raised in guinea pig; 1:2000; Sigma-Aldrich (I8510)	Goat anti-guinea pig A488; 1:800; Life Technologies (A11073)	[#ab91353] Selected Validation Data available on supplier website
IgG Isotype Control (Rabbit)	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Rabbit IgG, monoclonal [EPR25A] - Isotype Control; 1:100; Abcam (ab172730)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#ab172730] Selected Validation Data available on supplier website
Laminin alpha 4	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Laminin alpha 4 monoclonal; Raised in mouse; 1:200; Abcam (ab242198)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	Insulin polyclonal; Raised in guinea pig; 1:2000; Sigma-Aldrich (I8510)	Goat anti-guinea pig A488; 1:800; Life Technologies (A11073)	[#ab242198] Immunogen: Recombinant fragment corresponding to Human Laminin alpha 4/LAMA4 aa 670-809. Sequence:ENLLNQARELQAKAESSSDEAVADTSRRVGGALARKSALKTRLSDAVKQLQAAERGDAQQRLGQSRLITEEANRTTM EVQQATAPMANNLTNWSQNLQHFDSSAYNTAVNSARDAVRNLTEVVPQLLDQLRTVEQKRPAS. Selected Validation Data available on supplier website
Laminin alpha 5	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Laminin alpha 5 monoclonal; Raised in mouse; 1:100; Abcam (ab77175)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	Insulin polyclonal; Raised in guinea pig; 1:2000; Sigma-Aldrich (I8510)	Goat anti-guinea pig A488; 1:800; Life Technologies (A11073)	[#ab77175] Immunogen: Full length native protein (purified) corresponding to Human Laminin alpha 5/LAMA5. Selected Validation Data available on supplier website
Mimecan (OGN)	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Mimecan polyclonal; Raised in rabbit; 1:50; Proteintech (12755-1-AP)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#12755-1-AP] Selected Validation Data available on supplier website
Periostin (POSTN)	Bovine Serum Albumin; 10%; Fisher Scientific (BP1600)	Periostin polyclonal; Raised in rabbit; 1:200; Proteintech (19899-1-AP)	Donkey anti-rabbit A488; 1:800; Life Technologies (A21206)	Insulin monoclonal; Raised in mouse; 1:5000; Sigma-Aldrich (I2018)	Goat anti-mouse A568; 1:800; Life Technologies (A11031)	[#19899-1-AP] Selected Validation Data available on supplier website
Insulin				Insulin monoclonal; Raised in mouse; Sigma-Aldrich (I2018)		[#I2018] The antibody reacts specifically against insulin by RIA and immunocytochemistry. It exhibits cross-reactivity with human proinsulin. The antibody binds to insulin with an affinity constant of 8.8 × 109 M-1 in RIA.
Insulin				Insulin polyclonal; Raised in guinea pig; Sigma-Aldrich (I8510)		[#I8510] Selected Validation Data available on supplier website and in previous publications.

Chapter 3

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

Adapted from: Shi, Y.[#]; **Li, Z.[#]**; Wang, B.; Shi, X.; Ye, H.; Delafield, D.; Lv, L.; Ye, Z.; Chen, Z.; Ma, F.; Li, L., Enabling global analysis of protein citrullination and homocitrullination via biotin thiol tag-assisted mass spectrometry. **2021.** Under review.

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 691 citrullination sites and 81 homocitrullination sites from 432 and 63 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 for further deciphering 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) (**Figure 1a**)[1-4]. 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* (**Figure 1b**)[5]. 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, 6].

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 responses[5, 7-13]. 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 transduction[14-17]. Moreover, recent accumulating evidence has revealed that citrullination and homocitrullination are associated with the development of diverse pathological states including prion disease[18], psoriasis[19], Alzheimer's disease (AD)[20-22] and cancers[23-28], 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 PTMs[29-31]. However, these approaches are neither suitable for high-

throughput analysis nor able to pinpoint exact sites of the PTMs with confidence[32-33]. 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[32-36]. 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 ^{13}C 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[13, 37-38]. Delicate searching algorithms and statistical modeling have also been developed to aid in the direct analysis[39-43]. Chemical derivatization of the PTMs prior to MS analysis is an alternative to enlarge the mass shift but usually suffers from incomplete reaction[44-46]. The above-mentioned strategies did not address the challenge associated with 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[47]. Nevertheless, previous designs led to unsatisfying fragmentation of the peptide backbones and thus limited the identified citrullination and homocitrullination sites[9, 47-48].

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[9, 29, 46-48]. Here, we design a biotin thiol tag that can be easily synthesized with low cost (**Figure S1**) and can specifically react with citrulline or homocitrulline residues together with 2,3-butanedione (**Figure 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 and two arginine residues within the sequence (SAVRACitSSVPGVR) (**Figure S2a**). After 6 h, the reaction was complete without any observable side products (**Figure S2b**), suggesting 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 tested an identical peptide containing an arginine at the modified position (SAVRARSSVPGVR) (**Figure S2c**) and no interfering reaction was observed on any arginine residues after incubation with biotin thiol tag (**Figure S2d**). To further verify the specificity and efficiency of this reaction, we repeated the experiment on six other standard peptides with varying lengths and sequences. The results proved

that our biotin thiol tag-enabled reaction is highly selective towards only ureido groups on citrulline or homocitrulline instead of other amino acid residues (**Figure S3**), which lays the foundation for further applications. Moreover, the reaction with multiple citrulline residues in the proximity also exhibited very good efficiency though with more complex in-source fragmentation pattern (**Figure S4**). We then evaluated the enrichment performance by spiking the derivatized peptide standard (SAVRACitSSVPGVR) into a complex peptide mixture (1:400, w/w) (**Figure S5a**) followed by enrichment with streptavidin beads (**Figure S5b**). 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 lower identification rates[9]. 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) (**Figure 1d** and **Figure S6a-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) (**Figure S6a, d-g**), respectively. The collected tandem MS spectra of the derivatized peptide standard delivered nearly full sequence coverage under HCD (**Figure 1e**), ETD (**Figure S6h**) or electron-transfer/higher-energy collision dissociation (EThcD) (**Figure S6i**) 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 (**Figure 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 and LysC was used during the digestion step. As a negative control, non-treated histone protein resulted in no peptides being identified as citrullinated, demonstrating again the high specificity of the biotin tag reaction (**Figure 2b**). After *in vitro* PAD treatment, we found many arginine residues were catalyzed to citrulline with abundant peptides confidently identified as citrullinated (**Figure 2c**), indicating the high efficacy of our method. A rich series of b/y ions were generated upon HCD fragmentation, allowing for unambiguous characterization of both singly (**Figure 2d**) and multiply (**Figure 2e**) citrullinated peptides.

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 (**Figure S7a**) 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 (**Figure S7b, 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 (**Figure S7d-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 in-depth characterization of citrullinated peptides in 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 (**Figure S8a**) probably because LysC digestion produces longer peptides, which results in lower fragmentation efficiency. Though trypsin digestion produced higher identification rate, different procedures generated complementary results (**Figure S8b**), suggesting the potential benefits to use a combination of multiple digestion methods. Interestingly, some peptides were identified with citrullination sites located at peptide C-termini, indicating that citrulline residues could potentially be cleaved by trypsin. 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 (**Figure S8c**), which demonstrates that no artificial cleavage of citrulline residues happens after enzymatic digestion. Some citrullination sites were confidently identified with different digestion protocols (**Figure S8d-f**), which further supported our observations of trypsin cleavable C-terminal citrullinated arginine. However, 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[37, 49], others have reported the presence of numerous C-terminal citrullination sites though manual inspection of the spectra is usually required[23, 40, 50]. These authors also suggested that the cleavage might be protein-specific and/or at a lower rate[35, 50]. Although some previous studies as well as ours provided evidence to support this observation, we believe that future investigations into the mechanisms are highly desirable and essential before including the peptide C-terminal citrullines as true hits, especially considering that Bennike *et al.* have reported the inability of trypsin to cleave citrulline *in vitro* using synthetic peptides[49].

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 strategy for interrogating the citrullination proteome. For our later applications, we remove the C-terminal citrullination identifications to ensure our results are as confident and reliable as possible. One can also utilize LysC digestion to avoid the dilemma but with a cost of lower identification rate and higher experimental expense. In addition, we evaluated the reproducibility by analyzing three biological replicates and the good overlap among these replicates indicates the reliability and robustness of our optimized methods (**Figure S9**).

Large-scale citrullinome profiling of different mouse tissues. Next, we ask whether the developed method can delineate the citrullination landscape from biological samples, and offers 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 691 citrullination sites from 432 citrullinated proteins with high confidence (**Figure 3a** and **Table S1**), which is a dramatic increase

in identification number and coverage of citrullinated proteome compared to previous studies. More importantly, about 60% of these proteins were not reported to be proteins with PTMs retrievable from the UniProt database (**Figure S10**), which suggests that our results greatly expand the knowledge and 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 (**Figure 3a**); however, the total number of citrullinated proteins in the brain is lower than that in the body (**Figure S10**). 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 (**Figure S11a, b**). We observed a higher degree of overlap between brain regions with many more shared proteins and sites in between (**Figure S11c-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 (**Figure 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 (**Figure 3b**). **Figure 3c** captures the prevalence of singly- and multiply-citrullinated proteins where 70% of the identified proteins were observed with only one citrullination site.

The newly discovered citrullination proteome provides a valuable resource to conjecture the regulatory mechanisms of this PTM. For instance, we identified nine citrullination sites on MBP while there are only four reported in the UniProt database (**Figure 3d**). Our results provided

high-quality tandem MS spectra, which not only confirmed the presence of known modification sites (**Figure 3e**), but also identified unknown sites with confidence (**Figure 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 underlying 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 molecular 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[20, 22]. In this study, we identified nine citrullination sites on GFAP compared to four described in the UniProt database (**Figure S12**), which underscored 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 (**Figure S13a**) and microtubule-associated protein tau (**Figure S13b**). These two proteins have been proven to be closely associated with the initiation and progression of AD[51-53] and our results suggest the possible roles of their citrullinated forms in the pathogenesis of these neurodegenerative diseases.

We then performed a motif analysis and found there was no strongly conserved amino acid sequence patterns flanking identified citrullination sites (**Figure 3g**), though glutamic acid residues are present more frequently in the proximity especially on the N-terminal side, which is consistent with a recent study by Fert-Bober *et al.* who also investigated citrullination in mice[39]. On the other

hand, Tanikawa *et al.* found that approximately one-fifth of the PAD4 substrates contained an RG/RGG motif in human tissue and plasma[54]. Similarly, Lee *et al.* and other researchers observed that aspartate and glycine residues were overrepresented at the +1 position[37, 55-56]. Notably, these latter studies also used human cell lines or tissues which might lead to discrepancies, and our results could indicate the sequence patterns surrounding citrullination sites are different among species. 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 (**Figure 3h**) or Kyoto encyclopedia of genes and genomes (KEGG) pathways (**Figure 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, perikaryon 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 be enriched in mitochondria. In accordance with this, we also identified four citrullination sites on an essential glycolytic enzyme pyruvate kinase (PKM). Interestingly, many of these sites on PKM are located at critical positions (**Figure S14a**), which raises the likelihood that citrullination can influence the kinase activity and supports a recent study concluding that citrullination regulates glycolysis[57]. For instance, R455 is located in the proximity of one allosteric center (**Figure S14b**), which was known to regulate enzyme activity[58]. Another position R399 which we found to be citrullinated was also shown to be very important in stabilizing the highly active tetrameric form (**Figure S14c, d**)[58-59]. Our results greatly expand current understandings of protein citrullination by demonstrating its widespread

distribution (**Figure S15**) and involvement in many other biological processes (**Figure S16**), molecular functions (**Figure S17**) and KEGG pathways (**Figure S18**).

Additionally, we noticed that 24 citrullination sites are colocalized with other arginine modifications especially omega-n-methylarginine (**Figure S19a**). 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 (**Figure S19b**). Hnrnpa2b1 was shown to influence RNA metabolism and transport, and arginine methylation could regulate the nucleocytoplasmic distribution of this protein[60]. Our results raise the possibility that citrullination indirectly participates in biological processes through an interplay with other protein modifications such as arginine methylation. Future studies will examine potential PTM crosstalk involving citrullination and their role in protein function and biological processes.

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 abundance[32]. 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 81 homocitrullination sites from 63 proteins across all the tissues (**Table S2**), which fills in the blank of protein homocitrullination database. Similarly, more sites and proteins were identified in brain regions compared to body organs, which suggests its intimate association with brain functions (**Figure 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 (**Figure 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 (**Figure 4b, c**) while both are colocalized with several lysine modifications (**Figure 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[61-66]. 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 (**Figure 4d**). As expected, homocitrullinated proteins are concentrated in myelin sheath and may function importantly in the nervous system (**Figure 4e**). They also participate in functions related to oxygen binding (**Figure 4e**), which is consistent with the observation that more homocitrullinated proteins are identified in the heart (**Figure 4a**). Interestingly, we found that they are more likely to be localized in mitochondria and correspondingly associated with processes such as tricarboxylic acid cycle (**Figure 4e**). Homocitrullinated proteins may interfere with cell-cell interactions as well which can be discerned from their prevalence in extracellular matrix components (**Figure 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 (**Figure 5a**). For isotopic labeling such as reductive dimethylation, quantification can be achieved during survey scans[67]. While for isobaric labeling approaches such as tandem mass tag (TMT)[68] or *N*,*N*-dimethyl leucine (DiLeu)[69],

quantitative information can be obtained from reporter ions upon tandem MS fragmentation (**Figure 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 (**Figure S20**). We first tested this pipeline with citrullinated peptide standard and found that the standard could be completely dimethylated without showing any observable side reactions (**Figure S21a, b**), which ensured no interference to the following steps. When differentially labeling the standard and mixing with known ratios, accurate quantitation was achieved (**Figure S21c-e**) and reliable results were obtained after biotin thiol tag derivatization (**Figure 5b**). We moved forward to evaluate this strategy using complex biological samples (**Figure S22a**). Although we observed lower identification and quantification numbers which was likely due to increased complexity of the spectra (**Figure S22b**), quantification of the citrullinated peptides exhibited both great accuracy and precision compared to theoretical ratios (**Figure 5c**). This method also provided reproducible identifications (**Figure S23a**) and consistently accurate quantitation results (**Figure S23b**) of multiple replicates, suggesting its reliability and utility in real applications. In addition, the identification and quantification rates can be 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 691citrullination sites and 81 homocitrullination sites from 432 and 63 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 multiplexed quantitative analysis. We will also integrate isobaric labeling strategies to alleviate these shortcomings and further increase the multiplexing capability for quantitative PTM analyses in future studies. 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 CH_2Cl_2 (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). ^{1}H 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 mg cm^{-3} CD_3OD 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: $C_{12}H_{22}N_3O_2S_2$; $[M+H]^+$: *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 standards (Genscript) were 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 modifications[70]. Briefly, 75 µL streptavidin agarose (Sigma) was washed with 1 mL 1× 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) or a MALDI TOF/TOF RapifleX mass spectrometer (Bruker). For MALDI-LTQ Orbitrap detection, ionization was performed using a laser energy of 15 µJ. Spectra were acquired with appropriate mass ranges at a resolution of 30k (at *m/z* 400). For MALDI TOF/TOF RapifleX detection, ionization was performed using a laser power of 80% with 2000 laser shots at scan frequency of 5000. Spectra were acquired with appropriate mass ranges with a sample rate of 5 GS/s in reflectron mode.

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 a ratio of 2 µg enzyme per mg of histone in a buffer containing 100 mM Tris and 5 mM $CaCl_2$ (pH 7.5). The catalysis was quenched by adding 50 mM EDTA solution. LysC (Promega) was then added to histone samples with or without PAD treatment in a 50:1 ratio (protein:enzyme, w/w) and incubated overnight at 37 °C. Digestion was quenched by adding TFA to pH <3 and desalted using Omix Tips (Agilent) before drying in vacuo.

Protein extraction and digestion of mouse tissues. For method optimization, brain tissue 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 ℃) 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 ℃ 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 ℃ 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 μL 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 μm inner diameter homemade microcapillary column which is packed with 15 cm of Bridged Ethylene Hybrid C18 particles (1.7 μm, 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% to 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 Metascape[71] (version 3.5) while GO analysis for homocitrullination was accomplished using DAVID bioinformatics resources[72] with a FDR cutoff of 0.05. Sequence motif analyses were done using WebLogo[73]. 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.

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. The MALDI TOF/TOF RapifleX mass spectrometer was purchased through the support of an NIH shared instrument grant S10OD025084. 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.

References

(1) Fuhrmann, J.; Clancy, K. W.; Thompson, P. R. *Chem. Rev.* **2015,** *115*, 5413-61.

(2) Fuhrmann, J.; Thompson, P. R. *ACS chemical biology* **2016,** *11*, 654-68.

(3) Witalison, E.; Thompson, P.; Hofseth, L. *Current Drug Targets* **2015,** *16*, 700-710.

(4) Mondal, S.; Thompson, P. R. *Acc Chem Res* **2019,** *52*, 818-832.

(5) Pruijn, G. J. *Front Immunol* **2015,** *6*, 192.

(6) Gyorgy, B.; Toth, E.; Tarcsa, E.; Falus, A.; Buzas, E. I. *The international journal of biochemistry & cell biology* **2006,** *38*, 1662-77.

(7) Schellekens, G. A.; de Jong, B. A.; van den Hoogen, F. H.; van de Putte, L. B.; van Venrooij, W. J. *The Journal of clinical investigation* **1998,** *101*, 273-81.

(8) Elkon, K. B. *Sci Transl Med* **2013,** *5*, 209fs39.

(9) Tutturen, A. E.; Fleckenstein, B.; de Souza, G. A. *Journal of proteome research* **2014,** *13*, 2867-73.

(10) Turunen, S.; Huhtakangas, J.; Nousiainen, T.; Valkealahti, M.; Melkko, J.; Risteli, J.; Lehenkari, P. *Arthritis Res Ther* **2016,** *18*, 239.

(11) Tilvawala, R.; Nguyen, S. H.; Maurais, A. J.; Nemmara, V. V.; Nagar, M.; Salinger, A. J.; Nagpal, S.; Weerapana, E.; Thompson, P. R. *Cell Chem Biol* **2018,** *25*, 691-704 e6.

(12) Fert-Bober, J.; Darrah, E.; Andrade, F. *Immunol Rev* **2020,** *294*, 133-147.

(13) Raijmakers, R.; van Beers, J. J.; El-Azzouny, M.; Visser, N. F.; Bozic, B.; Pruijn, G. J.; Heck, A. J. *Arthritis Res Ther* **2012,** *14*, R114.

(14) Moscarello, M. A.; Mastronardi, F. G.; Wood, D. D. *Neurochem Res* **2007,** *32*, 251-6.

(15) Bradford, C. M.; Ramos, I.; Cross, A. K.; Haddock, G.; McQuaid, S.; Nicholas, A. P.; Woodroofe, M. N. *J Neuroimmunol* **2014,** *273*, 85-95.

(16) Yang, L.; Tan, D.; Piao, H. *Neurochem Res* **2016,** *41*, 1845-56.

(17) Gs Chirivi, R. *J. Clin. Cell. Immunol.* **2013,** *04*.

(18) Jang, B.; Kim, E.; Choi, J. K.; Jin, J. K.; Kim, J. I.; Ishigami, A.; Maruyama, N.; Carp, R. I.; Kim, Y. S.; Choi, E. K. *The American journal of pathology* **2008,** *173*, 1129-42.

(19) Ishida-Yamamoto, A.; Senshu, T.; Takahashi, H.; Akiyama, K.; Nomura, K.; Iizuka, H. *J Invest Dermatol* **2000,** *114*, 701-5.

(20) Ishigami, A.; Ohsawa, T.; Hiratsuka, M.; Taguchi, H.; Kobayashi, S.; Saito, Y.; Murayama, S.; Asaga, H.; Toda, T.; Kimura, N.; Maruyama, N. *Journal of neuroscience research* **2005,** *80*, 120-8.

(21) Acharya, N. K.; Nagele, E. P.; Han, M.; Coretti, N. J.; DeMarshall, C.; Kosciuk, M. C.; Boulos, P. A.; Nagele, R. G. *J Autoimmun* **2012,** *38*, 369-80.

(22) Ishigami, A.; Masutomi, H.; Handa, S.; Nakamura, M.; Nakaya, S.; Uchida, Y.; Saito, Y.; Murayama, S.; Jang, B.; Jeon, Y. C.; Choi, E. K.; Kim, Y. S.; Kasahara, Y.; Maruyama, N.; Toda, T. *Journal of neuroscience research* **2015,** *93*, 1664-74.

(23) Yuzhalin, A. E.; Gordon-Weeks, A. N.; Tognoli, M. L.; Jones, K.; Markelc, B.; Konietzny, R.; Fischer, R.; Muth, A.; O'Neill, E.; Thompson, P. R.; Venables, P. J.; Kessler, B. M.; Lim, S. Y.; Muschel, R. J. *Nat. Commun.* **2018,** *9*, 4783.

(24) Yuzhalin, A. E. *Cancer Res* **2019,** *79*, 1274-1284.

(25) Stadler, S. C.; Vincent, C. T.; Fedorov, V. D.; Patsialou, A.; Cherrington, B. D.; Wakshlag, J. J.; Mohanan, S.; Zee, B. M.; Zhang, X.; Garcia, B. A.; Condeelis, J. S.; Brown, A. M.; Coonrod, S. A.; Allis, C. D. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110*, 11851-6.

(26) Chang, X.; Han, J.; Pang, L.; Zhao, Y.; Yang, Y.; Shen, Z. *BMC Cancer* **2009,** *9*, 40.

(27) Wang, L.; Song, G.; Zhang, X.; Feng, T.; Pan, J.; Chen, W.; Yang, M.; Bai, X.; Pang, Y.; Yu, J.; Han, J.; Han, B. *Cancer Res* **2017,** *77*, 5755-5768.

(28) Thalin, C.; Lundstrom, S.; Seignez, C.; Daleskog, M.; Lundstrom, A.; Henriksson, P.; Helleday, T.; Phillipson, M.; Wallen, H.; Demers, M. *PloS one* **2018,** *13*, e0191231.

(29) Senshu, T.; Sato, T.; Inoue, T.; Akiyama, K.; Asaga, H. *Anal. Biochem.* **1992,** *203*, 94-100.

(30) Nicholas, A. P.; King, J. L.; Sambandam, T.; Echols, J. D.; Gupta, K. B.; McInnis, C.; Whitaker, J. N. *The Journal of comparative neurology* **2003,** *459*, 251-66.

(31) Moelants, E. A.; Van Damme, J.; Proost, P. *PloS one* **2011,** *6*, e28976.

(32) Verheul, M. K.; van Veelen, P. A.; van Delft, M. A. M.; de Ru, A.; Janssen, G. M. C.; Rispens, T.; Toes, R. E. M.; Trouw, L. A. *Autoimmun Rev* **2018,** *17*, 136-141.

(33) Hensen, S. M.; Pruijn, G. J. *Mol Cell Proteomics* **2014,** *13*, 388-96.

(34) Clancy, K. W.; Weerapana, E.; Thompson, P. R. *Curr Opin Chem Biol* **2016,** *30*, 1-6.

(35) Vitorino, R.; Guedes, S.; Vitorino, C.; Ferreira, R.; Amado, F.; Van Eyk, J. E. *Journal of proteome research* **2020**.

(36) De Ceuleneer, M.; Van Steendam, K.; Dhaenens, M.; Deforce, D. *Proteomics* **2012,** *12*, 752-60.

(37) Lee, C. Y.; Wang, D.; Wilhelm, M.; Zolg, D. P.; Schmidt, T.; Schnatbaum, K.; Reimer, U.; Ponten, F.; Uhlen, M.; Hahne, H.; Kuster, B. *Mol Cell Proteomics* **2018,** *17*, 1378-1391.

(38) Stobernack, T.; Glasner, C.; Junker, S.; Gabarrini, G.; de Smit, M.; de Jong, A.; Otto, A.; Becher, D.; van Winkelhoff, A. J.; van Dijl, J. M. *Journal of proteome research* **2016,** *15*, 4532-4543.

(39) Fert-Bober, J.; Venkatraman, V.; Hunter, C. L.; Liu, R.; Crowgey, E. L.; Pandey, R.; Holewinski, R. J.; Stotland, A.; Berman, B. P.; Van Eyk, J. E. *Journal of proteome research* **2019**.

(40) Wang, X.; Swensen, A. C.; Zhang, T.; Piehowski, P. D.; Gaffrey, M. J.; Monroe, M. E.; Zhu, Y.; Dong, H.; Qian, W. J. *Journal of proteome research* **2020,** *19*, 1863-1872.

(41) Villacres, C.; Spicer, V.; Krokhin, O. V. *Journal of proteome research* **2021**.

(42) Huh, S.; Hwang, D.; Kim, M. S. *Anal. Chem.* **2020,** *92*, 12975-12986.

(43) De Ceuleneer, M.; Van Steendam, K.; Dhaenens, M.; Elewaut, D.; Deforce, D. *Journal of proteome research* **2012,** *11*, 5245-51.

(44) Stensland, M.; Holm, A.; Kiehne, A.; Fleckenstein, B. *Rapid Commun Mass Spectrom* **2009,** *23*, 2754-62.

(45) De Ceuleneer, M.; De Wit, V.; Van Steendam, K.; Van Nieuwerburgh, F.; Tilleman, K.; Deforce, D. *Rapid Commun Mass Spectrom* **2011,** *25*, 1536-42.

(46) Choi, M.; Song, J. S.; Kim, H. J.; Cha, S.; Lee, E. Y. *Analytical biochemistry* **2013,** *437*, 62-7.

(47) Lewallen, D. M.; Bicker, K. L.; Subramanian, V.; Clancy, K. W.; Slade, D. J.; Martell, J.; Dreyton, C. J.; Sokolove, J.; Weerapana, E.; Thompson, P. R. *ACS chemical biology* **2015,** *10*, 2520-8.

(48) Tutturen, A. E.; Holm, A.; Fleckenstein, B. *Anal Bioanal Chem* **2013,** *405*, 9321-31.

(49) Bennike, T.; Lauridsen, K. B.; Olesen, M. K.; Andersen, V.; Birkelund, S.; Stensballe, A. *Journal of Proteomics & Bioinformatics* **2013,** *6*.

(50) Jin, Z.; Fu, Z.; Yang, J.; Troncosco, J.; Everett, A. D.; Van Eyk, J. E. *Proteomics* **2013,** *13*, 2682-91.

(51) Yamazaki, Y.; Zhao, N.; Caulfield, T. R.; Liu, C.-C.; Bu, G. *Nature Reviews Neurology* **2019,** *15*, 501-518.

(52) Liu, C. C.; Liu, C. C.; Kanekiyo, T.; Xu, H.; Bu, G. *Nat Rev Neurol* **2013,** *9*, 106-18.

(53) Iqbal, K.; Liu, F.; Gong, C. X.; Grundke-Iqbal, I. *Curr Alzheimer Res* **2010,** *7*, 656-64.

(54) Tanikawa, C.; Ueda, K.; Suzuki, A.; Iida, A.; Nakamura, R.; Atsuta, N.; Tohnai, G.; Sobue, G.; Saichi, N.; Momozawa, Y.; Kamatani, Y.; Kubo, M.; Yamamoto, K.; Nakamura, Y.; Matsuda, K. *Cell Rep* **2018,** *22*, 1473-1483.

(55) Stensland, M. E.; Pollmann, S.; Molberg, O.; Sollid, L. M.; Fleckenstein, B. *Biol. Chem.* **2009,** *390*, 99-107.

(56) Assohou-Luty, C.; Raijmakers, R.; Benckhuijsen, W. E.; Stammen-Vogelzangs, J.; de Ru, A.; van Veelen, P. A.; Franken, K. L.; Drijfhout, J. W.; Pruijn, G. J. *Biochimica et biophysica acta* **2014,** *1844*, 829-36.

(57) Coassolo, S.; Davidson, G.; Negroni, L.; Gambi, G.; Daujat, S.; Romier, C.; Davidson, I. *bioRxiv* **2020**, 718486.

(58) Zhang, Z.; Deng, X.; Liu, Y.; Liu, Y.; Sun, L.; Chen, F. *Cell Biosci* **2019,** *9*, 52.

(59) Gao, X.; Wang, H.; Yang, J. J.; Liu, X.; Liu, Z. R. *Mol Cell* **2012,** *45*, 598-609.

(60) Friend, L. R.; Landsberg, M. J.; Nouwens, A. S.; Wei, Y.; Rothnagel, J. A.; Smith, R. *PloS one* **2013,** *8*, e75669.

(61) Ye, J.; Ai, X.; Eugeni, E. E.; Zhang, L.; Carpenter, L. R.; Jelinek, M. A.; Freitas, M. A.; Parthun, M. R. *Mol Cell* **2005,** *18*, 123-30.

(62) Yan, Q.; Dutt, S.; Xu, R.; Graves, K.; Juszczynski, P.; Manis, J. P.; Shipp, M. A. *Mol Cell* **2009,** *36*, 110-20.

(63) Bao, X.; Liu, Z.; Zhang, W.; Gladysz, K.; Fung, Y. M. E.; Tian, G.; Xiong, Y.; Wong, J. W. H.; Yuen, K. W. Y.; Li, X. D. *Mol Cell* **2019,** *76*, 660-675 e9.

(64) Xie, Z.; Dai, J.; Dai, L.; Tan, M.; Cheng, Z.; Wu, Y.; Boeke, J. D.; Zhao, Y. *Mol Cell Proteomics* **2012,** *11*, 100-7.

(65) Dai, L.; Peng, C.; Montellier, E.; Lu, Z.; Chen, Y.; Ishii, H.; Debernardi, A.; Buchou, T.; Rousseaux, S.; Jin, F.; Sabari, B. R.; Deng, Z.; Allis, C. D.; Ren, B.; Khochbin, S.; Zhao, Y. *Nat Chem Biol* **2014,** *10*, 365-70.

(66) Chen, Y.; Sprung, R.; Tang, Y.; Ball, H.; Sangras, B.; Kim, S. C.; Falck, J. R.; Peng, J.; Gu, W.; Zhao, Y. *Mol Cell Proteomics* **2007,** *6*, 812-9.

(67) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. *Nat Protoc* **2009,** *4*, 484-94.

(68) Li, J.; Van Vranken, J. G.; Pontano Vaites, L.; Schweppe, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H.; Kuhn, K.; Pike, I.; Bomgarden, R. D.; Rogers, J. C.; Gygi, S. P.; Paulo, J. A. *Nature methods* **2020,** *17*, 399-404.

(69) Frost, D. C.; Greer, T.; Li, L. *Anal. Chem.* **2015,** *87*, 1646-54.

(70) Schiapparelli, L. M.; McClatchy, D. B.; Liu, H. H.; Sharma, P.; Yates, J. R., 3rd; Cline, H. T. *Journal of proteome research* **2014,** *13*, 3966-78.

(71) Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. *Nat. Commun.* **2019,** *10*, 1523.

(72) Huang da, W.; Sherman, B. T.; Lempicki, R. A. *Nat Protoc* **2009,** *4*, 44-57.

(73) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res* **2004,** *14*, 1188-90.

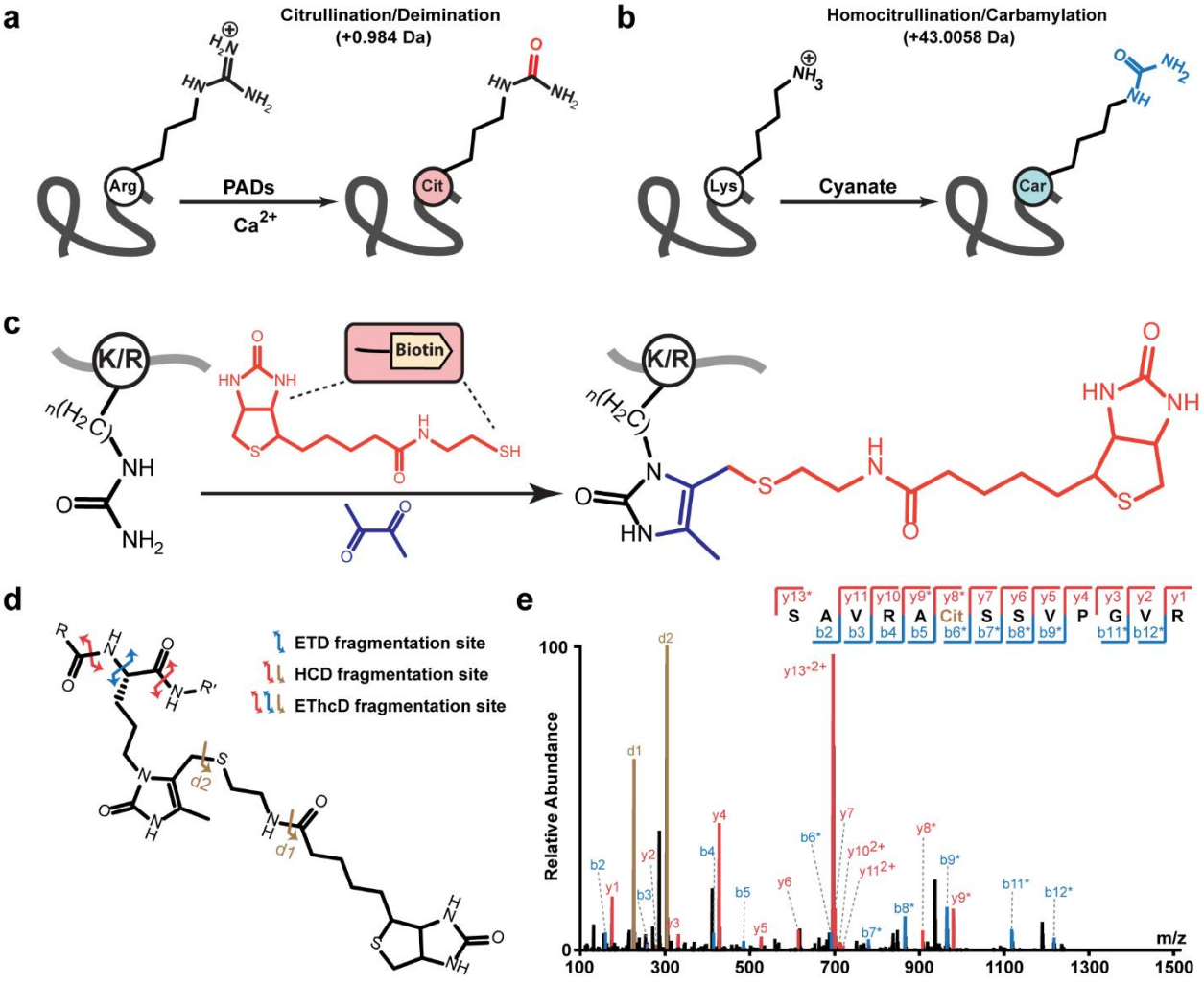

Figure 1. Design of biotin thiol tag for citrullination and homocitrullination analysis. a, PAD enzymes catalyze the hydrolytic conversion of peptidyl arginine residues into citrulline. **b**, Schematic showing homocitrullination process resulting from the chemical modification of lysine residues. **c**, 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 SAVRACitSSVPGVR upon HCD fragmentation.

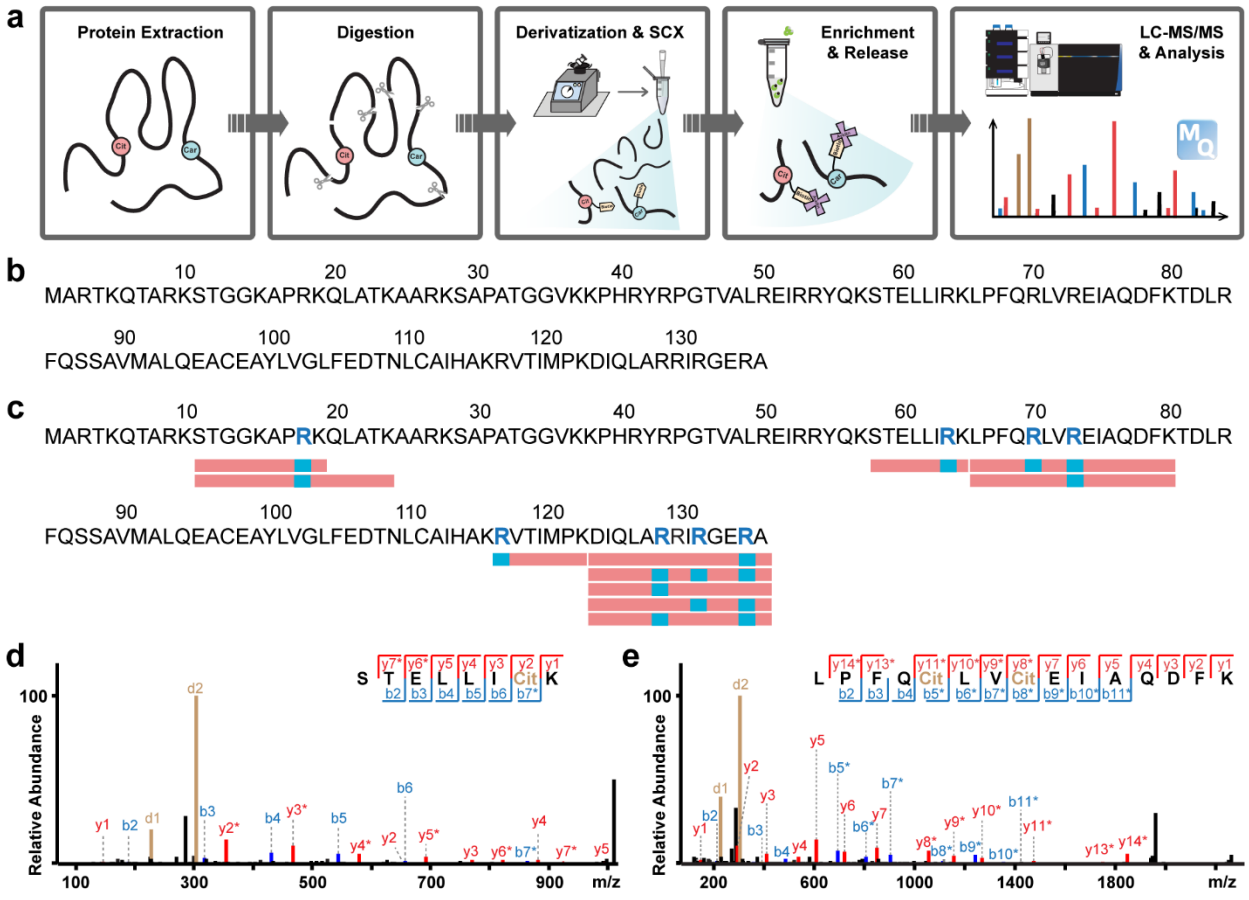

Figure 2. Improved *in vitro* protein citrullination analysis with biotin thiol tag. a, Experimental workflow of protein citrullination analysis with biotin thiol tag. **b**,**c**, Citrullination analysis on histone H3 protein before (**b**) and after (**c**) *in vitro* PAD treatment. Identified citrullination 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 (R64Cit). **e**, Example tandem MS spectrum of two citrullination sites (R70Cit and R73Cit) identified on the same peptide from PAD treated histone H3.

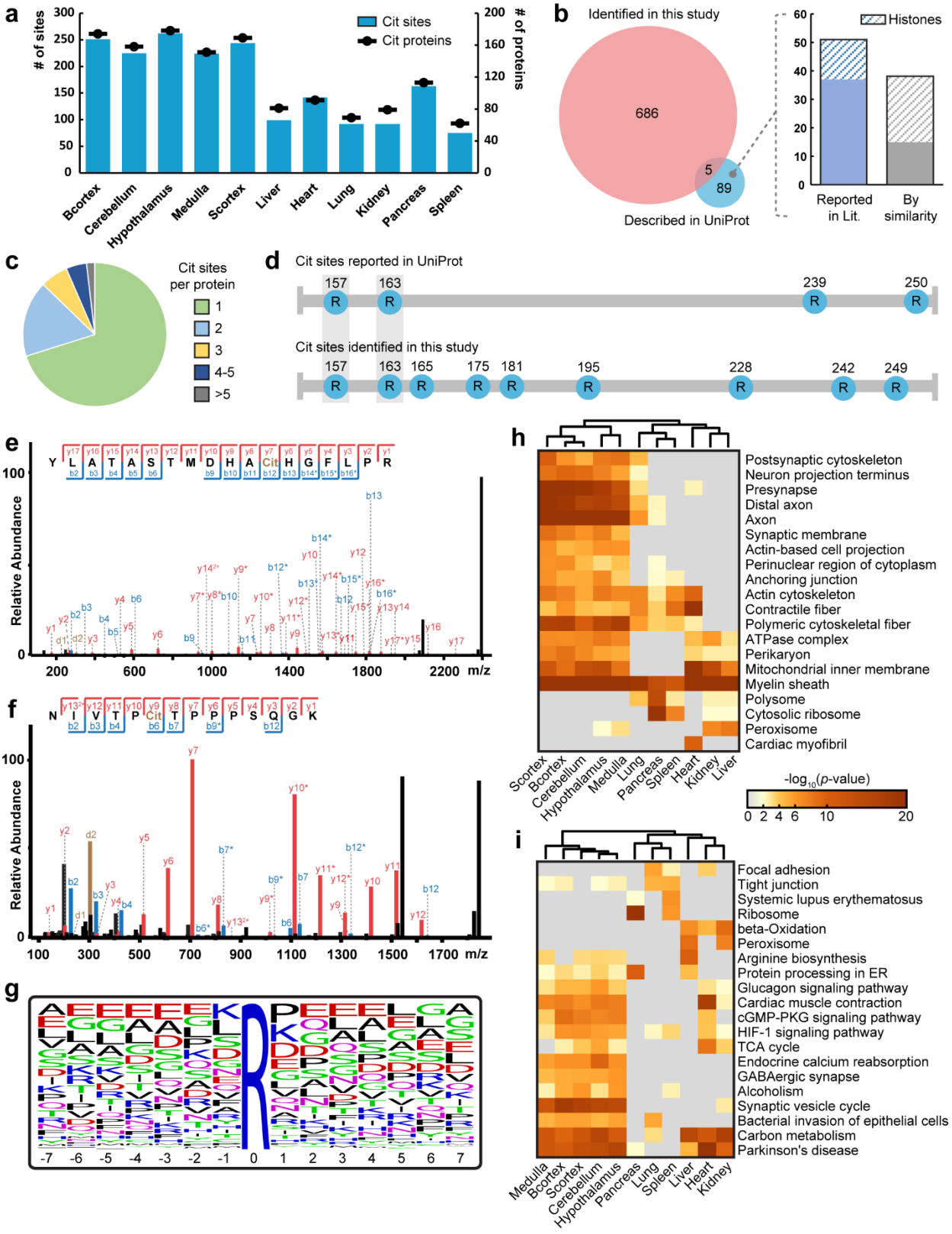
a
# of sites
# of proteins
Cit sites
Cit proteins
Bcortex
Cerebellum
Hypothalamus
Medulla
Scortex
Liver
Heart
Lung
Kidney
Pancreas
Spleen
b
Identified in this study
686
5
89
Described in UniProt
Histones
Reported in Lit.
By similarity
c
Cit sites per protein
1
2
3
4-5
>5
d
Cit sites reported in UniProt
157
163
239
250
Cit sites identified in this study
157
163
165
175
181
195
228
242
249
e
Relative Abundance
m/z
f
Relative Abundance
m/z
g
h
Postsynaptic cytoskeleton
Neuron projection terminus
Presynapse
Distal axon
Axon
Synaptic membrane
Actin-based cell projection
Perinuclear region of cytoplasm
Anchoring junction
Actin cytoskeleton
Contractile fiber
Polymeric cytoskeletal fiber
ATPase complex
Perikaryon
Mitochondrial inner membrane
Myelin sheath
Polysome
Cytosolic ribosome
Peroxisome
Cardiac myofibril
Scortex
Bcortex
Cerebellum
Hypothalamus
Medulla
Lung
Pancreas
Spleen
Heart
Kidney
Liver
-log10(p-value)
i
Focal adhesion
Tight junction
Systemic lupus erythematosus
Ribosome
beta-Oxidation
Peroxisome
Arginine biosynthesis
Protein processing in ER
Glucagon signaling pathway
Cardiac muscle contraction
cGMP-PKG signaling pathway
HIF-1 signaling pathway
TCA cycle
Endocrine calcium reabsorption
GABAergic synapse
Alcoholism
Synaptic vesicle cycle
Bacterial invasion of epithelial cells
Carbon metabolism
Parkinson's disease
Medulla
Bcortex
Scortex
Cerebellum
Hypothalamus
Pancreas
Lung
Spleen
Liver
Heart
Kidney

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 citrullination sites identified in this study with those reported in the UniProt database. Many of the sites only reported in UniProt have that description based on similarity prediction or location on histone proteins. **c**, Distribution of the number of citrullination sites per citrullinated proteins identified. **d**, Comparison of citrullination sites identified in this study and those reported in the UniProt database on myelin basic protein. **e**,**f**, Example tandem MS spectra of two citrullination sites identified on myelin basic protein, R157Cit (**e**) and R228Cit (**f**). **g**, Sequence motif of identified citrullinated peptides. Citrullination sites are centered in the middle as "0" position. The height of letters indicates the relative frequency of each amino acid at certain positions. **h**,**i**, Heatmaps generated using Metascape showing the significantly enriched (p value < 0.01 in at least one tissue region) cellular components (**h**) and Kyoto encyclopedia of genes and genomes (KEGG) pathways (**i**) in different mouse tissues. The most significant 20 terms are shown in each heatmap. Color coding indicates $-\log_{10}$ (p values). Columns are clustered based on their profile similarity.

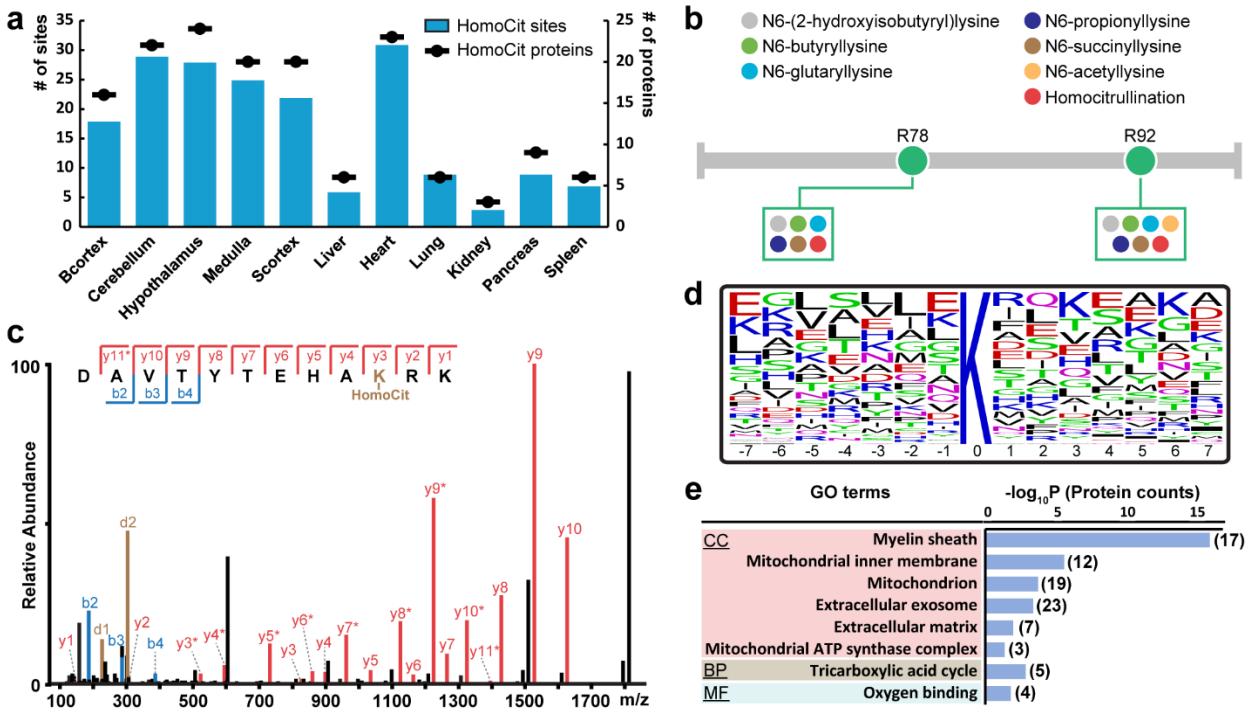

Figure 4. Profiling of protein homocitrullination in different mouse tissues. a, Number of identified homocitrullinated proteins and homocitrullination sites in different mouse tissues. Identification numbers are presented as a total from three mice. HomoCit, homocitrullination; Bcortex, cerebral cortex; Scortex, hippocampus and thalamus. **b**, Colocalization of homocitrullination sites identified in this study with other lysine modifications reported in UniProt database on histone H4. **c**, Example tandem MS spectrum of a homocitrullination site identified on histone H4 (K78HomoCit). **d**, Sequence motif of identified homocitrullinated peptides. Homocitrullination sites are centered in the middle as "0" position. The height of letters indicates the relative frequency of each amino acid at certain positions. Black rectangles within 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.

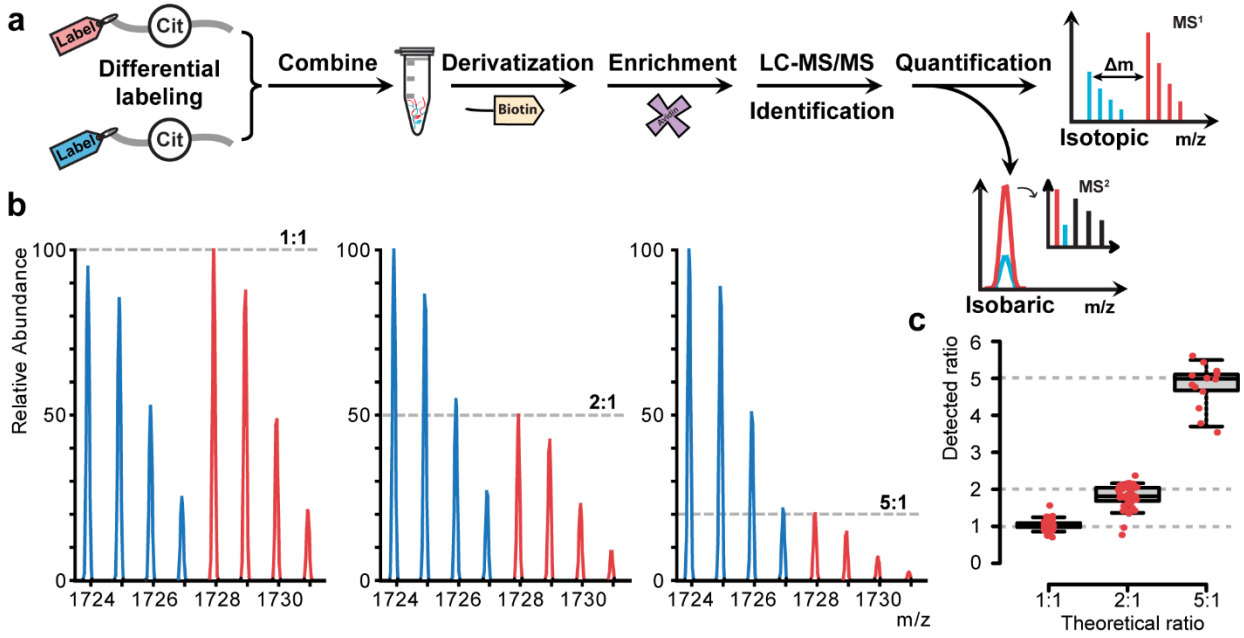

Figure 5. Multiplexed quantitative citrullination analysis using chemical labeling strategies. **a**, Schematic showing the pipeline for simultaneous qualitative and quantitative analysis of citrullination using isotopic or isobaric labeling. Relative quantification can be achieved during survey scans or tandem MS scans, respectively. **b**, Spectra showing quantitation accuracy in duplex dimethyl labeling using citrullinated peptide standard SAVRACitSSVPGVR. 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.

Supplemental Information

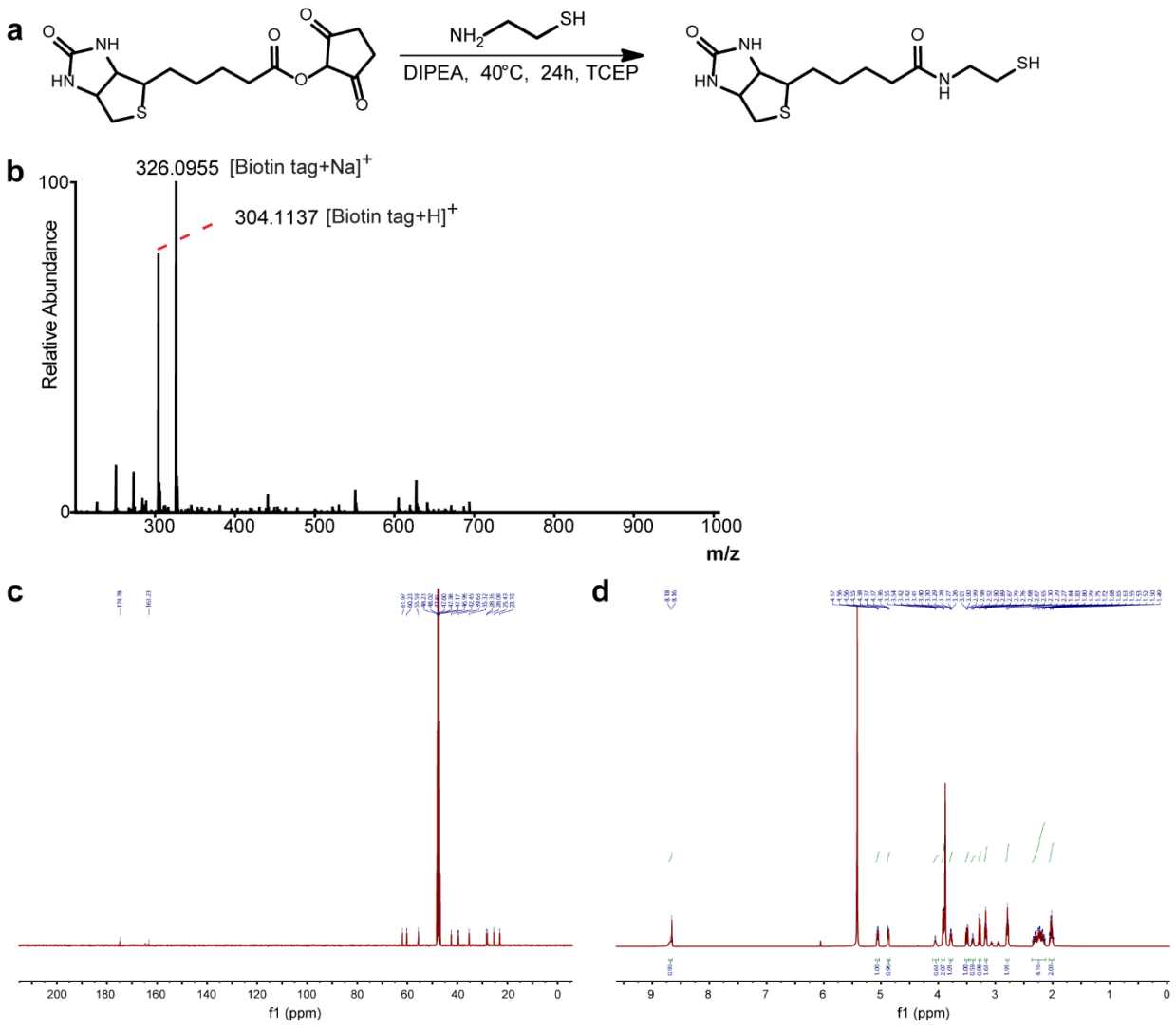

Figure S1. Synthesis and characterization of biotin thiol tag. a, One-step synthesis of biotin thiol tag using biotin-NHS ester and cysteamine. **b**, MS spectrum of biotin thiol tag. Low-intensity peaks come from the impurity of the ingredients. **c**, ^{13}C-NMR spectrum of biotin thiol tag. **d**, ^{1}H-NMR spectrum of biotin thiol tag.

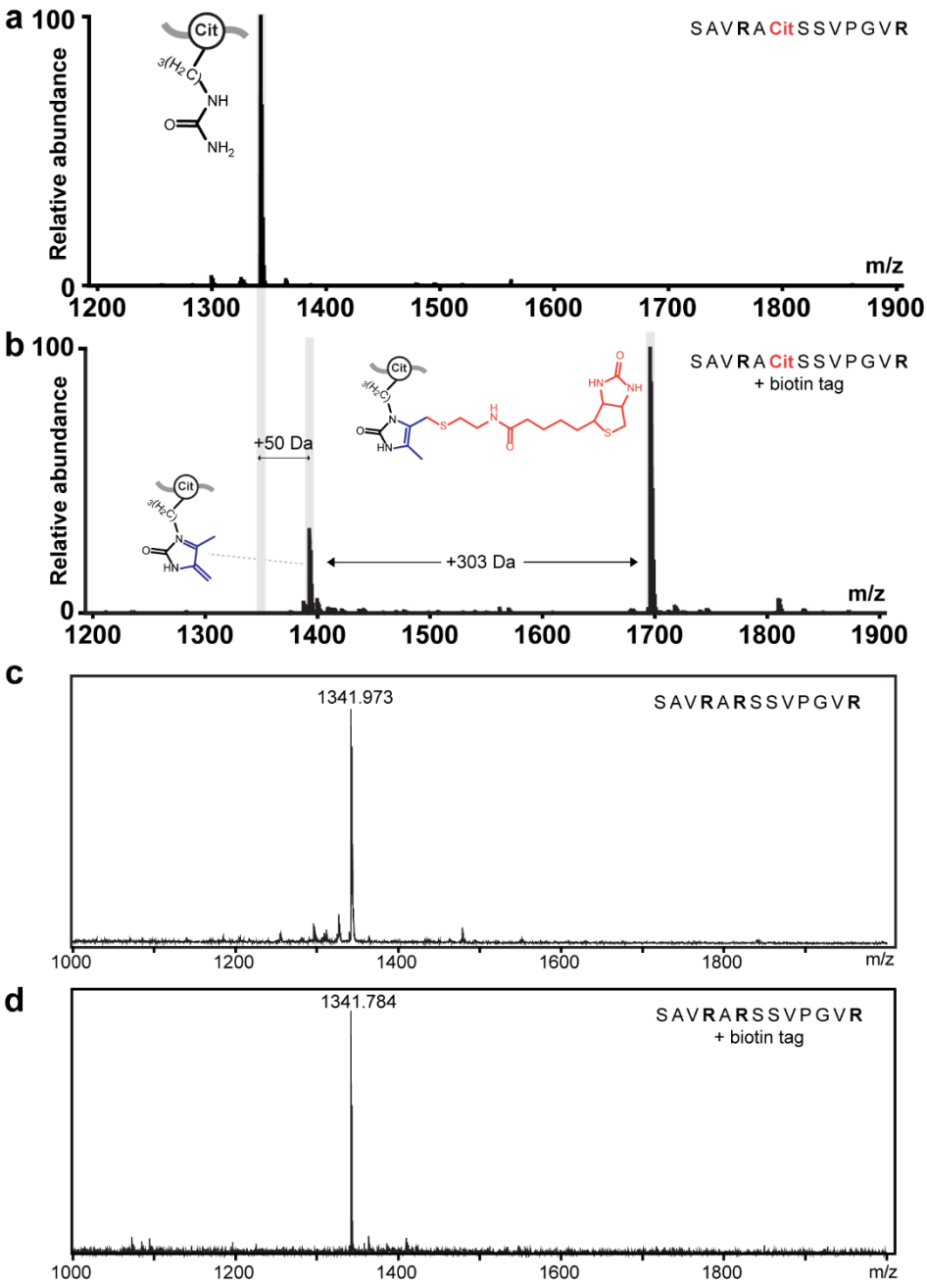

Figure S2. Proof-of-principle test of biotin thiol tag using citrullinated peptide standard. a, Spectrum of citrullinated peptide standard SAVRACitSSVPGVR. **b**, Spectrum of derivatized citrullinated peptide standard SAVRACitSSVPGVR with biotin thiol tag and 2,3-butanedione (+ 353 Da). The lower peak originates from the loss of biotin moiety due to in-source fragmentation when using a MALDI source (+ 50 Da). Citrullinated residue is indicated in red font. **c**, Spectrum

of peptide standard SAVRARSSVPGVR. **d**, Spectrum of peptide standard SAVRARSSVPGVR after incubation with biotin thiol tag and 2,3-butanedione. No mass shift is observed due to a lack of citrulline residue in the peptide sequence.

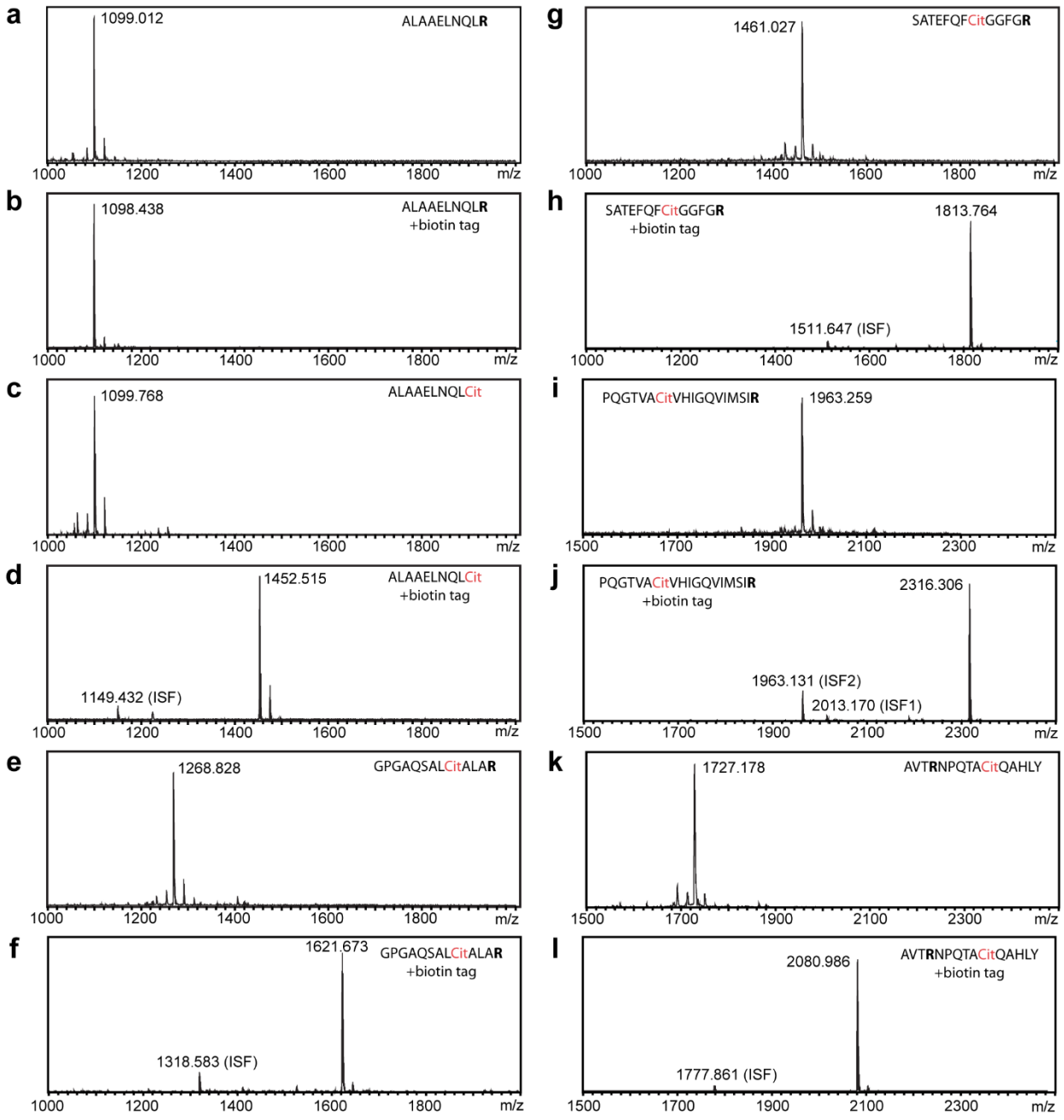

Figure S3. Verification of reaction specificity and efficiency of biotin thiol tag. Spectra of six peptide standards before and after incubation with biotin thiol tag and 2,3-butanedione. Citrullinated residue is indicated in red font. ISF, in-source fragmentation.

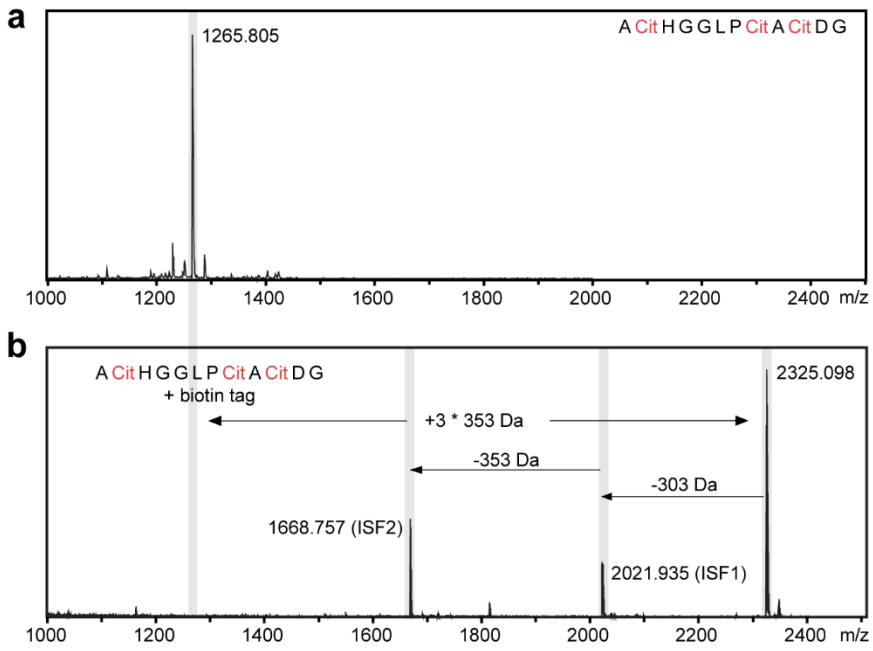

Figure S4. Reaction of multiple citrulline residues in the proximity. a, Spectrum of a peptide standard containing three citrulline residues within the sequence. **b**, Spectrum of the peptide standard after derivatization with biotin thiol tag and 2,3-butanedione. ISF, in-source fragmentation. Citrullinated residue is indicated in red font.

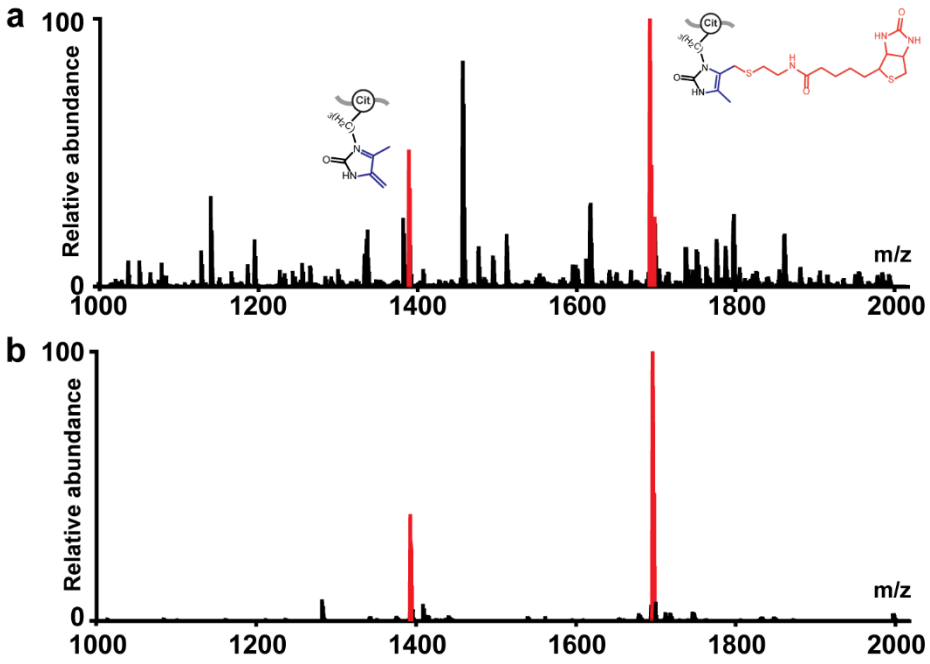

Figure S5. Evaluation of enrichment performance using citrullinated peptide standard. a, Spectrum of derivatized citrullinated peptide standard mixed with 400 μg complex peptide mixture. The higher red peak indicates the biotin thiol tag-derivatized standard peptide SAVRACitSSVPGVR and the lower peak indicates its in-source fragmentation, similar to that in Supplementary Fig. 2b. **b**, Spectrum of enriched derivatized citrullinated peptide standard SAVRACitSSVPGVR from the peptide mixture.

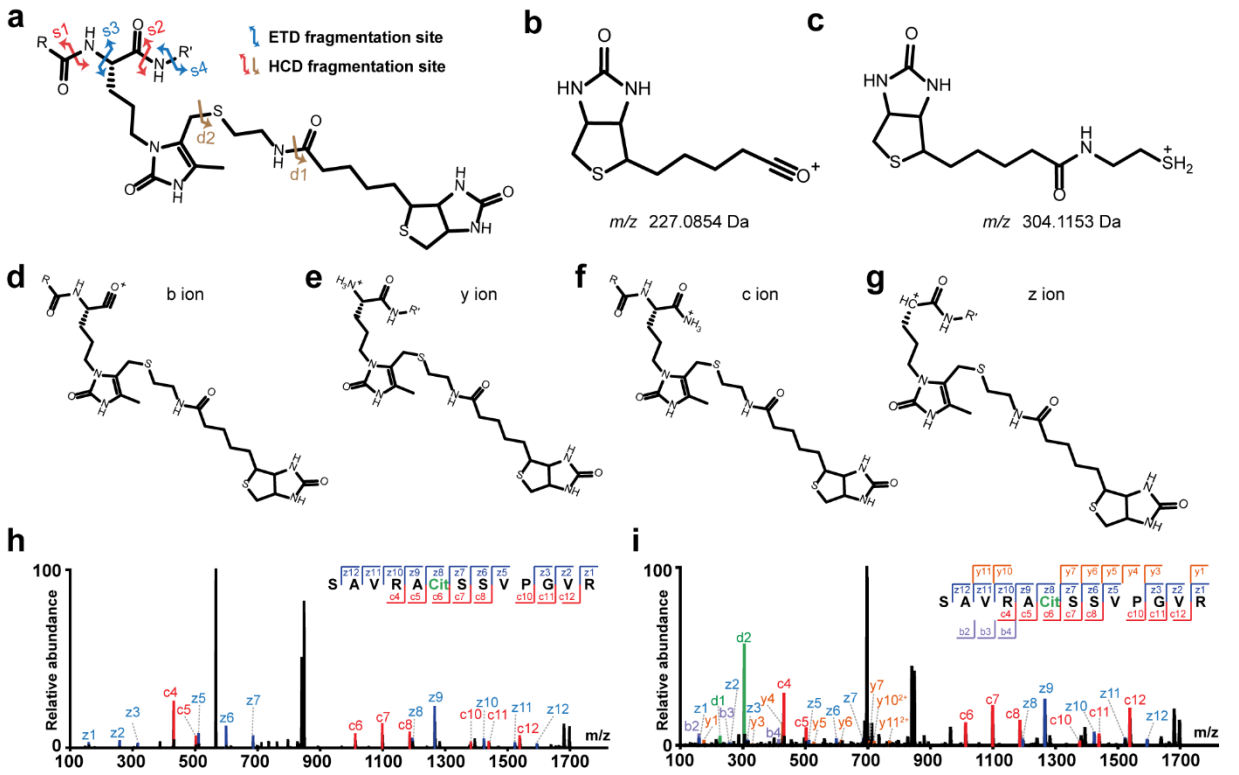

Figure S6. Fragmentation behavior of biotin thiol tag-derivatized citrullinated peptides. a, Fragmentation sites of biotin thiol tag-derivatized citrullinated peptides upon HCD or ETD fragmentations. **b**,**c**, Two observed diagnostic ions at *m/z* 227 (**b**) and *m/z* 304 (**c**) upon HCD fragmentation which correspond to d1 or d2 in (**a**). **d**, Fragment b ion generated at s2 position in (**a**). **e**, Fragment y ion generated at s1 position in (**a**). **f**, Fragment c ion generated at s4 position in (**a**). **g**, Fragment z ion generated at s3 position in (**a**). **h**, Tandem MS spectrum of the biotin thiol tag-derivatized citrullinated peptide standard upon ETD fragmentation. **i**, Tandem MS spectrum of the derivatized citrullinated peptide standard upon EThcD fragmentation. Citrullinated residue is indicated in green font, and diagnostic ion pair d1 and d2 are highlighted in green in the tandem MS spectrum.

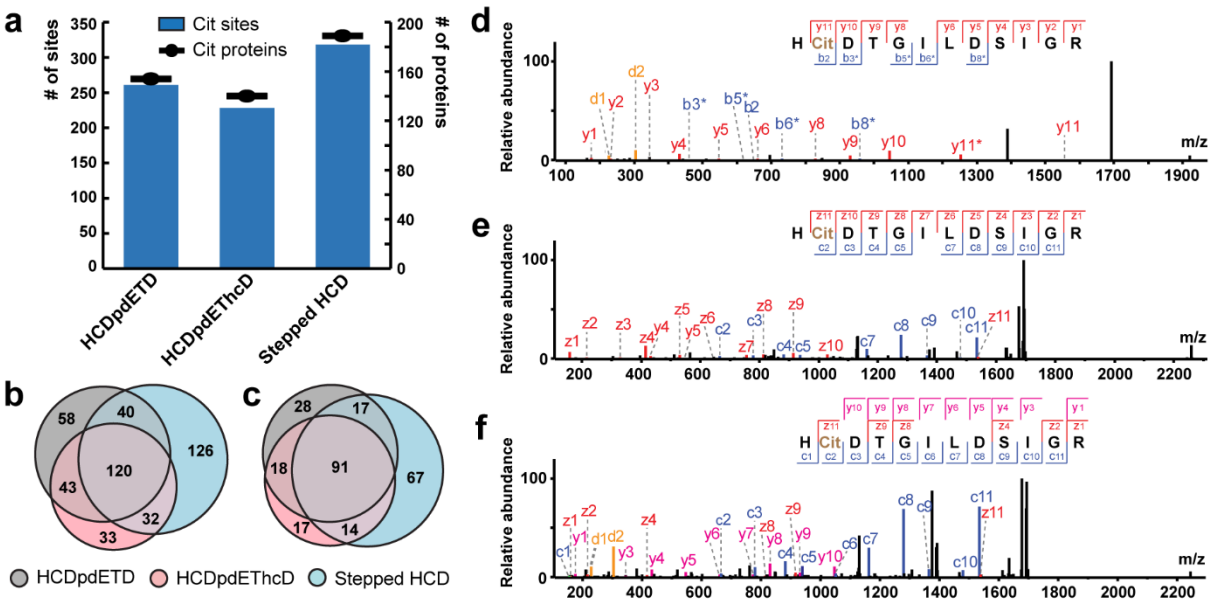

Figure S7. Optimization of different fragmentation methods in detecting protein citrullination. **a**, Number of identified citrullination sites and citrullinated proteins with three different fragmentation methods. Cit, citrullination; pd, product ion triggered. **b**, Venn diagram showing the overlap of citrullination sites identified with different fragmentation methods. **c**, Venn diagram showing the overlap of citrullinated proteins identified with different fragmentation methods. **d**, Tandem MS spectrum of one citrullinated peptide (R165Cit) on myelin basic protein identified with HCD fragmentation. **e**, Tandem MS spectrum of R165Cit on myelin basic protein identified with ETD fragmentation. **f**, Tandem MS spectrum of R165Cit on myelin basic protein identified with EThcD fragmentation.

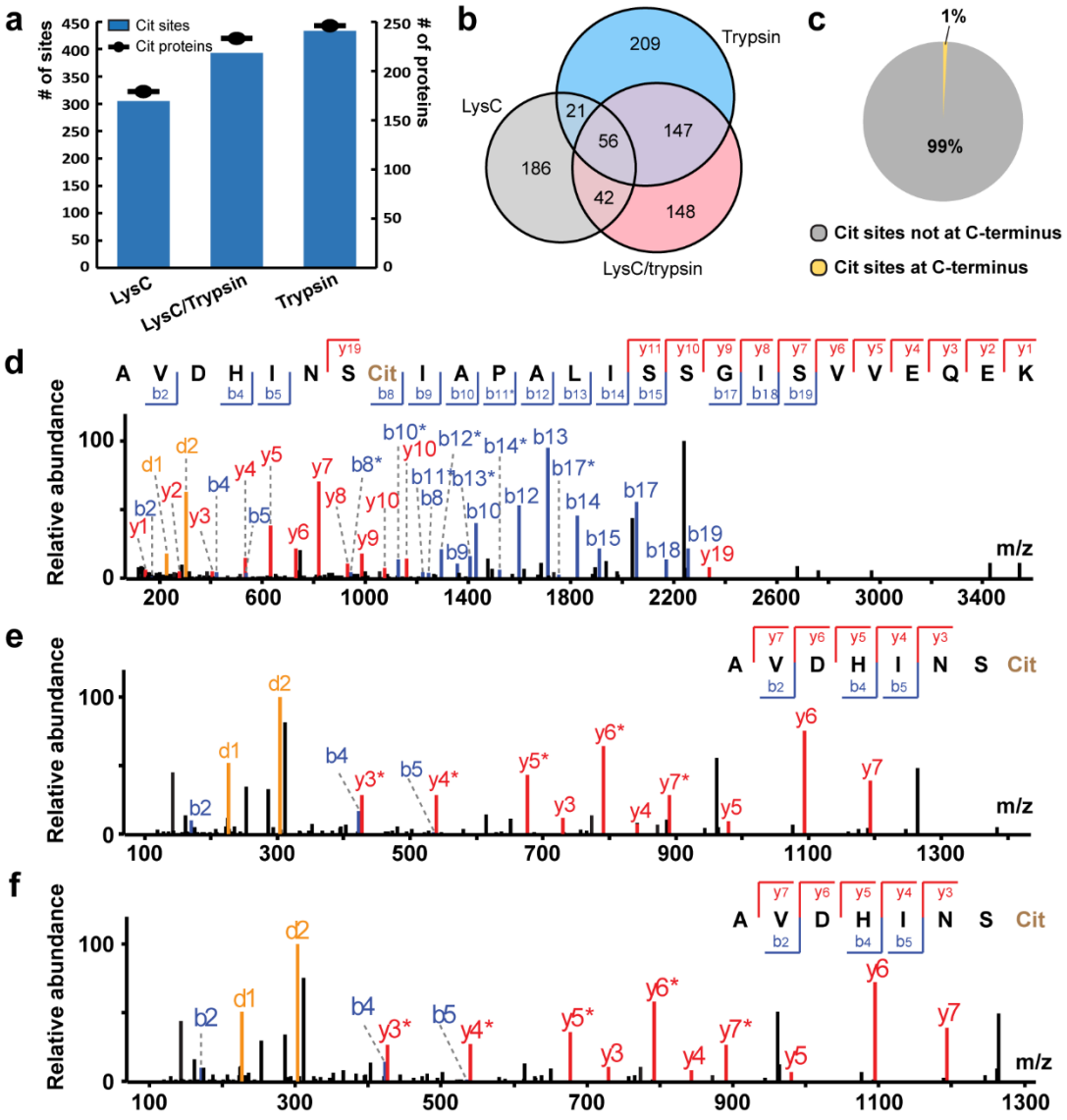

Figure S8. Optimization of different enzymatic digestion methods in detecting protein citrullination. a, Number of identified citrullination sites and citrullinated proteins with three different enzymatic digestion methods. Cit, citrullination. **b**, Venn diagram showing the overlap of citrullination sites identified with different digestion methods. **c**, Percentage of citrullination sites identified at or not at peptide C-termini in LysC digestion when searching with tryptic peptide parameters. **d-f**, Tandem MS spectra of the citrullinated peptides (R72Cit) on gamma-enolase protein identified in LysC (**d**), LysC/trypsin (**e**) and trypsin (**f**) digestion.

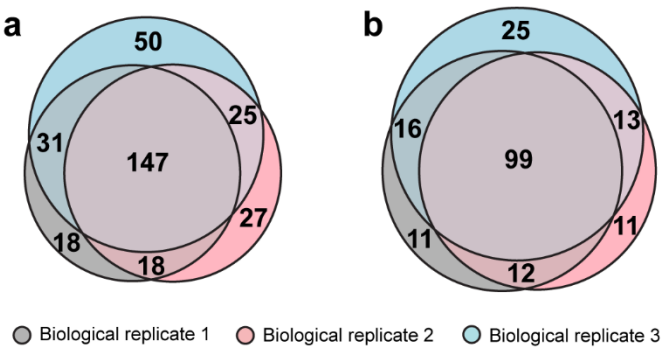

Figure S9. Evaluation of method reproducibility in citrullination analysis. a,**b**, Overlap of identified citrullination sites (**a**) and citrullinated proteins (**b**) of three biological replicates from mouse brain digest.

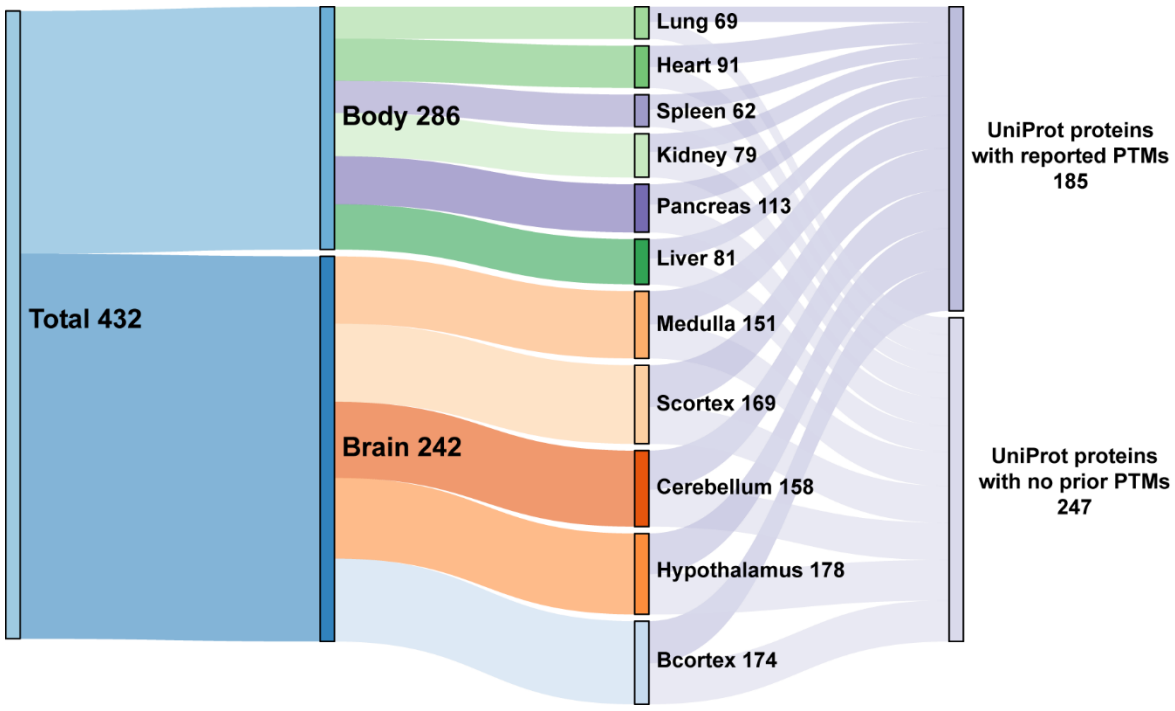

Figure S10. Tissue-specific distribution of identified citrullinated proteins in this study. Sankey diagram showing the distribution of identified citrullinated proteins in different tissues. Values indicate the number of proteins. Size of nodes of different tissues are not scaled to their values.

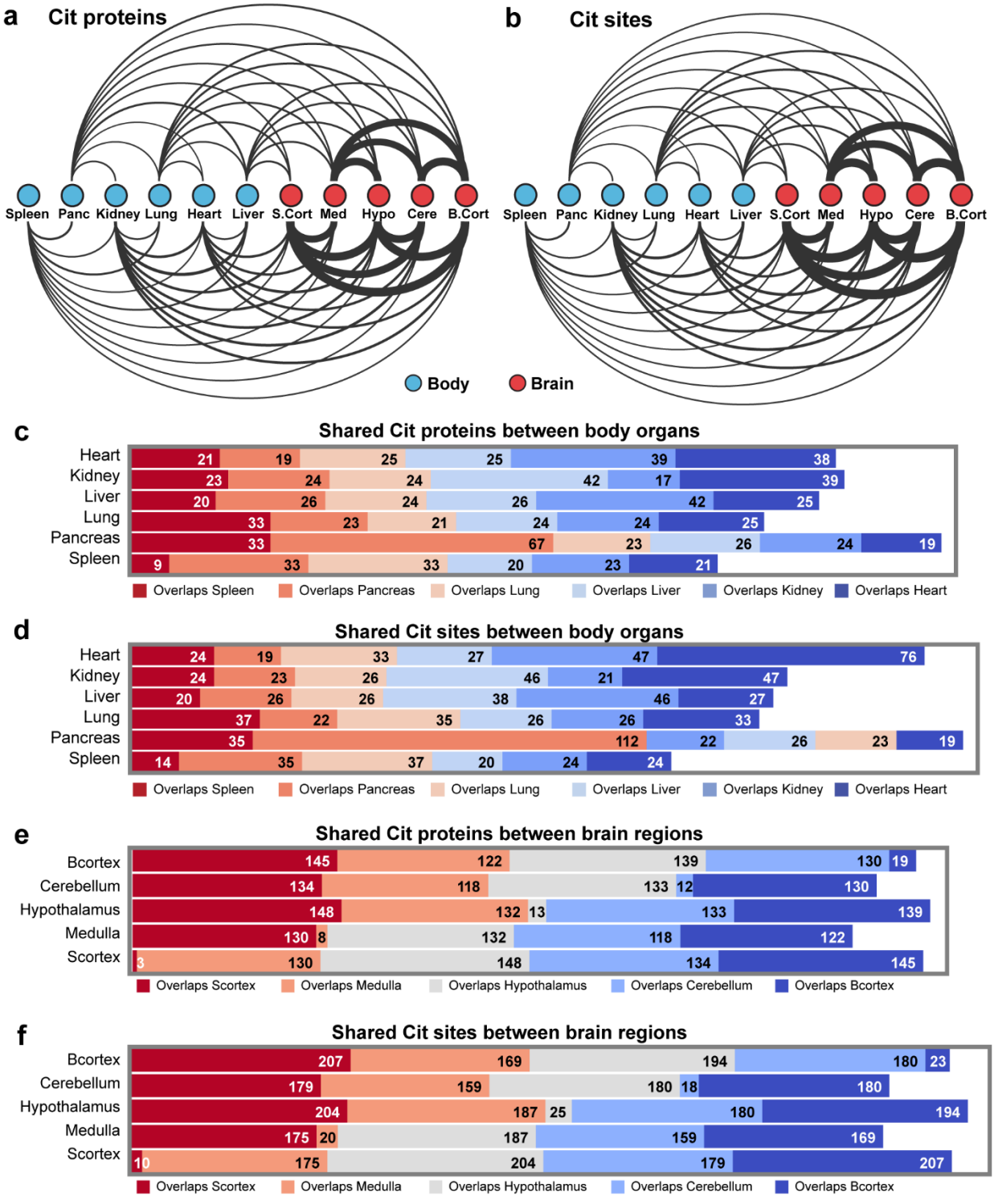
a
Cit proteins
b
Cit sites
Spleen Panc Kidney Lung Heart Liver S.Cort Med Hypo Cere B.Cort
Body
Brain
c
Shared Cit proteins between body organs
Heart 21 19 25 25 39 38
Kidney 23 24 24 42 17 39
Liver 20 26 24 26 42 25
Lung 33 23 21 24 24 25
Pancreas 33 67 23 26 24 19
Spleen 9 33 33 20 23 21
Overlaps Spleen
Overlaps Pancreas
Overlaps Lung
Overlaps Liver
Overlaps Kidney
Overlaps Heart
d
Shared Cit sites between body organs
Heart 24 19 33 27 47 76
Kidney 24 23 26 46 21 47
Liver 20 26 26 38 46 27
Lung 37 22 35 26 26 33
Pancreas 35 112 22 26 23 19
Spleen 14 35 37 20 24 24
Overlaps Spleen
Overlaps Pancreas
Overlaps Lung
Overlaps Liver
Overlaps Kidney
Overlaps Heart
e
Shared Cit proteins between brain regions
Bcortex 145 122 139 130 19
Cerebellum 134 118 133 12 130
Hypothalamus 148 132 13 133 139
Medulla 130 8 132 118 122
Scortex 3 130 148 134 145
Overlaps Scortex
Overlaps Medulla
Overlaps Hypothalamus
Overlaps Cerebellum
Overlaps Bcortex
f
Shared Cit sites between brain regions
Bcortex 207 169 194 180 23
Cerebellum 179 159 180 18 180
Hypothalamus 204 187 25 180 194
Medulla 175 20 187 159 169
Scortex 10 175 204 179 207
Overlaps Scortex
Overlaps Medulla
Overlaps Hypothalamus
Overlaps Cerebellum
Overlaps Bcortex

Figure S11. Overlap of identified citrullinated proteins and citrullination sites in different tissues. a,**b**, Arcplots showing the overlap of identified citrullinated proteins (**a**) and citrullination sites (**b**) between different tissues. Cit, citrullination; Panc, pancreas; Scort, hippocampus and thalamus; Med, medulla; Hypo, hypothalamus; Cere, cerebellum; Bcort, cerebral cortex. The width of ribbons connecting two tissues is proportional to the number of the overlap. **c**-**f**, Stacked bar graphs showing the number of shared citrullinated proteins and citrullination sites between different body organs (**c**, **d**) and brain regions (**e**, **f**). The overlap of one tissue with itself indicates the number of unique proteins or sites only identified in this tissue.

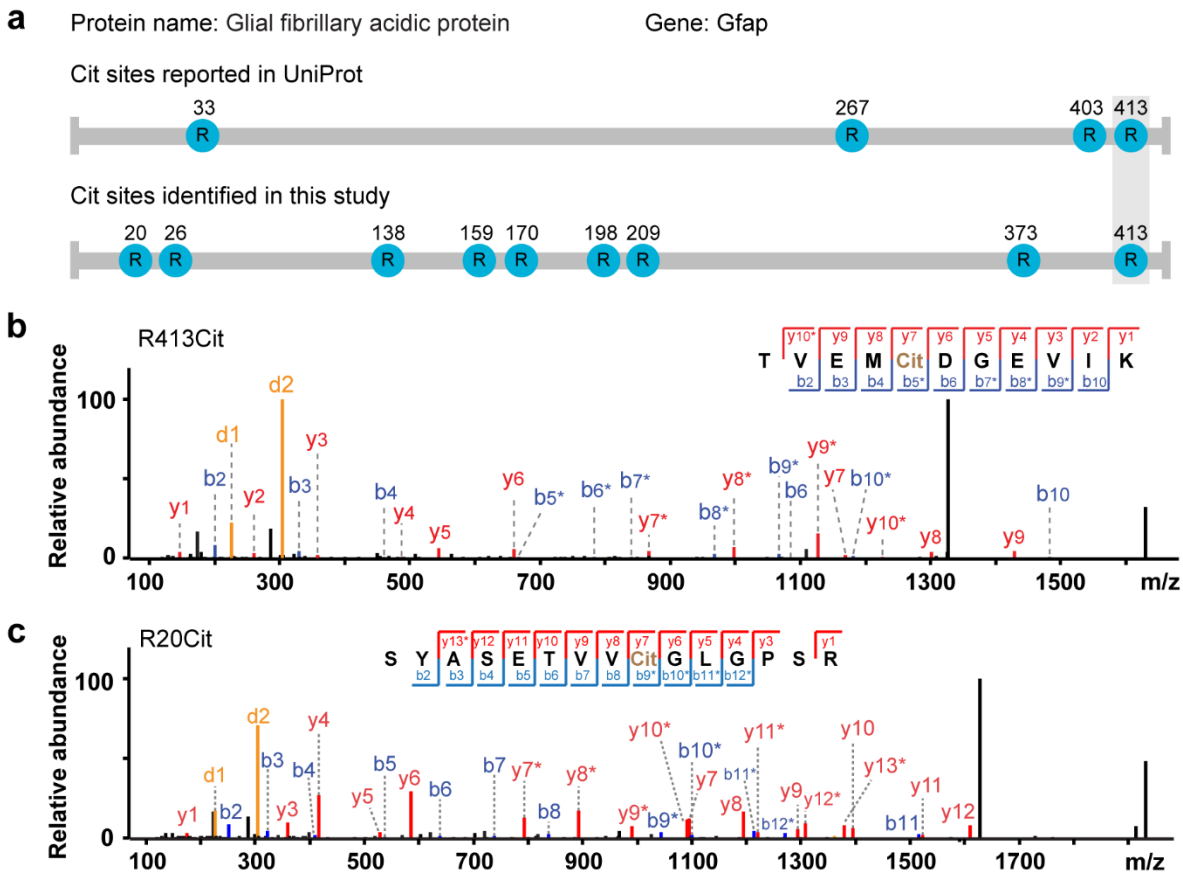

Figure S12. Examples of novel citrullination sites identified on glial fibrillary acidic protein in this study. a, Comparison of citrullination sites identified in this study and those reported in the UniProt database on glial fibrillary acidic protein. Cit, citrullination. **b**,**c**, Tandem MS spectra of two citrullinated peptides R413Cit (**b**) and R20Cit (**c**) on glial fibrillary acidic protein, with diagnostic ion pairs d1 and d2 highlighted.

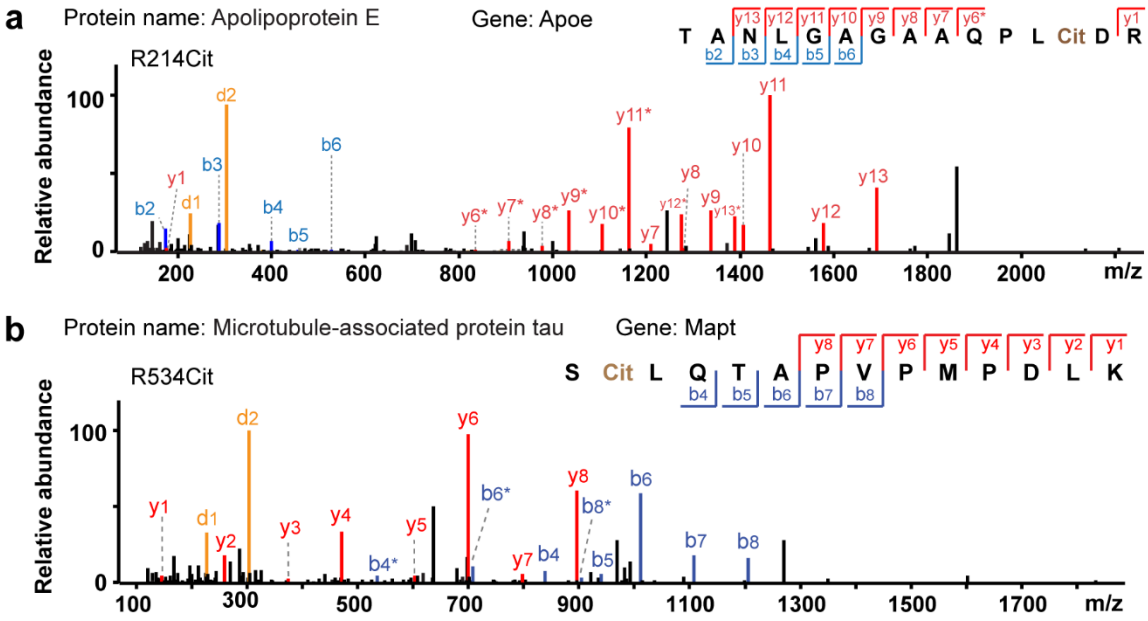

Figure S13. Examples of novel citrullinated proteins identified in this study. a, Tandem MS spectrum of one citrullinated peptide (R214Cit) on apolipoprotein E. **b**, Tandem MS spectrum of one citrullinated peptide (R534Cit) on microtubule-associated protein tau.

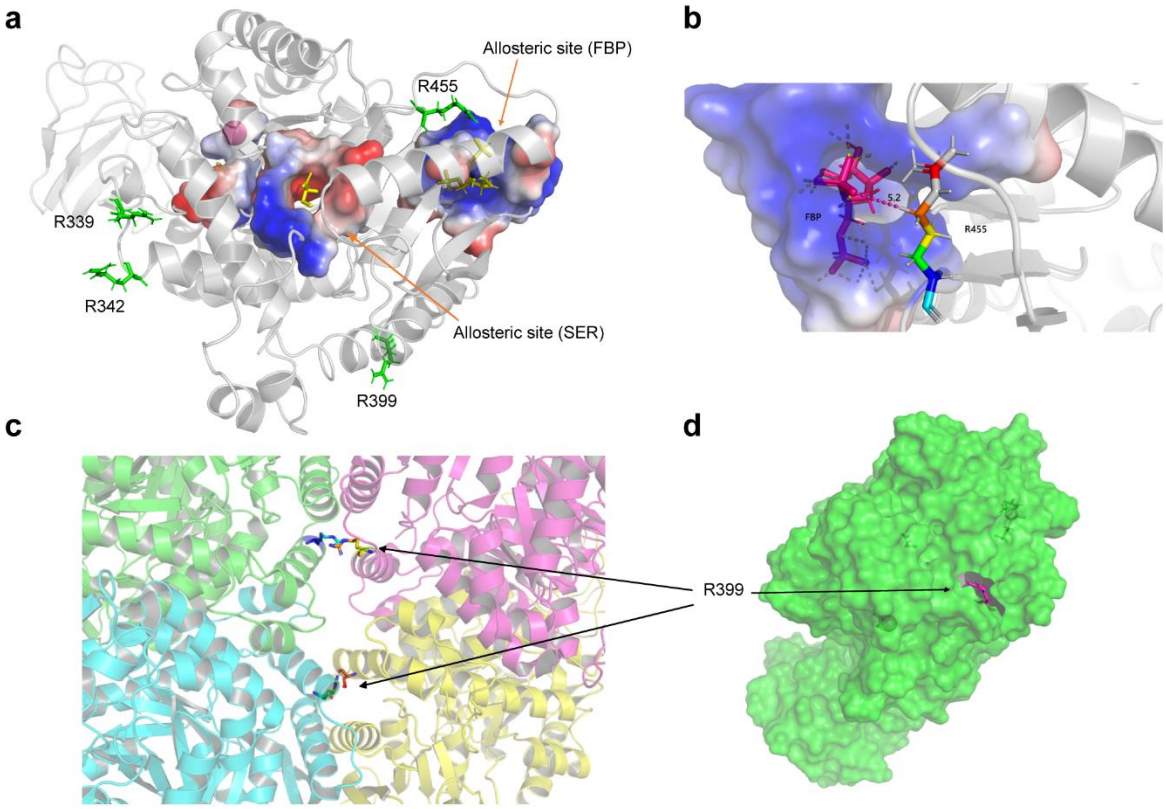

Figure S14. Homology modeling of mouse pyruvate kinase isozymes M2 (PKM2). a, Global view of citrullinated arginine sites of mouse PKM2. Green sticks represent the side chains of identified citrulline residues and yellow sticks represent given ligands of PKM2. Topological surface of PKM2 is generated based on protein contact potential of vacuum electrostatics using PyMOL. FBP, fructose 1,6-bisphosphate; SER, serine. **b**, Enlarged figure of allosteric binding pocket of FBP. Pink stick represents FBP; sticks with rainbow color represent R455. Pink dash line represents the distance from R455 to FBP (5.2 Å). **c**, Human PKM2 tetramer. Sticks with rainbow color represent R399. They are located at the interaction interface of the tetramer. **d**, Mouse PKM2 monomer. Pink stick represents the residue of R399 which is located in the tetramerization interface based on the homology model of human PKM2.

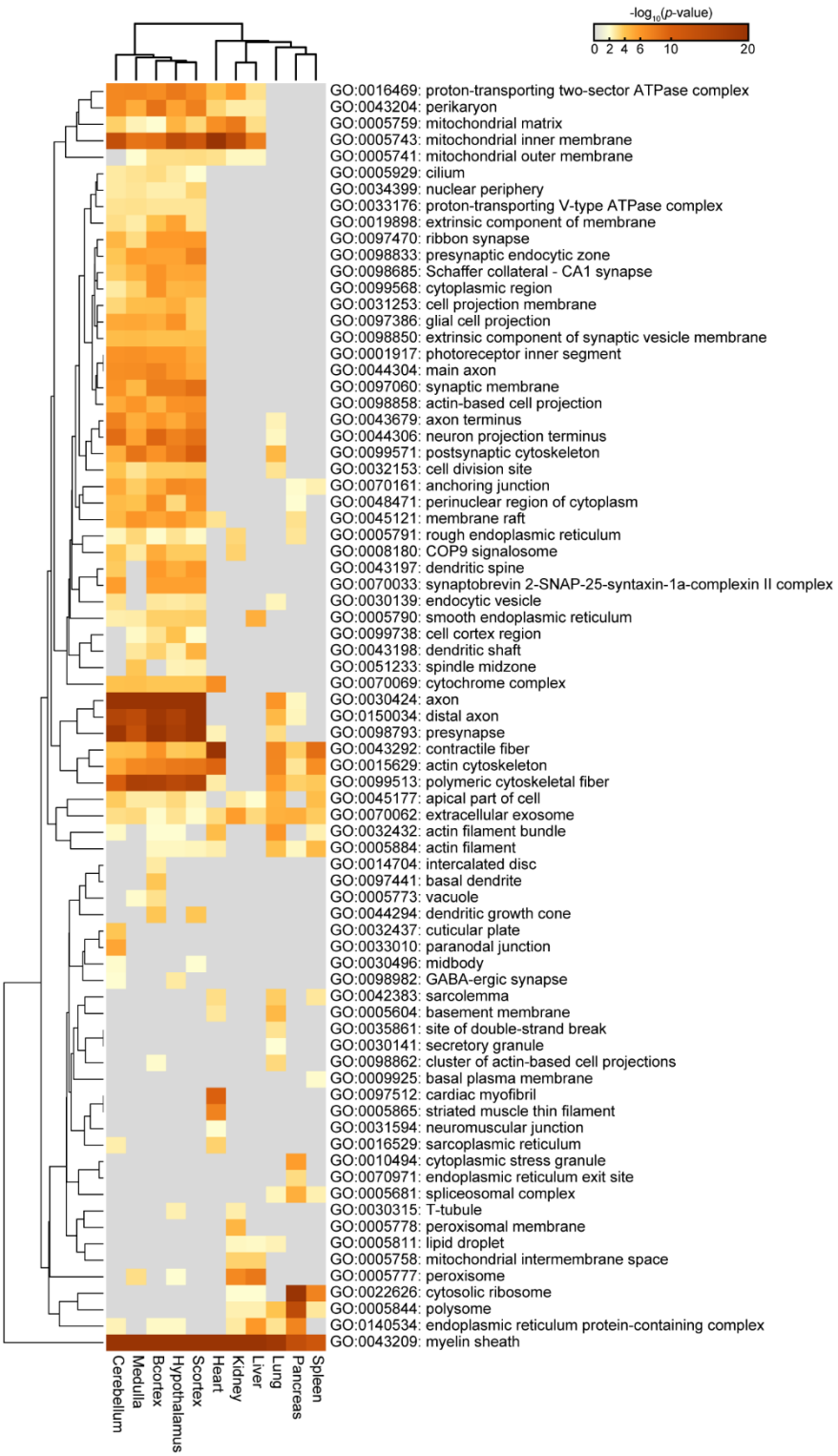

Figure S15. Enrichment of cellular component in different mouse tissues. Heatmaps generated using Metascape showing the significantly enriched (p value < 0.01 in at least one tissue region)

cellular components in different mouse tissues. The most significant 100 terms are shown. Color coding indicates $-\log_{10}$ (*p* values). Columns and rows are clustered based on their profile similarity.

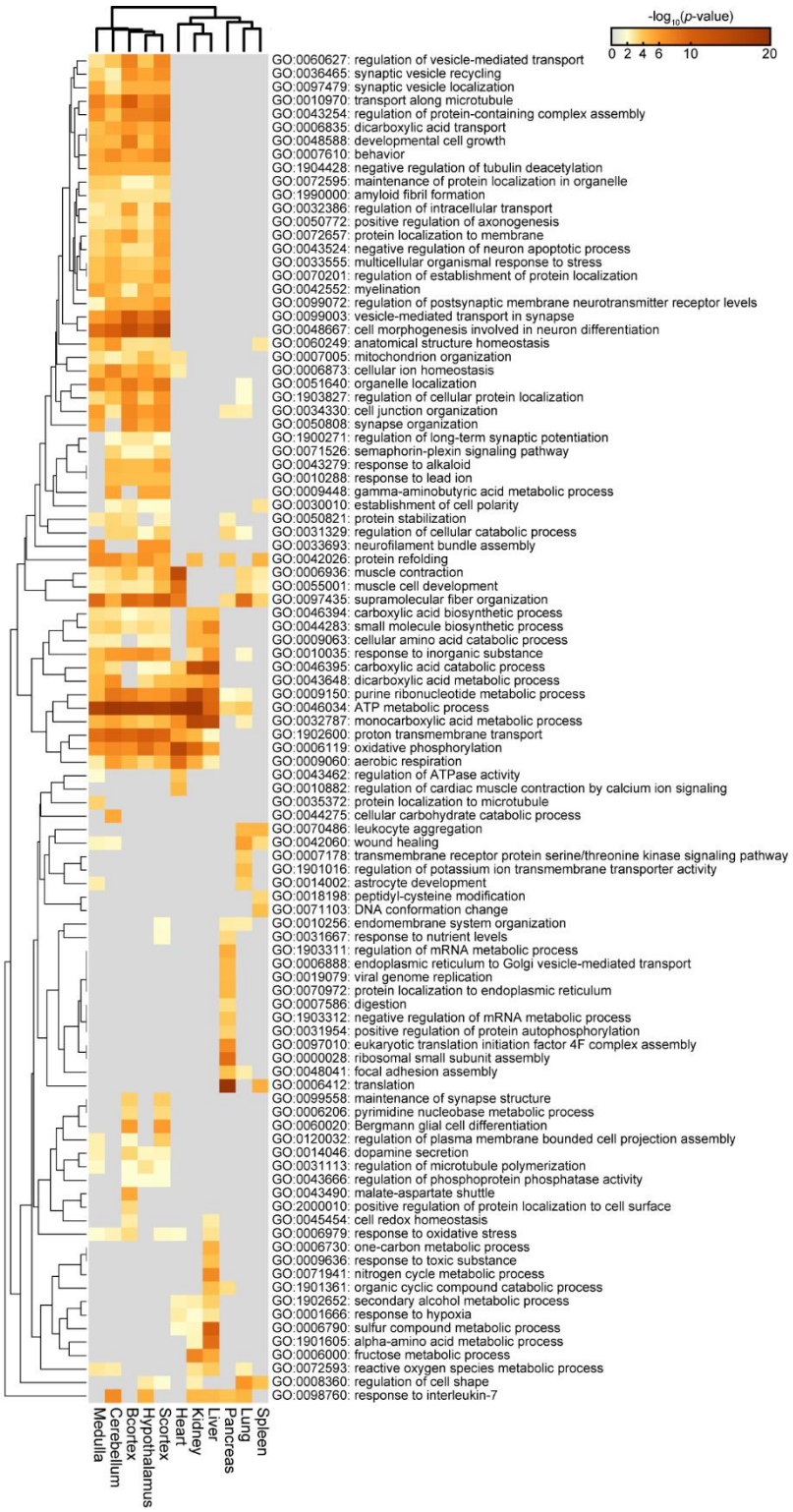

Figure S16. Enrichment of biological processes in different mouse tissues. Heatmaps generated using Metascape showing the significantly enriched (p value < 0.01 in at least one tissue region)

biological processes in different mouse tissues. The most significant 100 terms are shown. Color coding indicates $-\log_{10}$ (*p* values). Columns and rows are clustered based on their profile similarity.

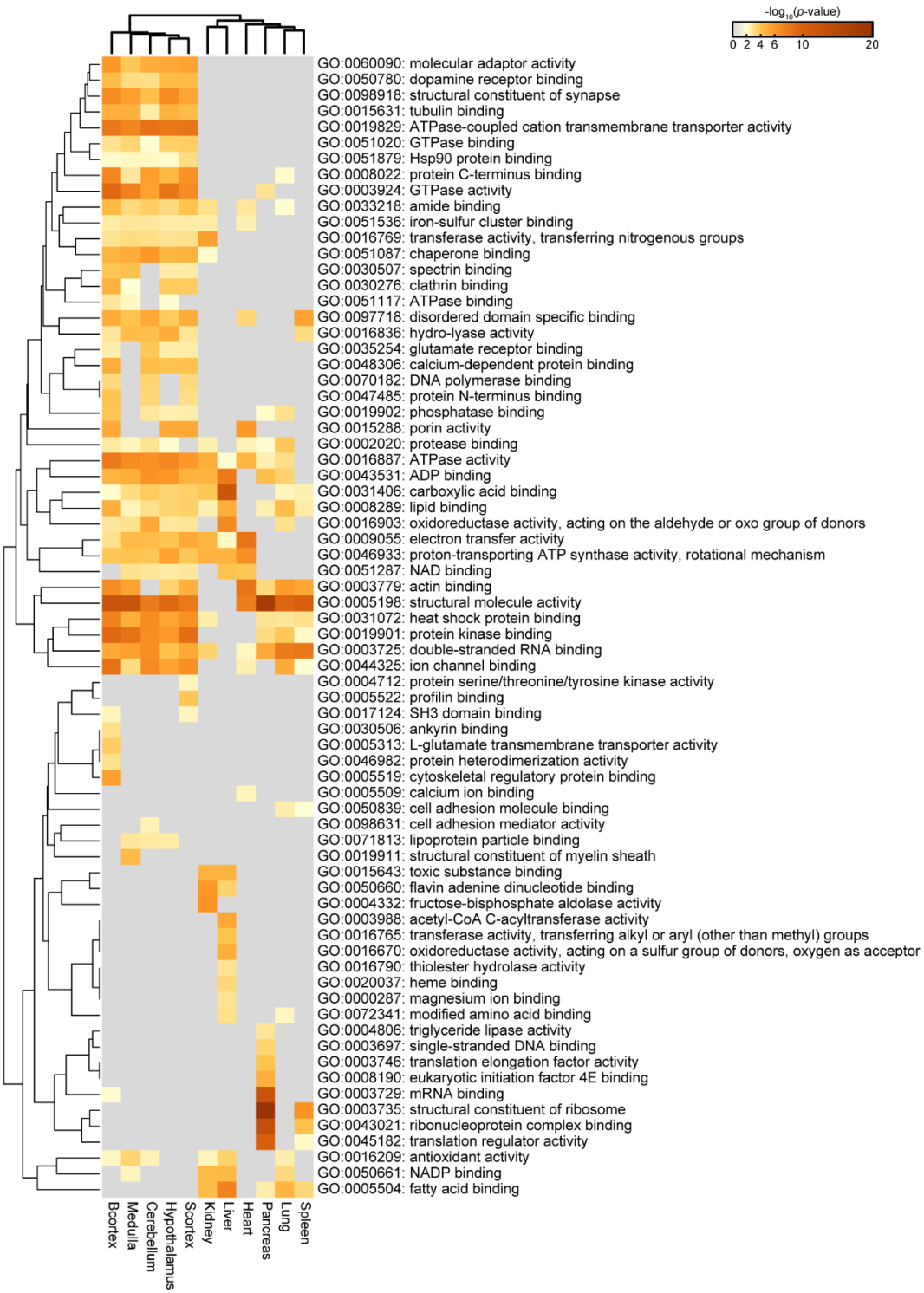

Figure S17. Enrichment of molecular function in different mouse tissues. Heatmaps generated using Metascape showing the significantly enriched (p value < 0.01 in at least one tissue region)

molecular functions in different mouse tissues. The most significant 100 terms are shown. Color coding indicates $-\log_{10}$ (*p* values). Columns and rows are clustered based on their profile similarity.

Figure S18. Enrichment of KEGG pathway in different mouse tissues. Heatmaps generated using Metascape showing the significantly enriched (p value < 0.01 in at least one tissue region) KEGG pathways in different mouse tissues. Color coding indicates -$\log_{10}$ (p values). Columns and rows are clustered based on their profile similarity.

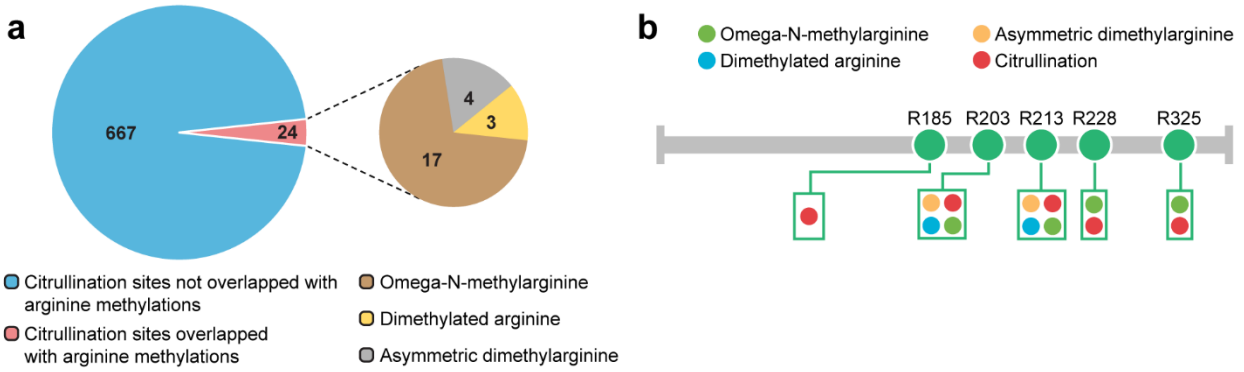

Figure S19. Colocalization of citrullination with arginine methylation. a, Some citrullination sites identified in this study are also reported as arginine methylation sites in the UniProt database. **b**, Colocalization of citrullination with arginine methylation on heterogeneous nuclear ribonucleoproteins A2/B1.

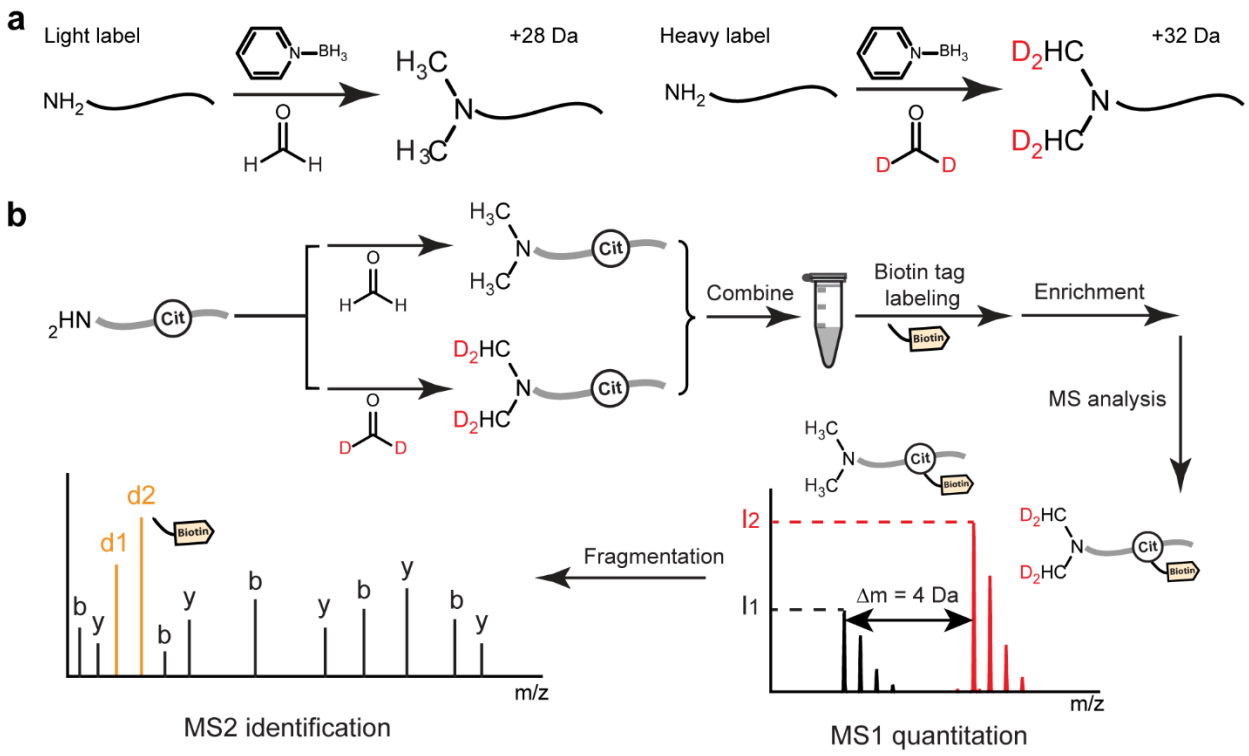

Figure S20. Principle of quantitative citrullination analysis using duplex isotopic dimethyl labeling strategy. a, Chemical reactions of duplex dimethylation which create a 4 Da mass difference between the light and heavy labels. **b**, Schematic showing the process of quantitative citrullination analysis using duplex dimethyl labeling and biotin thiol tag-assisted mass spectrometry.

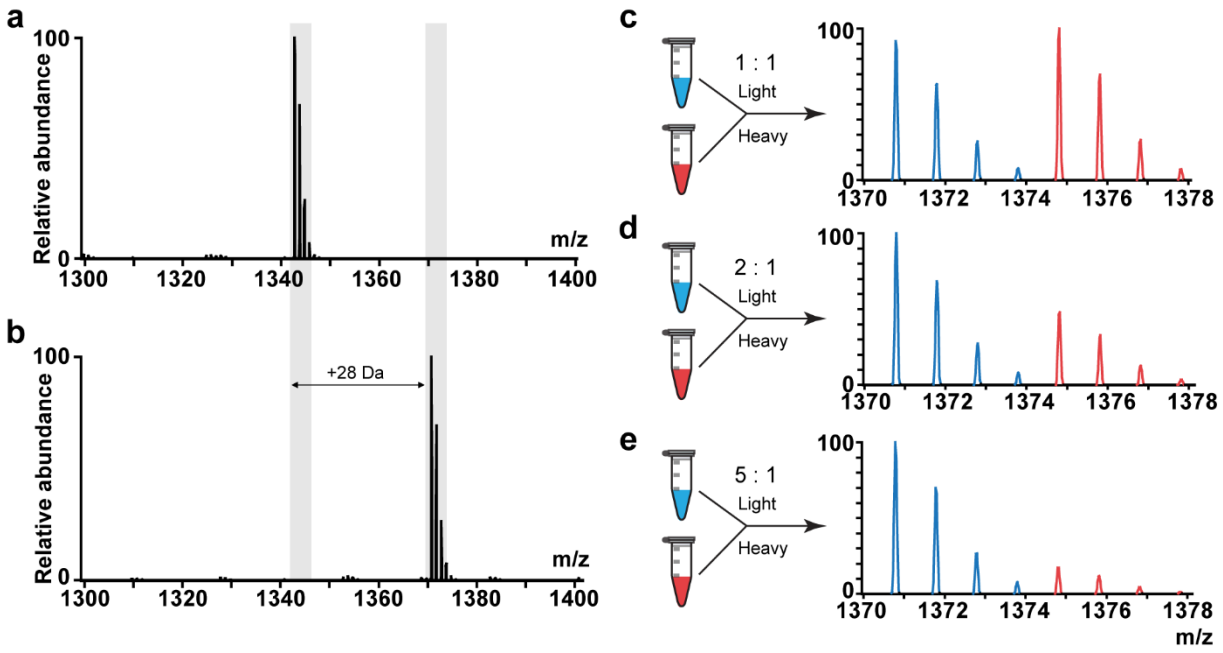

Figure S21. Test of quantitative citrullination analysis using citrullinated peptide standard. **a**, Spectrum of citrullinated peptide standard SAVRACitSSVPGVR. **b**, Spectrum of dimethylated citrullinated peptide standard SAVRACitSSVPGVR (light label, +28 Da). **c-e**, Spectra showing the labeling accuracy. Citrullinated peptide standard was dimethylated by either heavy isotope or light isotope labeling. Heavy and light labels were then mixed in three known ratios, 1:1 (**c**), 2:1 (**d**) and 5:1 (**e**).

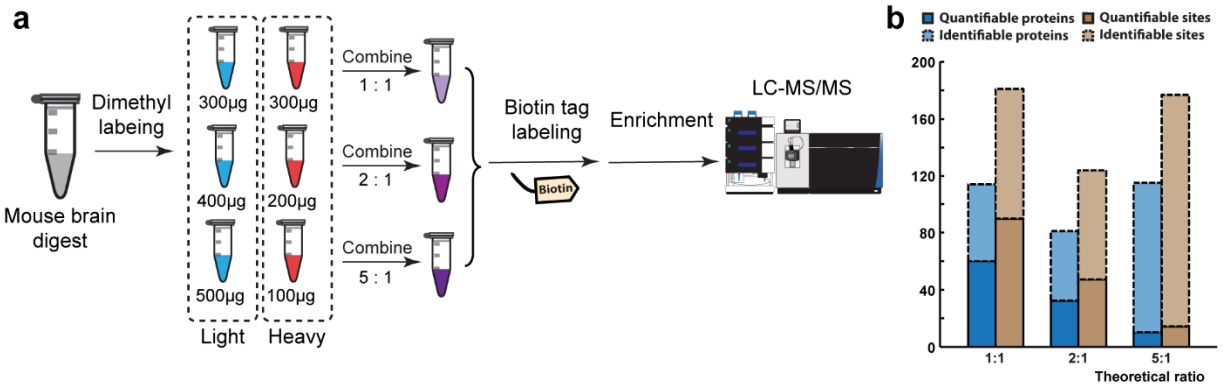

Figure S22. Test of quantitative citrullination analysis using mouse brain digest. a, Schematic showing the workflow of testing quantitative citrullination analysis using mouse brain digest. Digests with known peptide amount ratios (1:1, 2:1 and 5:1) were dimethylated by either light or heavy isotope labeling. Combined samples were subjected to biotin thiol tag derivatization, enrichment, and LC-MS/MS analysis. **b**, Number of quantifiable and identifiable citrullinated proteins and citrullination sites when testing quantitation accuracy and precision using mouse brain digest.

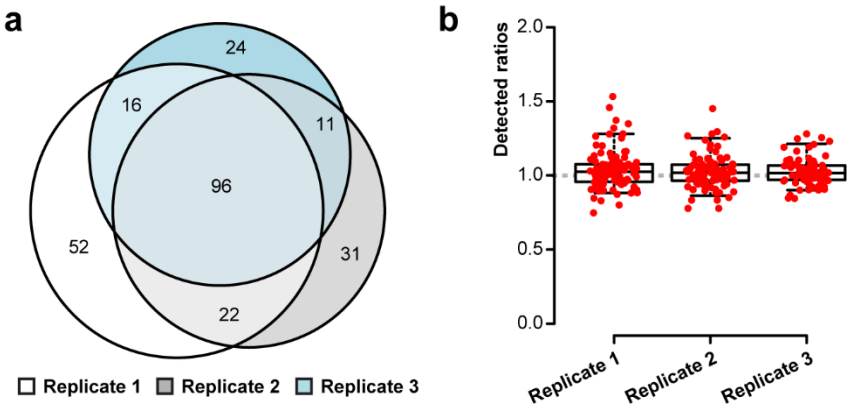

Figure S23. Test of reproducibility of quantitative citrullination analysis using dimethyl labeling. a, Venn diagram showing the overlap of identified citrullinated peptides in three biological replicates using duplex dimethyl labeling. Each replicate contains heavy and light isotope labeled samples mixed in 1:1 ratio. **b**, Boxplots showing reproducible quantitative results of three replicates. 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.

Table S1. List of all identified citrullinated proteins and sites from five brain regions and six body organs in mice. Citrullination localization probability, MaxQuant (Andromeda) searching score and peptide sequence for each site are also shown.

Protein	Position	Protein name	Localization probability	Andromeda score	Peptide sequence
P62259	19	14-3-3 protein epsilon	1.000	88.87	LAEQAER(1)YDEMVESMKK
P62259	141	14-3-3 protein epsilon	1.000	173.03	YLAEFATGNDR(1)K
P68254	138	14-3-3 protein theta	1.000	142.45	YLAEVACGDDR(1)K
Q9CQ62	119	2,4-dienoyl-CoA reductase, mitochondrial	1.000	171.72	CDVR(1)DPDMVHNTVLELIK
Q9CQX8	27	28S ribosomal protein S36, mitochondrial	0.849	99.09	VVQVVKPHAPLIKFPNR(0.849)R(0.151)
Q60597	323	2-oxoglutarate dehydrogenase, mitochondrial	1.000	98.16	LNVLANVIR(1)K
Q921H8	249	3-ketoacyl-CoA thiolase A, peroxisomal	1.000	101.45	TITVSQDEGVR(1)PSTTMQGLAK
Q8BWT1	181	3-ketoacyl-CoA thiolase, mitochondrial	0.990	98.54	YNISR(0.003)EDCDR(0.99)YALQSQQR(0.007)
Q8BWT1	224	3-ketoacyl-CoA thiolase, mitochondrial	1.000	134.68	QTMQVDEHAR(1)PQTTLEQLQK
P63325	119	40S ribosomal protein S10	0.999	106.38	FTR(0.999)GEADRDTYRR
P63325	124	40S ribosomal protein S10	0.999	106.93	FTRGEADR(0.999)DTYR(0.001)R
P63325	116	40S ribosomal protein S10	0.963	115.68	GPEGER(0.003)PAR(0.963)FTR(0.034)
P62281	132	40S ribosomal protein S11	1.000	134.68	DVQIGDIVTVGECR(1)PLSK
P63323	45	40S ribosomal protein S12	1.000	183.09	R(1)QAHLCVLASNCDEPMYVK
P62843	140	40S ribosomal protein S15	1.000	74.36	HGRPGIGATHSSR(1)FIPLK
P62852	111	40S ribosomal protein S25	1.000	113.11	HRAQVIYTR(1)NTK
Q6ZWY3	5	40S ribosomal protein S27-like	1.000	178.03	PLAR(1)DLLHPSLEEEKK
Q6ZWY3	23	40S ribosomal protein S27-like	1.000	58.08	R(1)LVQSPNSYFMDVK
P97461	198	40S ribosomal protein S5	1.000	159.12	KDELER(1)VAK
P62242	178	40S ribosomal protein S8	0.876	92.26	LLACIASR(0.876)PGQCGR(0.124)
P14206	117	40S ribosomal protein SA	0.994	158.52	FTPGTFTNQIQAAFR(0.994)EPR(0.006)
P61922	436	4-aminobutyrate aminotransferase, mitochondrial	1.000	107.50	GR(1)GTFCSFDTPDEAIR
P61922	285	4-aminobutyrate aminotransferase, mitochondrial	1.000	53.48	R(1)TVAGIIVEPIQSEGGDNHASDDFFRK
P61922	310	4-aminobutyrate aminotransferase, mitochondrial	1.000	105.71	RTVAGIIVEPIQSEGGDNHASDDFFR(1)K
P63038	309	60 kDa heat shock protein, mitochondrial	1.000	92.79	VGLQVVAVKAPGFGDNR(1)K
P99027	38	60S acidic ribosomal protein P2	1.000	218.88	KILDSVGIEADDDR(1)LNK
P35979	146	60S ribosomal protein L12	1.000	57.48	EILGTAQSVGCNVDGR(1)HPHDIIDDINSGAVECPAS
P47963	5	60S ribosomal protein L13	1.000	141.07	APSR(1)NGMILKPHFHK
P47963	60	60S ribosomal protein L13	0.995	99.06	IAPRPASGPIR(0.005)PIVR(0.995)CPTVR
Q9CPR4	135	60S ribosomal protein L17	1.000	76.29	AHGR(1)INPYMSSPCHIEMILTEK
P14115	65	60S ribosomal protein L27a	1.000	151.95	R(1)NQSFCPTVNLDK
P62889	56	60S ribosomal protein L30	1.000	162.93	LVILANNCPALR(1)K
P62900	92	60S ribosomal protein L31	1.000	148.41	R(1)NEDEDSPNK
O55142	22	60S ribosomal protein L35a	1.000	118.52	NQR(1)EHTALLK
Q9D8E6	33	60S ribosomal protein L4	1.000	90.46	APIR(1)PDIVNFVHTNLRK
Q9D8E6	45	60S ribosomal protein L4	1.000	183.01	APIRPDIVNFVHTNLR(1)K
Q9D8E6	49	60S ribosomal protein L4	1.000	221.85	NNR(1)QPYAVSELAGHQTSAESWGTGR
P47962	248	60S ribosomal protein L5	1.000	96.10	AHAAIR(1)ENPVYEK
P14148	258	60S ribosomal protein L7	1.000	271.66	TTHFVEGGDAGNR(1)EDQINR
P62918	123	60S ribosomal protein L8	1.000	117.15	AQLNIGNVLPVGTMPEGTIVCCLEEKPGDR(1)GK
P51410	129	60S ribosomal protein L9	1.000	78.40	MR(1)TGVACSVSQAQK
P20029	50	78 kDa glucose-regulated protein	1.000	159.26	NGR(1)VEIIANDQGNR
Q99KI0	555	Aconitate hydratase, mitochondrial	1.000	103.90	DSSGQR(1)VDVSPTSQR
Q99KI0	607	Aconitate hydratase, mitochondrial	1.000	94.77	FR(1)GHLDNISNNLLIGAINIENGK
Q99KI0	58	Aconitate hydratase, mitochondrial	1.000	128.76	R(1)LNRPLTLSEK
P62737	258	Actin, aortic smooth muscle	1.000	131.58	FR(1)CPETLFQPSFIGMESAGIHETTYNSIMK
P63260	256	Actin, cytoplasmic 2	1.000	102.72	FR(1)CPEALFQPSFLGMESCGIHETTFNSIMK
P63260	95	Actin, cytoplasmic 2	1.000	193.06	IWHHTFYNELR(1)VAPEEHPVLLTEAPLNPK
Q8K0L3	105	Acyl-coenzyme A synthetase ACSM2, mitochondrial	0.978	134.29	QTANVLSGACGLHR(0.978)GDR(0.022)
Q9QYR9	137	Acyl-coenzyme A thioesterase 2, mitochondrial	1.000	49.77	R(1)DVQTPFLVELEVLDGHEPDGGQR
Q9WTP6	186	Adenylate kinase 2, mitochondrial	1.000	103.31	TR(1)LEAYHTQTTPLVEYYRK
Q9WUR9	188	Adenylate kinase 4, mitochondrial	1.000	84.25	SR(1)GVLHQFSGTETNR
Q9R0Y5	149	Adenylate kinase isoenzyme 1	1.000	66.15	R(1)LETYYNATEPVISFYDK
P48962	259	ADP/ATP translocase 1	1.000	157.98	GADIMYTGTLDCWR(1)K
P48962	152	ADP/ATP translocase 1	1.000	146.60	GSSQR(1)EFNGLGDCLTK
Q9D8S3	399	ADP-ribosylation factor GTPase-activating protein 3	1.000	156.73	DSSRDPEPAMR(1)STGSSDRPSAR
Q9D8S3	313	ADP-ribosylation factor GTPase-activating protein 3	1.000	101.34	LGMGFGSCR(1)SGISHSVTSDMQTIEQESPTLAKPR
Q9WUL7	99	ADP-ribosylation factor-like protein 3	1.000	58.09	R(1)FEETGQELTELLEEEK
Q62151	115	Advanced glycosylation end product-specific receptor	1.000	54.02	VR(1)VYQIPGKPEIVDPASELTASVPNK
O54931	263	A-kinase anchor protein 2	1.000	76.38	TIEEQLDEEHLESHR(1)K
O35945	476	Aldehyde dehydrogenase, cytosolic 1	1.000	152.88	MSGNGR(1)ELGEHGLYEYTELK
P45376	70	Aldose reductase	1.000	126.32	R(1)QDLFIVSK
Q8R4G6	56	Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase A	1.000	56.68	EQILDLSKR(1)YIKALAEENR
Q7TPR4	614	Alpha-actinin-1	1.000	115.04	R(1)DQALTEEHAR
Q9JI91	353	Alpha-actinin-2	0.961	81.02	LR(0.961)ISNR(0.039)PAFMPSEGK
O88990	73	Alpha-actinin-3	1.000	198.24	AGTQIENIEEDFR(1)NGLK
O88990	59	Alpha-actinin-3	1.000	82.65	TFTAWCNSHLR(1)K
Q9QYC0	507	Alpha-adducin	1.000	156.24	IR(1)EQNLQDIK
P23927	107	Alpha-crystallin B chain	1.000	156.10	HEER(1)QDEHGFISR
P17182	50	Alpha-enolase	1.000	58.09	AAVPSGASTGIYEALELR(1)DNDK
P46660	16	Alpha-internexin	1.000	136.68	SFGSEHYLCSASSYR(1)K
P97449	218	Aminopeptidase N	1.000	58.98	VVATTQMQAADAR(1)K
Q8VHG2	554	Angiomotin	1.000	64.45	HIEIR(1)DQALSNAQAK

P16460	95	Argininosuccinate synthase	0.977	183.23	YLLGTSLAR(0.977)PCIAR(0.023)
Q61024	550	Asparagine synthetase [glutamine-hydrolyzing]	1.000	109.37	WINATDPSAR(1)TLTHYK
P05202	68	Aspartate aminotransferase, mitochondrial	1.000	114.24	MNLGVGAYR(1)DDNGKPYVLPSVR
Q9CQQ7	196	ATP synthase F(0) complex subunit B1, mitochondrial	1.000	76.59	NR(1)LDYHISVQNMMR
Q03265	405	ATP synthase subunit alpha, mitochondrial	1.000	84.07	GIR(1)PAINVGLSVSR
Q03265	186	ATP synthase subunit alpha, mitochondrial	1.000	103.60	ISVR(1)EPMQTGIK
Q03265	231	ATP synthase subunit alpha, mitochondrial	1.000	166.60	R(1)FNDGTDEK
P56480	241	ATP synthase subunit beta, mitochondrial	1.000	235.69	TR(1)EGNDLYHEMIESGVINLK
P97450	54	ATP synthase-coupling factor 6, mitochondrial	1.000	117.79	R(1)QASGGPVDIGPEYQQDLDRELYK
P97450	73	ATP synthase-coupling factor 6, mitochondrial	1.000	144.43	RQASGGPVDIGPEYQQDLDR(1)ELYK
P47857	108	ATP-dependent 6-phosphofructokinase, muscle type	1.000	59.54	R(1)GITNLCVIGGDGSLTGADTFR
Q62167	622	ATP-dependent RNA helicase DDX3X	1.000	89.47	ASSSR(1)SGGGGHGGSR
Q62167	75	ATP-dependent RNA helicase DDX3X	0.998	149.19	DAYSSFGSR(0.998)GDSR(0.002)
Q62167	99	ATP-dependent RNA helicase DDX3X	0.978	135.76	GRFDDR(0.978)GR(0.021)GDYDGIGGR
Q62167	110	ATP-dependent RNA helicase DDX3X	0.998	109.65	GRGDYDGIGGR(0.998)GDR(0.002)
Q62167	113	ATP-dependent RNA helicase DDX3X	1.000	156.88	GRGDYDGIGGRGDR(1)SGFGK
Q62167	600	ATP-dependent RNA helicase DDX3X	0.999	122.16	FSGGFGAR(0.999)DYR(0.001)
O70318	163	Band 4.1-like protein 2	1.000	86.50	SSVETQPAEEVR(1)K
Q9WV92	749	Band 4.1-like protein 3	1.000	70.57	EQR(1)GEEVDQSAPEQEQPATVSHEEEQASTIR
Q9WV92	579	Band 4.1-like protein 3	1.000	47.08	R(1)TFLETSTETALTNEWEK
Q9R069	224	Basal cell adhesion molecule	1.000	139.19	DDR(1)DANFHCAAHYDLPSGQHGR
Q05793	812	Basement membrane-specific heparan sulfate proteoglycan core protein	0.872	66.47	ATATACR(0.872)PCPCPYIDASR(0.064)R(0.064)
Q9QYB8	370	Beta-adducin	0.751	81.36	SRLGEHEFEALMR(0.751)MLDNLGYR(0.249)
O35490	361	Betaine--homocysteine S-methyltransferase 1	0.994	64.27	IASGR(0.994)PYNPSMSR(0.006)PDAWGVTK
O35490	318	Betaine--homocysteine S-methyltransferase 1	1.000	61.48	AIAEELAPER(1)GFLPPASEK
Q8VC30	152	Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)	1.000	121.18	R(1)GLCGTVLIHK
P34914	103	Bifunctional epoxide hydrolase 2	1.000	86.94	SINR(1)PMLQAAIALKK
Q64285	474	Bile salt-activated lipase	0.935	118.27	WMGADHADDLQYVFGKPFATPLGYR(0.935)PQDR(0.065)
Q64285	67	Bile salt-activated lipase	1.000	126.02	TLENPQR(1)HPGWQGTLK
Q80YN3	613	Breast carcinoma-amplified sequence 1 homolog	1.000	132.36	R(1)MSDAQVQTDPVSIGPVGK
Q6WVG3	207	BTB/POZ domain-containing protein KCTD12	0.800	64.50	SPSGGAAGPLLTPSQSLDGSR(0.8)R(0.2)
P12658	68	Calbindin	1.000	165.16	SFVDQYGQR(1)DDGK
P28652	507	Calcium/calmodulin-dependent protein kinase type II subunit beta	0.992	235.22	LTQYIDGQGR(0.992)PR(0.008)
Q8BH59	565	Calcium-binding mitochondrial carrier protein Aralar1	1.000	108.45	AGQTTYSGVVDCFR(1)K
P0DP28	87	Calmodulin-3	0.971	63.29	DTDSEEEIR(0.971)EAFR(0.029)VFDKDGNGYISAAELR
P0DP28	91	Calmodulin-3	1.000	93.63	EAFR(1)VFDKDGNGYISAAELR
Q80VC9	794	Calmodulin-regulated spectrin-associated protein 3	1.000	81.97	LAPLTR(1)VLTPPHDVDSLPHLRK
O08529	77	Calpain-2 catalytic subunit	1.000	52.81	R(1)PTEICADPQFIIGGATR
Q08331	244	Calretinin	1.000	98.28	KEMNIQQLTTYR(1)K
O09161	48	Calsequestrin-2	1.000	128.88	R(1)YDLLCLYYHEPVSSDK
Q3UHL1	23	CaM kinase-like vesicle-associated protein	1.000	88.87	NYNQPSEVTDR(1)YDLGQVIK
Q8C196	112	Carbamoyl-phosphate synthase [ammonia], mitochondrial	1.000	245.87	GQILTMANPIIGNGGAPDTTAR(1)DELGLNK
Q8C196	286	Carbamoyl-phosphate synthase [ammonia], mitochondrial	1.000	51.47	ILESDR(1)KEPLFGISTGNIITGLAAGAK
Q8C196	1292	Carbamoyl-phosphate synthase [ammonia], mitochondrial	1.000	59.37	R(1)LPTLEQPIIPSDYVAIK
P13634	247	Carbonic anhydrase 1	0.968	154.26	GLLSSAEGEPAVPVLSNHR(0.968)PPQPLKGR(0.032)
P00920	245	Carbonic anhydrase 2	0.982	94.35	TLNFNEEGDAEEAMVDNWR(0.982)PAQPLKNR(0.018)K
P16015	245	Carbonic anhydrase 3	1.000	206.93	LRSLFSSAENEPPVPLVGNWR(1)PPQPVK
P28651	276	Carbonic anhydrase-related protein	0.946	88.40	GAELVEGCDGILGDNFR(0.946)PTQPLSDR(0.054)
P08074	173	Carbonyl reductase [NADPH] 2	1.000	145.77	IR(1)VNSVNPTVVLTDMGK
Q91WU0	458	Carboxylesterase 1F	1.000	165.13	GHRDTGAPTYMYEYQYYPSFSSPQR(1)PK
Q7TPZ8	169	Carboxypeptidase A1	1.000	82.86	FSTGGTNR(1)PAIWIDTGIHSR
Q7TPZ8	155	Carboxypeptidase A1	1.000	121.26	IQIGSTFEGR(1)PINVLK
Q6P9K8	934	Caskin-1	0.999	77.07	NVNR(0.999)SQSFAVR(0.001)PR
P24270	23	Catalase	1.000	126.96	ASQR(1)PDVLTTGGGNPIGDK
P24270	431	Catalase	1.000	162.77	R(1)FNSANEDNVTQVR
P24270	458	Catalase	1.000	131.56	R(1)LCENIAGHLK
P30999	233	Catenin delta-1	0.945	158.61	HYEDGYPGGSDNYGSLSR(0.945)VTR(0.055)
P10605	314	Cathepsin B	1.000	76.21	ILR(1)GENHCGIESEIVAGIPR
P06797	318	Cathepsin L1	1.000	143.40	DR(1)DNHCGLATAASYPVVN
P49817	19	Caveolin-1	1.000	145.65	YVDSEGHLYTVPIR(1)EQGNIYKPNNK
O35566	133	CD151 antigen	1.000	71.85	R(1)YHQSGHEGVSSAVDK
Q9JM96	35	Cdc42 effector protein 4	0.999	128.73	ADLTAEMISAPLGDFR(0.999)HTMHVGR(0.001)
P53996	161	Cellular nucleic acid-binding protein	0.999	67.76	TSEVNCYR(0.999)CGESGHLAR(0.001)
Q8BXK9	109	Chloride intracellular channel protein 5	1.000	144.72	HR(1)ESNTAGIDIFSK
Q68FD5	437	Clathrin heavy chain 1	1.000	170.61	YESLELCR(1)PVLQQGRK
O08585	128	Clathrin light chain A	1.000	130.20	LEALDANSR(1)K
Q6IRU5	113	Clathrin light chain B	1.000	147.56	R(1)LQELDAASK
Q9QZE5	590	Coatomer subunit gamma-1	1.000	67.55	SVPLATTPMAEQR(1)PESTATAAVKQPEK
Q9CQC2	44	Colipase	1.000	141.60	SR(1)CCQHDTILGIAR
P11087	1055	Collagen alpha-1(I) chain	1.000	149.04	NGDR(1)GETGPAGPAGPIGPAGAR
P02463	1495	Collagen alpha-1(IV) chain	1.000	72.77	AHGQDLGTAGSCLR(1)K
P84086	96	Complexin-2	1.000	151.04	AALEQPCEGSLTR(1)PK
Q8BLR2	509	Copine-4	1.000	105.03	DIVQFVPFR(1)NFK
Q04447	96	Creatine kinase B-type	1.000	149.99	DLFDPIIEER(1)HGGYQPSDEHK
Q04447	13	Creatine kinase B-type	1.000	187.21	LR(1)FPAEDEFPDLSSHNNHMAK
Q04447	320	Creatine kinase B-type	1.000	48.50	R(1)GTGGVDTAAVGGVFDVSNADR
P07310	96	Creatine kinase M-type	1.000	140.82	DLFDPIIQDR(1)HGGYKPTDK
P21460	44	Cystatin-C	0.837	77.08	MLGAPEEADANEEGVR(0.837)R(0.163)
P97315	99	Cysteine and glycine-rich protein 1	1.000	100.28	HEEAPGHR(1)PTTNPNASK
P50462	41	Cysteine and glycine-rich protein 3	1.000	134.13	TCFHCMACR(1)K
P56395	52	Cytochrome b5	1.000	120.34	FLEEHPGGEEVLR(1)EQAGGDATENFEDVGHSTDAR
Q9DB77	70	Cytochrome b-c1 complex subunit 2, mitochondrial	1.000	83.40	AGSR(1)YEDSNNLGTSHLLR
Q9DB77	301	Cytochrome b-c1 complex subunit 2, mitochondrial	1.000	167.99	R(1)GNNTTSLLSQSVAK
Q9CR68	196	Cytochrome b-c1 complex subunit Rieske, mitochondrial	1.000	187.29	KEIDQEAAVEVSQLR(1)DPQHDLDR
P12787	127	Cytochrome c oxidase subunit 5A, mitochondrial	1.000	61.10	AGPHKEIYPYVIQELR(1)PTLNELGISTPEELGLDKV

P12787	94	Cytochrome c oxidase subunit 5A, mitochondrial	1.000	153.81	R(1)LNDFASAVR
P56391	20	Cytochrome c oxidase subunit 6B1	1.000	102.89	TAPFDSR(1)FPNQNQTK
Q9D0M3	167	Cytochrome c1, heme protein, mitochondrial	1.000	95.20	ALAEEVEVQDGPNDDGEMFMR(1)PGK
P33267	250	Cytochrome P450 2F2	1.000	89.37	SVR(1)EHQDSLDPNSPR
Q9JHU4	3217	Cytoplasmic dynein 1 heavy chain 1	0.872	76.82	IKETVDQVEELR(0.872)R(0.128)
Q9JHU4	1270	Cytoplasmic dynein 1 heavy chain 1	1.000	59.54	TKPVTGNLR(1)PEEALQALTIYEGK
Q8R0Y6	742	Cytosolic 10-formyltetrahydrofolate dehydrogenase	0.995	51.81	IGNPLDR(0.995)DTNHGPQNHEAHLR(0.005)K
Q8R0Y6	756	Cytosolic 10-formyltetrahydrofolate dehydrogenase	1.000	68.66	IGNPLDRDTNHGPQNHEAHLR(1)K
Q8R0Y6	551	Cytosolic 10-formyltetrahydrofolate dehydrogenase	0.854	98.06	IQGATIPINQAR(0.854)PNR(0.146)
Q8R0Y6	191	Cytosolic 10-formyltetrahydrofolate dehydrogenase	1.000	164.01	R(1)PQPEEGATYEGIQK
P18894	284	D-amino-acid oxidase	0.838	120.90	IVGELTGFR(0.155)PVR(0.838)PQVR(0.007)
Q91XC8	25	Death-associated protein 1	1.000	122.47	AGGMR(1)IVQKHPHTGDGK
Q91XC8	93	Death-associated protein 1	1.000	144.97	HVSPR(1)TQHIQQPR
O35459	73	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	1.000	139.58	HVLHVQLNR(1)PEK
P31001	179	Desmin	0.998	83.42	ARVDVER(0.998)DNLIDDLQR(0.002)
P31001	200	Desmin	1.000	86.80	LQEEIQLR(1)EEAENNLAAFR
P97427	552	Dihydropyrimidinase-related protein 1	0.837	123.44	NLHQSNFSLSGAQIDDNNPR(0.837)R(0.163)
O08553	227	Dihydropyrimidinase-related protein 2	1.000	84.15	ILDLGITGPEGHVLSR(1)PEEVEAEAVNR
O08553	557	Dihydropyrimidinase-related protein 2	0.992	117.18	R(0.008)TTQR(0.992)IVAPPGGR
O08553	268	Dihydropyrimidinase-related protein 2	1.000	179.35	SAAEVIAQAR(1)K
P50285	280	Dimethylaniline monooxygenase [N-oxide-forming] 1	1.000	88.56	TQLR(1)EPVLNDELPGR
P97872	277	Dimethylaniline monooxygenase [N-oxide-forming] 5	1.000	116.45	HR(1)ALSQHPTVNDDLPNR
Q9R1V6	827	Disintegrin and metalloproteinase domain-containing protein 22	0.989	122.64	HISDICENGR(0.989)PR(0.011)
Q99KV1	56	DnaJ homolog subfamily B member 11	1.000	80.79	LALQLHPDR(1)NPDDPQAQEK
Q9QXS6	384	Drebrin	1.000	51.47	MAPTPIPTR(1)SPSDSSTASTPIAEQIER
P31938	47	Dual specificity mitogen-activated protein kinase kinase 1	1.000	144.29	LEELELDEQQR(1)K
Q99KJ8	213	Dynactin subunit 2	1.000	98.66	ATAGAPPDSSLVTYELHSR(1)PEQDK
Q99KJ8	78	Dynactin subunit 2	1.000	52.81	R(1)TGYESGDYEMLGEGLGVK
P39053	343	Dynamin-1	1.000	159.46	R(1)IEGSGDQIDTYELSGGAR
Q8K1M6	61	Dynamin-1-like protein	1.000	67.42	R(1)PLILQLVHVSPEDK
P11531	2669	Dystrophin	1.000	67.65	R(1)VSEQEAALEETHR
E9PVA8	962	eIF-2-alpha kinase activator GCN1	1.000	79.47	AVSLLHTHTIPSR(1)VGK
P10126	430	Elongation factor 1-alpha 1	1.000	164.53	DMR(1)QTVAVGVIK
P58252	767	Elongation factor 2	1.000	127.96	R(1)GHVFEESQVAGTPMFVVK
Q8BFR5	53	Elongation factor Tu, mitochondrial	1.000	56.17	TYVR(1)DKPHVNVGTIGHVDHGK
P08113	51	Endoplasmin	1.000	91.24	TDDEVVQR(1)EEEAIQLDGLNASQIR
P84089	13	Enhancer of rudimentary homolog	0.949	87.68	SHTILLVQPTKR(0.949)PEGR(0.051)
Q8BH95	54	Enoyl-CoA hydratase, mitochondrial	1.000	93.42	NSSVGLIQLNR(1)PK
O88735	266	Ensconsin	1.000	114.27	AAHSR(1)NPVDRPK
O88735	212	Ensconsin	0.953	147.67	RLSSSSATLLNSPDR(0.953)AR(0.047)
O54782	715	Epididymis-specific alpha-mannosidase	0.865	70.41	IEQQYR(0.865)VGPLDLNR(0.135)
P23116	1259	Eukaryotic translation initiation factor 3 subunit A	0.867	164.13	DLRDDRDR(0.867)R(0.133)
P23116	1111	Eukaryotic translation initiation factor 3 subunit A	0.982	65.21	GADDDR(0.018)GPWR(0.982)NMDDDRVPR
P23116	956	Eukaryotic translation initiation factor 3 subunit A	1.000	121.10	LGGDDEERESSLRPDDDR(1)IPR
P23116	1091	Eukaryotic translation initiation factor 3 subunit A	1.000	131.33	RGLDDDRGPWR(1)NAAEDR
P23116	1221	Eukaryotic translation initiation factor 3 subunit A	1.000	129.02	RGPAEESSSWR(1)DSSR
P23116	979	Eukaryotic translation initiation factor 3 subunit A	0.799	164.53	RGPDEDRFSR(0.799)R(0.2)
Q6NZJ6	1019	Eukaryotic translation initiation factor 4 gamma 1	1.000	119.54	EAEMEEHR(1)EHIK
Q8BGD9	318	Eukaryotic translation initiation factor 4B	1.000	145.29	DDYSR(1)DDYRR
Q8BGD9	273	Eukaryotic translation initiation factor 4B	1.000	140.35	DYDRGYDSR(1)IGSGR
Q8BGD9	356	Eukaryotic translation initiation factor 4B	1.000	86.00	EDDASASTSQSSR(1)AASIFGGAKPVDTAAR
Q8BGD9	292	Eukaryotic translation initiation factor 4B	1.000	134.29	RDDDYR(1)GGGDRYEDR
Q8BGD9	225	Eukaryotic translation initiation factor 4B	0.961	120.90	RGDDSFGDKYR(0.961)DR(0.039)
Q9WUK2	210	Eukaryotic translation initiation factor 4H	1.000	186.76	AQRPR(1)LQLKPR
Q9WUK2	216	Eukaryotic translation initiation factor 4H	1.000	98.15	LQLKPR(1)TVATPLNQVANPNSAIFGGARPR
Q9WUK2	185	Eukaryotic translation initiation factor 4H	0.999	94.02	FR(0.999)DGPPLR(0.001)GSNMDFREPTEEER
Q9WUK2	191	Eukaryotic translation initiation factor 4H	1.000	140.04	FRDGPPLR(1)GSNMDFREPTEEER
P56564	505	Excitatory amino acid transporter 1	1.000	125.03	NR(1)DVEMGNSVIEENEMK
P12710	60	Fatty acid-binding protein, liver	1.000	104.18	VVR(1)NEFTLGEECELETMTGEK
Q923D2	39	Flavin reductase (NADPH)	1.000	92.46	DSSR(1)LPSEGPQPAHVVVGDVR
P97447	44	Four and a half LIM domains protein 1	1.000	106.35	FCANTCVDCR(1)KPISADAK
Q9QXD6	230	Fructose-1,6-bisphosphatase 1	1.000	117.00	DFDPAINEYLQR(1)K
P05064	331	Fructose-bisphosphate aldolase A	1.000	164.15	R(1)ALANSLACQGK
P05064	43	Fructose-bisphosphate aldolase A	1.000	181.64	R(1)LQSIGTENTEENRR
Q91Y97	56	Fructose-bisphosphate aldolase B	0.998	171.67	VENTEENR(0.998)R(0.002)
P05063	331	Fructose-bisphosphate aldolase C	1.000	100.93	R(1)AEMNGLAAQGR
P05063	43	Fructose-bisphosphate aldolase C	1.000	134.29	R(1)LSQIGVENTEENRR
P05063	201	Fructose-bisphosphate aldolase C	1.000	226.40	R(1)CQYVTEK
Q3UNH4	672	G protein-regulated inducer of neurite outgrowth 1	1.000	66.25	SIGSLPER(1)EPSASTSQK
P17183	50	Gamma-enolase	0.995	153.39	AAVPSGASTGIYEALELR(0.995)DGDKQR(0.005)
P17183	429	Gamma-enolase	1.000	130.47	FAGHNFR(1)NPSVL
P23242	293	Gap junction alpha-1 protein	0.988	100.43	LVTGDR(0.988)NNSSCR(0.012)
P23242	362	Gap junction alpha-1 protein	0.973	212.90	VAAGHELQPLAIVDQR(0.973)PSSR(0.027)
Q8BUV3	485	Gephyrin	0.801	51.47	ILVQAR(0.801)PGQDIR(0.199)PIGHDIK
P03995	20	Glial fibrillary acidic protein	1.000	236.96	SYASETVVR(1)GLGPSR
P03995	26	Glial fibrillary acidic protein	1.000	121.82	GLGPSR(1)QLGTMPR
P03995	138	Glial fibrillary acidic protein	1.000	98.18	LEVER(1)DNFAQDLGTLR
P03995	159	Glial fibrillary acidic protein	0.999	99.97	LQDETNLR(0.999)LEAENNLAAYR(0.001)QEADEATLAR
P03995	170	Glial fibrillary acidic protein	1.000	171.08	LEAENNLAAYR(1)QEADEATLAR
P03995	198	Glial fibrillary acidic protein	1.000	121.82	VESLEEEIQFLR(1)K
P03995	209	Glial fibrillary acidic protein	1.000	62.96	ELR(1)EQLAQQQVHVEMDVAKPDLTAALR
P03995	373	Glial fibrillary acidic protein	1.000	58.48	LLEGEENR(1)ITIPVQTFSNLQIR
P03995	413	Glial fibrillary acidic protein	1.000	153.48	TVEMR(1)DGEVIKDSK
P06745	417	Glucose-6-phosphate isomerase	1.000	140.73	MIPCDFLIPVQTQHPIR(1)K
P06745	472	Glucose-6-phosphate isomerase	1.000	87.68	VFEGNR(1)PTNSIVFTK

P26443	476	Glutamate dehydrogenase 1, mitochondrial	1.000	108.08	DSNYHLLMSVQESLER(1)K
P26443	123	Glutamate dehydrogenase 1, mitochondrial	0.833	77.08	IIKPCNHVLSLSFPIR(0.833)R(0.167)
P15105	340	Glutamine synthetase	0.782	143.07	GYFEDR(0.782)R(0.218)PSANCDPYAVTEAIVR
P15105	341	Glutamine synthetase	1.000	143.07	R(1)PSANCDPYAVTEAIVR
P15105	299	Glutamine synthetase	1.000	104.53	R(1)LTGFHETSNINDFSAGVANR
P47791	159	Glutathione reductase, mitochondrial	1.000	80.41	SHIEIIHGYATFADGPR(1)PTVEVNGK
P30115	45	Glutathione S-transferase A3	1.000	65.86	LR(1)SDGSLMFQQVPMVEIDGMK
P10649	43	Glutathione S-transferase Mu 1	1.000	82.18	RYTMGDAPDFDR(1)SQWLNEK
P19157	101	Glutathione S-transferase P 1	1.000	139.69	NQREAAQMDMVNDGVEDLR(1)GK
P19157	85	Glutathione S-transferase P 1	1.000	140.13	NQR(1)EAAQMDMVNDGVEDLRGK
P19157	202	Glutathione S-transferase P 1	1.000	244.72	AFLSSPEHVNR(1)PINGNGKQ
P16858	198	Glyceraldehyde-3-phosphate dehydrogenase	1.000	87.32	DGR(1)GAAQNIIPASTGAAK
P16858	78	Glyceraldehyde-3-phosphate dehydrogenase	1.000	221.21	LVINGKPITIFQER(1)DPTNIK
P16858	232	Glyceraldehyde-3-phosphate dehydrogenase	1.000	90.91	LTGMAFR(1)VPTPNVSVVDLTCR
P16858	246	Glyceraldehyde-3-phosphate dehydrogenase	1.000	230.30	VPTPNVSVVDLTCR(1)LEKPAK
Q9QXF8	230	Glycine N-methyltransferase	1.000	96.59	AHMVTLDYTVQVPGTGR(1)DGSPGFSK
Q9WUB3	520	Glycogen phosphorylase, muscle form	1.000	135.09	IGEDYISDLDQLR(1)K
P01117	41	GTPase KRas	1.000	162.87	SALTIQLIQNHFVDEYDPTIQDSYR(1)K
P63216	17	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-3	1.000	99.09	GETPVNSTMSIGQAR(1)K
P62874	8	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	1.000	159.75	SELDQLR(1)QEAEQLK
P62880	8	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	1.000	118.31	SELEQLR(1)QEAEQLR
P18872	145	Guanine nucleotide-binding protein G(o) subunit alpha	1.000	147.10	SR(1)EYQLNDSAK
P63017	100	Heat shock cognate 71 kDa protein	1.000	113.88	HWPFMVVNDAGR(1)PK
P63017	77	Heat shock cognate 71 kDa protein	1.000	90.71	R(1)FDDAVVQSDMK
P14602	131	Heat shock protein beta-1	1.000	89.47	HEER(1)QDEHGYISR
P07901	621	Heat shock protein HSP 90-alpha	1.000	91.64	AQALR(1)DNSTMGYMAAK
P07901	457	Heat shock protein HSP 90-alpha	1.000	182.37	LGIHEDSQNR(1)K
P11499	612	Heat shock protein HSP 90-beta	1.000	96.39	AQALR(1)DNSTMGYMMAK
Q6PGH2	15	Hematological and neurological expressed 1-like protein	1.000	72.19	SGSR(1)SMKPPGGESSDLFGSPEEGISSSKPNR
P51859	232	Hepatoma-derived growth factor	1.000	79.84	EEAEAQGVR(1)DHESL
P51859	189	Hepatoma-derived growth factor	1.000	50.78	ESGDHEEEDKEIAALEGER(1)PLPVEVEK
P49312	97	Heterogeneous nuclear ribonucleoprotein A1	1.000	103.67	EDSQR(1)PGAHLTVK
P49312	122	Heterogeneous nuclear ribonucleoprotein A1	1.000	222.62	EDTEEHHLR(1)DYFEQYGK
P49312	14	Heterogeneous nuclear ribonucleoprotein A1	1.000	111.64	SESPKEPEQLR(1)K
P49312	178	Heterogeneous nuclear ribonucleoprotein A1	1.000	79.65	YHTVNGHNCEVR(1)K
Q8BG05	9	Heterogeneous nuclear ribonucleoprotein A3	1.000	116.39	MEVKPPPGR(1)PQPDSGRR
Q8BG05	16	Heterogeneous nuclear ribonucleoprotein A3	0.825	90.73	MEVKPPPGR(0.014)PQPDSGR(0.825)R(0.161)
Q9Z2X1	206	Heterogeneous nuclear ribonucleoprotein F	1.000	84.17	FMSVQR(1)PGPYDRPGTAR
Q9Z2X1	212	Heterogeneous nuclear ribonucleoprotein F	0.992	107.72	FMSVQR(0.001)PGPYDR(0.992)PGTAR(0.007)
P61979	69	Heterogeneous nuclear ribonucleoprotein K	1.000	57.41	ALR(1)TDYNASVSVPDSSGPER
P61979	325	Heterogeneous nuclear ribonucleoprotein K	0.988	174.39	GGDLMAYDR(0.988)R(0.012)
Q9D0E1	428	Heterogeneous nuclear ribonucleoprotein M	1.000	91.24	MGAGLGHGMDR(1)VGSEIER
Q8VEK3	231	Heterogeneous nuclear ribonucleoprotein U	1.000	86.57	EDHGR(1)GYFEYIEENK
Q8VEK3	551	Heterogeneous nuclear ribonucleoprotein U	1.000	84.51	R(1)NFILDQTNVSAAAQR
O88569	228	Heterogeneous nuclear ribonucleoproteins A2/B1	0.999	86.19	GGGGNFGPGPGSNFR(0.999)GGSDGYGSGR(0.001)
O88569	325	Heterogeneous nuclear ribonucleoproteins A2/B1	1.000	71.02	SGNFGGSR(1)NMGGPYGGGNYGPGGSGGSGGYGGR
O88569	203	Heterogeneous nuclear ribonucleoproteins A2/B1	1.000	137.45	SGR(1)GGNFGFGDSR
O88569	213	Heterogeneous nuclear ribonucleoproteins A2/B1	0.998	70.67	SGR(0.002)GGNFGFGDSR(0.998)GGGGNFGPGPGSNFR
O88569	185	Heterogeneous nuclear ribonucleoproteins A2/B1	1.000	115.10	YHTINGHNAEVR(1)K
P63158	110	High mobility group protein B1	1.000	87.35	RPPSAFFLFCSEYR(1)PK
P27661	89	Histone H2AX	1.000	209.70	HLQLAIR(1)NDEELNK
Q6ZWY9	80	Histone H2B type 1-C/E/G	1.000	196.78	IAGEASR(1)LAHYNK
Q8CGP2	93	Histone H2B type 1-P	1.000	98.19	RSTITSR(1)EIQTAVR
Q8CGP2	80	Histone H2B type 1-P	1.000	96.25	IASEASR(1)LAHYNK
P62806	24	Histone H4	1.000	127.59	VLR(1)DNIQGITKPAIR
Q99JP6	285	Homer protein homolog 3	1.000	59.98	EAPDTAER(1)EETQQQVQDLETR
Q5BKQ4	127	Inactive pancreatic lipase-related protein 1	1.000	202.81	R(1)GSQTTYTQAANNVR
Q9D6R2	115	Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	1.000	99.06	GPLKTPIAAGHPSMNLLLR(1)K
P54071	89	Isocitrate dehydrogenase [NADP], mitochondrial	1.000	102.21	YFDLGLPNR(1)DQTNDQVTIDSALATQK
Q9D8B7	303	Junctional adhesion molecule C	1.000	156.60	TSEEGDFR(1)HK
P11679	23	Keratin, type II cytoskeletal 8	1.000	168.42	AFSSR(1)SFTSGPGAR
P11679	32	Keratin, type II cytoskeletal 8	1.000	146.60	SFTSGPGAR(1)ISSSSFSR
P11679	347	Keratin, type II cytoskeletal 8	1.000	79.43	ASLEAAIADAEQR(1)GEMAIK
P11679	40	Keratin, type II cytoskeletal 8	1.000	226.99	ISSSSFSR(1)VGSSSSSFR
P11679	94	Keratin, type II cytoskeletal 8	1.000	150.91	LEVDPNIQAVR(1)TQEK
P11679	407	Keratin, type II cytoskeletal 8	1.000	205.77	LLEGEESR(1)LESGMQNMSIHTK
P11679	18	Keratin, type II cytoskeletal 8	1.000	116.06	MSTSGPR(1)AFSSR
Q9CWU5	125	KH domain-containing protein 3	0.861	95.07	LMEIEALEAGVER(0.861)R(0.139)
Q61097	305	Kinase suppressor of Ras 1	1.000	113.42	SHESQLGNR(1)IDDVTPMK
P28740	183	Kinesin-like protein KIF2A	1.000	43.95	R(1)AQDVDATNPNYEIMCMIR
Q9WVM8	212	Kynurenine/alpha-aminoadipate aminotransferase, mitochondrial	1.000	70.35	FLYTVPNGNNPTGNSLTGDR(1)K
Q60675	75	Laminin subunit alpha-2	0.869	99.04	LVEHVPGQPVR(0.869)NPQCR(0.131)
Q6ZQ58	638	La-related protein 1	1.000	113.54	RHPGGDR(1)TGNHTSR
Q6A0A2	421	La-related protein 4B	1.000	88.39	SHLR(1)HAIPSTER
Q61792	130	LIM and SH3 domain protein 1	1.000	100.02	SR(1)MGPSGGEGVEPERR
P06151	157	L-lactate dehydrogenase A chain	1.000	100.13	NR(1)VIGSGCNLDSAR
P16125	318	L-lactate dehydrogenase B chain	1.000	153.95	DDEVAQLR(1)K
P16125	158	L-lactate dehydrogenase B chain	1.000	83.76	HR(1)VIGSGCNLDSAR
P0CW02	72	Lymphocyte antigen 6C1	1.000	55.45	TR(1)QCLSFCPAGVPIRDPNIR
P11588	163	Major urinary protein 1	1.000	85.74	HGILR(1)ENIIDLSNANR
P11588	78	Major urinary protein 1	1.000	156.23	FHTVR(1)DEECSELSMVADK
B5X0G2	78	Major urinary protein 17	1.000	137.64	VHTVR(1)DEECSELSMVADK
B5X0G2	140	Major urinary protein 17	1.000	102.21	DGETFQLMGLYGR(1)EPDLSSDIKER
P08249	191	Malate dehydrogenase, mitochondrial	1.000	91.28	GLDPAR(1)VNVPVIGGHAGK
A2AG50	318	MAP7 domain-containing protein 2	0.888	142.85	VATSAASGGHGSPLR(0.888)R(0.112)

Q8K310	231	Matrin-3	1.000	122.30	CR(1)DDSFFGETSHNYHK
Q8K310	186	Matrin-3	1.000	161.49	R(1)DSFDDRGPSLNPVLDYDHGSR
Q8K310	12	Matrin-3	1.000	134.29	SFQQSSLGR(1)DSQGHGR
Q8BI84	1736	Melanoma inhibitory activity protein 3	1.000	191.37	SEFGSLDR(1)HLPRPR
Q3ULD5	268	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	1.000	61.71	AATGEEVSAEDLGGADLHCR(1)K
Q9JLZ3	244	Methylglutaconyl-CoA hydratase, mitochondrial	1.000	130.15	AVGLISHVLEQNQEGDAAYR(1)K
Q9EQ20	458	Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	1.000	90.95	IVNDNPYGNGTAIFTTNGATAR(1)K
Q9EQ20	70	Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	1.000	89.43	WIDIHNPATNEVVGR(1)VPQSTK
Q8CAQ8	546	MICOS complex subunit Mic60	1.000	124.17	LR(1)GIEQAVQSHAVAEEEARK
Q8CAQ8	465	MICOS complex subunit Mic60	1.000	121.24	SEIQAEQDR(1)K
Q9QYR6	154	Microtubule-associated protein 1A	1.000	73.92	R(1)SIEEACLTLQHLNR
Q9QYR6	2390	Microtubule-associated protein 1A	1.000	124.82	SSRPDTLLSSEQR(1)PGK
Q9QYR6	1690	Microtubule-associated protein 1A	0.868	87.68	VWFPHELDGQGAR(0.868)PR(0.132)
P14873	996	Microtubule-associated protein 1B	0.998	65.18	R(0.002)ESVVSGDDR(0.998)AEEDMDDVLEK
P20357	1460	Microtubule-associated protein 2	1.000	183.74	KEPSTVSR(1)DEVR
P20357	51	Microtubule-associated protein 2	1.000	60.06	NANGFPYR(1)EEEEGAFGEHR
P20357	1488	Microtubule-associated protein 2	1.000	124.50	SEVQAHSPSR(1)K
Q7TSJ2	279	Microtubule-associated protein 6	1.000	148.48	ASGADQR(1)DTR
Q7TSJ2	435	Microtubule-associated protein 6	1.000	154.34	NEFR(1)AWTDIKPVKPIK
Q7TSJ2	431	Microtubule-associated protein 6	0.976	58.81	QIREEVASTVSSSYR(0.976)NEFR(0.024)
Q61166	162	Microtubule-associated protein RP/EB family member 1	0.924	68.42	KPLGSSTAAPQR(0.924)PIATQR(0.076)
P10637	534	Microtubule-associated protein tau	1.000	141.60	SR(1)LQTAPVPMPDLK
Q9D6M3	273	Mitochondrial glutamate carrier 1	1.000	134.86	GVNEDTYSGFLDCAR(1)K
P63085	299	Mitogen-activated protein kinase 1	1.000	80.18	R(1)IEVEQALAHPYLEQYYDPSDEPIAEAPFK
P26041	533	Moesin	1.000	82.88	ALTSELANAR(1)DESK
P26041	393	Moesin	1.000	188.48	ER(1)QEAEEAK
P26041	570	Moesin	1.000	174.23	QR(1)IDEFESM
P04370	157	Myelin basic protein	1.000	352.87	YLATASTMDHAR(1)HGFLPR(0.844)HR(0.156)
P04370	163	Myelin basic protein	1.000	278.16	HGFLPR(1)HR
P04370	165	Myelin basic protein	1.000	268.47	HR(1)DTGILDSIGR
P04370	175	Myelin basic protein	1.000	294.21	HRDTGILDSIGR(1)FFSGDRGAPK
P04370	181	Myelin basic protein	1.000	164.65	FFSGDR(1)GAPK
P04370	195	Myelin basic protein	1.000	296.81	DSHTR(1)TTHYGSLPQKSQHGR
P04370	228	Myelin basic protein	1.000	256.67	NIVTPR(1)TPPPSQGK
P04370	242	Myelin basic protein	1.000	218.58	DSR(1)SGSPMAR(0.826)R(0.174)
P04370	249	Myelin basic protein	0.998	165.18	SGSPMAR(0.998)R(0.002)
P60202	127	Myelin proteolipid protein	1.000	78.95	GSR(1)GQHQAHSLER
P20917	563	Myelin-associated glycoprotein	1.000	268.48	NVTESSSFSGGDNPHVLYSPEFR(1)ISGAPDK
P20917	587	Myelin-associated glycoprotein	1.000	224.91	LLGLR(1)GESPELDLSYSHSDLGK
P20917	582	Myelin-associated glycoprotein	1.000	185.44	R(1)LLGLRGESPELDLSYSHSDLGK
P20917	77	Myelin-associated glycoprotein	1.000	84.65	SR(1)TQVVHESFQGR
Q9D2P8	90	Myelin-associated oligodendrocyte basic protein	1.000	84.33	HQPAASPVVVR(1)APPAKPK
P04247	140	Myoglobin	1.000	152.34	ALELFR(1)NDIAAK
P04247	119	Myoglobin	1.000	105.52	R(1)HSGDFGADAQGAMSK
P97434	255	Myosin phosphatase Rho-interacting protein	0.940	72.57	EDESTISGDR(0.94)VDGGR(0.06)K
Q9QVP4	48	Myosin regulatory light chain 2, atrial isoform	1.000	78.71	EAFSCIDQNR(1)DGIICK
P97457	41	Myosin regulatory light chain 2, skeletal muscle isoform	1.000	73.57	EAFTVIDQNR(1)DGIIDKEDLR
P51667	50	Myosin regulatory light chain 2, ventricular/cardiac muscle isoform	1.000	119.45	NDLR(1)DTFAALGR
P51667	40	Myosin regulatory light chain 2, ventricular/cardiac muscle isoform	1.000	195.71	EAFTIMDQNR(1)DGFIDK
Q5SX40	1200	Myosin-1	1.000	79.14	DLEEATLQHEATAATLR(1)K
Q5SX40	1669	Myosin-1	1.000	55.84	DTQLHLDDALR(1)GQEDLK
Q61879	1710	Myosin-10	1.000	80.81	HAEQER(1)DELADEIANSASGK
O08638	809	Myosin-11	1.000	155.48	R(1)QQQLTAMK
Q6URW6	593	Myosin-14	1.000	91.28	NLR(1)DQADFSVLHYAGK
Q6URW6	968	Myosin-14	1.000	96.79	R(1)LQQHIQELESHLEAEEGAR
Q5SX39	1800	Myosin-4	1.000	156.67	DLQHR(1)LDEAEQLALK
Q5SX39	1884	Myosin-4	1.000	160.57	R(1)QAEEAEEQSNVNLAK
Q02566	1678	Myosin-6	0.860	155.55	DTQLQLDDAVHANDDLKENIAIVER(0.86)R(0.14)
Q02566	1679	Myosin-6	1.000	155.55	R(1)NNLLQAELEELR
Q02566	23	Myosin-6	1.000	115.56	ER(1)LEAQTRPFDIR
Q02566	29	Myosin-6	1.000	98.68	ERLEAQTR(1)PFDIR
Q02566	1346	Myosin-6	1.000	223.89	HDCDLLR(1)EQYEEEMEAK
Q02566	1319	Myosin-6	1.000	162.67	R(1)QLEEEGKAK
Q02566	370	Myosin-6	1.000	225.88	QR(1)EEQAEPDGTEDADK
Q02566	190	Myosin-6	1.000	119.54	R(1)VIQYFASIAAIGDR
Q02566	17	Myosin-6	1.000	184.78	TDAQMADFGAAAQYLR(1)K
Q02566	1262	Myosin-6	1.000	179.93	TLEDQANEYR(1)VK
Q02566	1384	Myosin-6	1.000	169.52	YETDAIQR(1)TEELEEAKK
Q02566	1195	Myosin-6	1.000	145.77	DLEEATLQHEATAAALR(1)K
Q02566	1178	Myosin-6	1.000	91.24	R(1)DLEEATLQHEATAAALR
Q02566	654	Myosin-6	1.000	220.69	KGSSFQTVSALHR(1)ENLNK
Q02566	1422	Myosin-6	1.000	240.25	HR(1)LQNEIEDLMVDVER
Q02566	741	Myosin-6	1.000	135.23	YRILNPAAIPEGQFIDSR(1)K
Q02566	1882	Myosin-6	1.000	280.85	R(1)QAEEAEEQANTNLSK
Q02566	147	Myosin-6	1.000	108.23	R(1)SEAPPHIFSISDNAYQYMLTDR
Q02566	169	Myosin-6	1.000	180.97	RSEAPPHIFSISDNAYQYMLTDR(1)ENQSILITGESGAGK
Q02566	1911	Myosin-6	1.000	157.68	VQHELDEAEER(1)ADIAESQVNK
Q02566	1822	Myosin-6	1.000	119.39	VR(1)ELENELEAEQK
Q02566	725	Myosin-6	1.000	98.18	YR(1)ILNPAAIPEGQFIDSRK
Q91Z83	1677	Myosin-7	1.000	114.72	R(1)NNLLQAELEELR
Q8VDD5	1697	Myosin-9	0.850	126.27	R(0.85)QAQQER(0.15)DELADEIANSSGK
Q8VDD5	1703	Myosin-9	0.949	118.05	R(0.051)QAQQER(0.949)DELADEIANSSGK
O70468	966	Myosin-binding protein C, cardiac-type	0.898	291.58	AHNVAGPGGPIVTKEPVTVQEILQR(0.898)PR(0.102)
O70468	941	Myosin-binding protein C, cardiac-type	1.000	45.27	VR(1)AHNVAGPGGPIVTKEPVTVQEILQRPR
Q9JHL1	292	Na(+)/H(+) exchange regulatory cofactor NHE-RF2	1.000	75.52	R(1)DPFQESGLHLSPTAAEAK

Q9JHL1	156	Na(+)/H(+) exchange regulatory cofactor NHE-RF2	1.000	98.55	R(1)GPQGYGFNLHSDK
Q91XE4	255	N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming)	1.000	162.19	TADGDLAGTVHPQLQDHDFEPLR(1)PGEPIFK
Q9D8B4	20	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 11	1.000	101.75	RFFESYHEVPDGTQCHR(1)K
Q7TMF3	9	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	1.000	87.99	R(1)GVQQVTGHGGLR
Q9Z1P6	34	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 7	1.000	127.49	R(1)TQPPPKLPVGPSHK
Q9DCJ5	147	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8	1.000	101.17	AR(1)PEPNPVIEGDLKPAK
O09111	127	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial	1.000	59.34	YR(1)EVNGLPIMESNYFDPSK
Q9D6J6	153	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	1.000	176.29	DSDSILETLQR(1)K
Q9DCT2	230	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	1.000	96.56	RVVAEPVELAQEFR(1)K
Q9DCN2	219	NADH-cytochrome b5 reductase 3	0.829	62.59	DILLR(0.829)PELEELR(0.17)NEHSAR(0.001)
Q9DCN2	226	NADH-cytochrome b5 reductase 3	0.947	134.68	DILLRPELEELR(0.947)NEHSAR(0.053)
P61082	157	NEDD8-conjugating enzyme Ubc12	0.892	109.44	EAAEVLQNNR(0.892)R(0.108)
Q6R891	370	Neurabin-2	1.000	86.95	GVDNGR(1)APDMAPEEVDESK
P13595	787	Neural cell adhesion molecule 1	1.000	105.53	TEEER(1)TPNHDGGK
P12961	55	Neuroendocrine protein 7B2	0.976	90.35	LLHGVMEQLGIAR(0.976)PR(0.024)
Q810U3	449	Neurofascin	1.000	67.65	TR(1)LDCPFFGSPIPTLR
P19246	229	Neurofilament heavy polypeptide	0.910	162.29	AQALQEECGYLR(0.91)R(0.09)
P19246	66	Neurofilament heavy polypeptide	1.000	73.68	FR(1)GAASSTDSLDTLSNGPEGCVVAAVAAR
P19246	190	Neurofilament heavy polypeptide	1.000	195.78	QR(1)EEAEAAAR
P19246	131	Neurofilament heavy polypeptide	0.996	136.26	SLEGEAAALR(0.996)QQQAGR(0.004)
P08551	30	Neurofilament light polypeptide	1.000	195.77	VHISSVR(1)SGYSTAR
P08551	23	Neurofilament light polypeptide	1.000	107.50	YVETPR(1)VHISSVR
P08553	722	Neurofilament medium polypeptide	1.000	110.03	AEEEGGSEEEGSDR(1)SPQESK
P08553	425	Neurofilament medium polypeptide	1.000	223.53	FSTFSGSITGPLYTHR(1)QPSVTISSK
P08553	369	Neurofilament medium polypeptide	1.000	235.57	HNHDLSSYQDTIQQLENELR(1)GTK
P08553	256	Neurofilament medium polypeptide	1.000	163.14	SNHEEEVADLLAQIQASHITVER(1)K
P08553	26	Neurofilament medium polypeptide	1.000	104.84	SSFSR(1)VSGSPSSGFR
P08553	380	Neurofilament medium polypeptide	1.000	176.04	HLR(1)EYQDLLNVK
P60761	38	Neurogranin	1.000	252.52	IQASFR(1)GHMAR
P60761	68	Neurogranin	1.000	149.29	KGPGPGGPGGAGGAR(1)GGAGGGPSGD
P60761	53	Neurogranin	1.000	86.95	SGECGR(1)KGPGPGGPGGAGGAR
P35802	269	Neuronal membrane glycoprotein M6-a	1.000	76.30	EEQELHDIHSTR(1)SK
Q9Z1S5	255	Neuronal-specific septin-3	1.000	99.41	IR(1)QESMPFAVVGSDK
P32020	33	Non-specific lipid-transfer protein	1.000	78.71	FMKPGGENSR(1)DYPDMAK
Q9CZ44	136	NSFL1 cofactor p47	1.000	82.35	EHGAVAVER(1)VTK
P62960	144	Nuclease-sensitive element-binding protein 1	0.954	166.62	YAADRNHYR(0.954)R(0.046)
P00688	71	Pancreatic alpha-amylase	0.942	82.13	GFGGVQVSPPNENVVVHNPSR(0.942)PWWER(0.058)
P00688	264	Pancreatic alpha-amylase	1.000	142.32	GSEYFGNGR(1)VTEFK
P00688	478	Pancreatic alpha-amylase	1.000	158.56	VDGNCTGLR(1)VNVGSDGK
Q6P8U6	330	Pancreatic triacylglycerol lipase	1.000	56.69	CFPCGSGGCPQMGHYADR(1)YPGK
P24369	109	Peptidyl-prolyl cis-trans isomerase B	1.000	83.87	DFMIQGGDFTR(1)GDGTGGK
P24369	190	Peptidyl-prolyl cis-trans isomerase B	1.000	191.41	TDSR(1)DKPLKDVIIVDSGK
Q9D1M7	80	Peptidyl-prolyl cis-trans isomerase FKBP11	1.000	127.01	IIDTSLTR(1)DPLVIELGQK
P26883	19	Peptidyl-prolyl cis-trans isomerase FKBP1A	1.000	179.66	R(1)GQTCVVHYTGMLEDGKK
P15331	304	Peripherin	0.983	70.56	YADLSDAANR(0.983)NHEALR(0.017)
P35700	110	Peroxiredoxin-1	1.000	123.76	R(1)TIAQDYGVLK
P99029	176	Peroxiredoxin-5, mitochondrial	1.000	80.63	R(1)FSMVIDNGIVK
Q9R0H0	31	Peroxisomal acyl-coenzyme A oxidase 1	0.807	64.59	ERAAATFNPELITHILDGSPENTR(0.807)R(0.193)
Q9R0H0	509	Peroxisomal acyl-coenzyme A oxidase 1	1.000	124.12	NLQAQVSHR(1)K
Q99P30	3	Peroxisomal coenzyme A diphosphatase NUDT7	0.999	59.54	SR(0.999)PCGLPEPVR(0.001)NNLIDDAK
Q8VEM8	184	Phosphate carrier protein, mitochondrial	1.000	124.29	VR(1)IQTQPGYANTLR
Q9D0F9	23	Phosphoglucomutase-1	1.000	87.35	TQAYPDQKPGTSGLR(1)K
Q9DBJ1	240	Phosphoglycerate mutase 1	1.000	85.88	NLKPIKPMQFLGDEETVR(1)K
Q9R0K7	249	Plasma membrane calcium-transporting ATPase 2	1.000	196.31	IDESSLTGESDQVR(1)K
Q9CY58	201	Plasminogen activator inhibitor 1 RNA-binding protein	1.000	199.99	HSGSDR(1)SSFSHYSGLK
Q9QXS1	3344	Plectin	1.000	93.58	CR(1)SDQLTGLSLLPLSEK
P60335	243	Poly(rC)-binding protein 1	1.000	92.37	LNQVAR(1)QQSHFAMMHGGTGFAGIDSSSPEVK
P22907	116	Porphobilinogen deaminase	1.000	78.33	R(1)ENPCDAVVFHPK
P48678	166	Prelamin-A/C	1.000	123.60	RTLEGELHDLR(1)GQVAK
Q925B0	247	PRKC apoptosis WT1 regulator protein	1.000	86.18	HNR(1)DANAPASFSSSSTLEK
Q61656	516	Probable ATP-dependent RNA helicase DDX5	1.000	142.58	DRENYDR(1)GYSNLLK
Q61656	509	Probable ATP-dependent RNA helicase DDX5	0.996	90.58	RGGFNTFR(0.996)DR(0.003)ENYDR(0.002)
Q7TNG8	431	Probable D-lactate dehydrogenase, mitochondrial	1.000	83.99	R(1)ALALGGTCTGEHGIGLGK
Q7TNG8	450	Probable D-lactate dehydrogenase, mitochondrial	1.000	127.92	R(1)QLLQEEVGPVGVETMR
Q9JJV2	56	Profilin-2	1.000	87.68	DR(1)EGFFTNGLTLGAK
O35129	111	Prohibitin-2	1.000	97.49	VLSR(1)PNAQELPSMYQR
P01193	170	Pro-opiomelanocortin	1.000	129.08	ELEGER(1)PLGLEQVLESDAEK
Q9QXV0	92	ProSAAS	1.000	128.35	AR(1)AEAQEAEDQQAR
Q9R1P4	51	Proteasome subunit alpha type-1	1.000	108.63	R(1)AQSELAAHQK
O88737	3605	Protein bassoon	1.000	112.36	FR(1)HHGGHTVSSSQK
Q4VAA2	27	Protein CDV3	1.000	59.12	SSR(1)AANAASGAGGSSAAAGSRPGDGGSLGSGAR
P09103	215	Protein disulfide-isomerase	1.000	172.81	KFDEGR(1)NNFEGEITKEK
D3Z6P0	153	Protein disulfide-isomerase A2	1.000	55.59	R(1)VGPSATHLEDEEGVQALMAK
P27773	207	Protein disulfide-isomerase A3	1.000	54.76	EYDDNGEGITIFR(1)PLHLANK
Q9D0F3	381	Protein ERGIC-53	1.000	48.60	R(1)GAGTPGQPGQVSQQELDTVVK
Q5SVD0	4	Protein FAM101B	1.000	71.31	MVGR(1)LSLQDVPELVDTKK
Q8CHT6	260	Protein FAM169B	1.000	58.71	LWEVEPPGAWGQQR(1)VNIWLK
E9Q8I9	1937	Protein furry homolog	1.000	109.84	NSDLFTVLSR(1)SSSPDLSSSSK
Q80TL4	327	Protein KIAA1045	1.000	72.69	R(1)LTEAPSCSVSISHVGPIADSSPAASSSK
Q61644	196	Protein kinase C and casein kinase substrate in neurons protein 1	1.000	154.81	CR(1)QDVQK
P63318	21	Protein kinase C gamma type	0.783	87.51	AGLGPGGGDSEGGPR(0.217)PLFCR(0.783)K
P63318	104	Protein kinase C gamma type	1.000	52.93	FR(1)LHSYSSPTFCDHCGSLLYGLVHQGMK
P63318	632	Protein kinase C gamma type	0.872	112.17	LEIAPPFR(0.872)PR(0.106)PCGR(0.022)
Q62433	56	Protein NDRG1	1.000	98.44	GNR(1)PVILTYHDIGMNHK
Q62433	3	Protein NDRG1	1.000	158.67	SR(1)ELHDVDLAEVKPLVEK

Q60829	167	Protein phosphatase 1 regulatory subunit 1B	0.988	46.19	GLEGPWER(0.002)PPPLDEPQR(0.988)DGNSEDQVEGR(0.01)
Q60829	40	Protein phosphatase 1 regulatory subunit 1B	1.000	181.88	RRPTPAMLFR(1)VSEHSSPEEEASPHQR
Q3UM45	77	Protein phosphatase 1 regulatory subunit 7	0.999	42.71	GAEDPEEEHELAVDMETINLDR(0.999)DAEDVDLTHYR(0.001)
P31725	25	Protein S100-A9	1.000	157.98	SITTIIDTFHQYSR(1)K
O88978	10	Protein tilB homolog	0.990	169.41	ITEDLIR(0.99)R(0.01)
E9QAT4	400	Protein transport protein Sec16A	1.000	180.97	RLNLDSFSSTSR(1)LGHPPPPGASGVYQAFPR
E9QAT4	1240	Protein transport protein Sec16A	0.768	65.43	YSEPER(0.22)PSSR(0.768)ASHYSDQLAPR(0.012)
Q3UPL0	1168	Protein transport protein Sec31A	1.000	118.96	LR(1)EQTLSPTIINGLHSIAR
P61620	24	Protein transport protein Sec61 subunit alpha isoform 1	1.000	80.02	FLEVIKPFCVILPEIQKPER(1)K
P63054	51	Purkinje cell protein 4	1.000	191.53	AAVAIQSQFR(1)K
O89086	128	Putative RNA-binding protein 3	0.932	85.06	YDSRPGGYGYGYGR(0.932)SR(0.068)
Q80TZ3	277	Putative tyrosine-protein phosphatase auxilin	1.000	83.67	NGCR(1)PYCDVLIGETK
Q9D051	49	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	1.000	88.57	EAINQGMDEELER(1)DEK
P52480	455	Pyruvate kinase PKM	0.993	116.06	APIIAVTR(0.993)NPQTAR(0.007)
P52480	399	Pyruvate kinase PKM	0.833	53.30	EAEAAIYHLQLFEELR(0.833)R(0.167)
P52480	339	Pyruvate kinase PKM	0.758	133.15	KPR(0.758)PTR(0.242)AEGSDVANAVLDGADCIMLSGETAK
P52480	342	Pyruvate kinase PKM	0.758	133.15	KPR(0.242)PTR(0.758)AEGSDVANAVLDGADCIMLSGETAK
P50396	169	Rab GDP dissociation inhibitor alpha	0.964	130.89	TFEGVDPQNTSMR(0.964)DVYR(0.036)K
P50396	173	Rab GDP dissociation inhibitor alpha	0.911	123.19	TFEGVDPQNTSMR(0.089)DVYR(0.911)K
P50396	35	Rab GDP dissociation inhibitor alpha	1.000	91.87	VLHMDR(1)NPYYGGESSSITPLEELYK
P50396	328	Rab GDP dissociation inhibitor alpha	1.000	208.72	NTNDANSCQIIIPQNQVNR(1)K
Q9JKF1	798	Ras GTPase-activating-like protein IQGAP1	1.000	89.23	AYQDR(1)LAYLHSHK
F6SEU4	1082	Ras/Rap GTPase-activating protein SynGAP	1.000	73.77	SQQLTVSAAQKPR(1)PSSGNLLQSPEPSYGPARPR
P63011	151	Ras-related protein Rab-3A	1.000	123.32	GR(1)QLADHLGFEFFEASAK
P63321	135	Ras-related protein Ral-A	1.000	128.61	R(1)QVSVEEAK
B9EKR1	447	Receptor-type tyrosine-protein phosphatase zeta	0.806	74.91	DNEEDTGLNPGR(0.806)DSVTNQIR(0.194)K
B9EKR1	455	Receptor-type tyrosine-protein phosphatase zeta	0.948	111.65	DNEEDTGLNPGR(0.052)DSVTNQIR(0.948)K
Q64374	274	Regucalcin	1.000	151.06	DGLNAEGLLR(1)QPDAGNIFK
Q8K0T0	237	Reticulon-1	1.000	74.03	AEGVR(1)APNQPAPVEGK
Q8K0T0	375	Reticulon-1	1.000	137.62	GLSYETTESPR(1)PVGQVADKPK
Q8K0T0	201	Reticulon-1	1.000	112.08	YIDITR(1)PQEAK
P24549	420	Retinal dehydrogenase 1	1.000	105.00	R(1)ANNTTYGLAAGLFTK
Q99PT1	49	Rho GDP-dissociation inhibitor 1	1.000	130.56	SIQEIQELDKDDESLR(1)K
Q99PL5	1394	Ribosome-binding protein 1	1.000	59.89	GELESSDQVR(1)EHTSHLEAELEK
Q9Z2I4	674	Roundabout homolog 3	1.000	67.65	GLAEVAVR(1)MQEPTVLGPR
Q91X83	265	S-adenosylmethionine synthase isoform type-1	1.000	71.59	FVIGGPQGDAGVTGR(1)K
O55143	246	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	1.000	154.84	IRDEMVATEQER(1)TPLQQK
Q8R0F9	275	SEC14-like protein 4	0.923	67.22	R(0.077)YYLSNQER(0.923)PQYEHSVVVGR
P17563	382	Selenium-binding protein 1	1.000	71.35	R(1)IPGGPQMIQLSLDGK
P17563	62	Selenium-binding protein 1	1.000	115.12	SPQYSQVIHR(1)LPMPYLK
Q9WUH7	708	Semaphorin-4G	0.979	135.00	R(0.021)R(0.979)KYSLGR
O55131	380	Septin-7	0.990	166.69	DSEAELQR(0.99)R(0.01)
O55131	41	Septin-7	1.000	98.54	NLEGYVGFANLPNQVYR(1)K
Q8CHH9	422	Septin-8	1.000	169.26	AAMEALQSQALHATSQQPLR(1)K
Q8CHH9	84	Septin-8	1.000	94.23	LR(1)PQTYDLQESNVHLK
Q03734	253	Serine protease inhibitor A3M	1.000	116.60	HFR(1)DEELSCSVLELK
Q9JKK8	2324	Serine/threonine-protein kinase ATR	0.998	50.83	DDLR(0.998)KDCR(0.002)LMEFNSLINK
Q9JLM8	15	Serine/threonine-protein kinase DCLK1	1.000	128.74	DMELEHFDER(1)DK
Q9JLM8	5	Serine/threonine-protein kinase DCLK1	1.000	65.18	SFGR(1)DMELEHFDERDK
P62137	187	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	0.849	157.68	IFCCHGGLSPDLQSMEQIR(0.849)R(0.151)
P62141	313	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	0.997	63.30	YQYGGLNSGR(0.997)PVTPPR(0.003)
Q921I1	239	Serotransferrin	1.000	65.62	ADR(1)DQYELLCLDNTRKPVDQYEDCYLAR
Q921I1	251	Serotransferrin	1.000	105.35	ADRDQYELLCLDNTR(1)KPVDQYEDCYLAR
Q921I1	610	Serotransferrin	1.000	112.85	DFASCHLAQAPNHVVVSR(1)K
Q921I1	588	Serotransferrin	1.000	69.99	QEDFELLCPDGTR(1)KPVK
P07724	141	Serum albumin	1.000	96.29	DDNPSLPPFER(1)PEAEAMCTSFK
P07724	34	Serum albumin	1.000	83.40	SEIAHR(1)YNDLGEQHFK
Q91VW3	47	SH3 domain-binding glutamic acid-rich-like protein 3	0.982	86.08	RIQYQLVDISQDNALR(0.982)DEMR(0.018)
P14576	485	Signal recognition particle 54 kDa protein	1.000	100.53	VLHHMGGMAGLQSMMR(1)QFQQGAAGNMK
P27048	49	Small nuclear ribonucleoprotein-associated protein B	1.000	84.65	HMNLILCDCDEFR(1)K
Q6PIE5	680	Sodium/potassium-transporting ATPase subunit alpha-2	1.000	115.01	DMTSEQLDEILR(1)DHTEIVFAR
Q6PIE5	493	Sodium/potassium-transporting ATPase subunit alpha-2	1.000	88.24	YQLSIHER(1)EDSPQSHVLVMK
Q6PIC6	43	Sodium/potassium-transporting ATPase subunit alpha-3	1.000	205.52	MSVEEVCR(1)K
Q6PIC6	492	Sodium/potassium-transporting ATPase subunit alpha-3	1.000	210.29	YQLSIHETEDPNDNR(1)YLLVMK
P14231	274	Sodium/potassium-transporting ATPase subunit beta-2	1.000	82.35	INAANIATDDER(1)DK
P55012	221	Solute carrier family 12 member 2	1.000	224.99	IDHYR(1)HTAAQLGEK
P16546	1803	Spectrin alpha chain, non-erythrocytic 1	1.000	84.37	DLTGVQNLR(1)K
P16546	566	Spectrin alpha chain, non-erythrocytic 1	1.000	68.97	R(1)AQLADSFHLQQFFR
P16546	883	Spectrin alpha chain, non-erythrocytic 1	1.000	65.43	R(1)QDLEDSLQAQQYFADANEAESWMR
P16546	2064	Spectrin alpha chain, non-erythrocytic 1	1.000	83.54	R(1)WTQLLANSATR
P16546	439	Spectrin alpha chain, non-erythrocytic 1	1.000	58.86	SADESGQALLAASHYASDEVR(1)EK
Q62261	797	Spectrin beta chain, non-erythrocytic 1	1.000	101.85	DVAEEITNYR(1)PTIDTLHEQASALPQAHAESPDVK
Q62261	1834	Spectrin beta chain, non-erythrocytic 1	1.000	90.46	KLPEELGR(1)DQNTVETLQR
Q62261	1245	Spectrin beta chain, non-erythrocytic 1	1.000	109.72	LVSDGNINSDR(1)IQEK
Q62261	2205	Spectrin beta chain, non-erythrocytic 1	1.000	77.87	TLETPAAQMEGFLNR(1)K
Q9WTX5	154	S-phase kinase-associated protein 1	1.000	179.14	NDFTEEEEAQVR(1)K
Q60598	97	Src substrate cortactin	1.000	146.81	MDR(1)SAVGHEYQSK
Q78PY7	503	Staphylococcal nuclease domain-containing protein 1	1.000	150.06	KEVPIHR(1)VADISGDTQK
Q78PY7	423	Staphylococcal nuclease domain-containing protein 1	1.000	117.62	KVNVTVDYIR(1)PASPATETVPAFSER
Q78PY7	485	Staphylococcal nuclease domain-containing protein 1	1.000	168.64	SSHYDELLAAEAR(1)AIK
Q78PY7	184	Staphylococcal nuclease domain-containing protein 1	1.000	41.48	YTIENPR(1)HFVDSHHQKPVNAIIEHVR
P54227	14	Stathmin	1.000	76.82	R(1)ASGQAFELILSPR
P38647	107	Stress-70 protein, mitochondrial	1.000	138.97	R(1)QAVTNPNNTFYATK
Q8K2B3	476	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	1.000	155.17	ACALSIAESCR(1)PGDKVPSIK
Q8BWF0	400	Succinate-semialdehyde dehydrogenase, mitochondrial	1.000	120.63	R(1)HQSGGNFFEPTLLSNVTR

Q8R086	269	Sulfite oxidase, mitochondrial	0.840	81.35	HEVTVTLQCAGNR(0.84)R(0.16)
O88935	587	Synapsin-1	1.000	59.57	ASGAPPGGQQR(1)QGPPQKPPGPAGPTR
O88935	430	Synapsin-1	1.000	203.82	DASPGR(1)GSHSQSSSPGALTLGR
O88935	534	Synapsin-1	1.000	97.39	QSR(1)PVAGGPGAPPAARPPASPSPQR
O88935	680	Synapsin-1	1.000	96.77	SQSLTNAFNLPEPAPPR(1)PSLSQDEVK
O88935	352	Synapsin-1	1.000	121.33	TNTGSAMLEQIAMSDR(1)YK
Q64332	59	Synapsin-2	1.000	83.36	R(1)PPPAQAPAPQPAPQPAPTPSVGSSFFSSLSQAVK
Q64332	429	Synapsin-2	1.000	130.49	TPALSPQR(1)PLTTQQPQSGTLKEPDSSK
Q9JIS5	136	Synaptic vesicle glycoprotein 2A	0.971	171.07	GGLSDGEGPPGGR(0.971)GEAQR(0.029)
Q8BG39	7	Synaptic vesicle glycoprotein 2B	1.000	82.64	YR(1)DNYEGYAPSDGYYR
P60879	8	Synaptosomal-associated protein 25	0.999	60.62	AEDADMR(0.999)NELEEMQR(0.001)
P60879	17	Synaptosomal-associated protein 25	1.000	141.07	R(1)ADQLADESLESTRR
P60879	142	Synaptosomal-associated protein 25	0.990	51.30	R(0.01)VTNDAR(0.99)ENEMDENLEQVSGIIGNLR
P46096	260	Synaptotagmin-1	1.000	46.85	HDIIGEFKVPMNTVDFGHVTEEWR(1)DLQSAEK
O08599	467	Syntaxin-binding protein 1	1.000	115.91	ER(1)ISEQTYQLSR
P26039	1919	Talin-1	1.000	96.85	HR(1)VQELGHGCSALVTK
P26039	2455	Talin-1	1.000	138.45	R(1)LQAAGNAVK
P52196	249	Thiosulfate sulfurtransferase	1.000	84.65	VDLSQPLIATCR(1)K
A2ASS6	24001	Titin	1.000	89.36	EACR(1)EDVGHYTVK
A2ASS6	4171	Titin	1.000	118.72	IDVQGGR(1)DHLSDAQK
A2ASS6	19790	Titin	1.000	71.62	NSSGHAQGSAIVNVLDR(1)PGPCQNLK
A2ASS6	467	Titin	1.000	152.47	TTTTAVHIQPAQEQAR(1)K
A2ASS6	19180	Titin	1.000	62.59	VAAENQYGR(1)GPFVETPKPIK
P62869	68	Transcription elongation factor B polypeptide 2	0.982	46.41	TLGECGFTSQTAR(0.982)PQAPATVGLAFR(0.018)
Q3UHF7	1696	Transcription factor HIVEP2	1.000	66.54	ITAEIYTLAAMHR(1)PGAGK
P42669	139	Transcriptional activator protein Pur-alpha	0.757	43.09	DYLGDFIEHYAQLGPSQPPDLAQAQDEPR(0.757)R(0.243)
P42669	170	Transcriptional activator protein Pur-alpha	0.772	94.06	IR(0.772)QTVNR(0.228)GPGLGSTQGQTIALPAQGLIEFR
Q01853	349	Transitional endoplasmic reticulum ATPase	0.984	88.98	AHVIVMAATNR(0.984)PNSIDPALR(0.012)R(0.003)
Q01853	210	Transitional endoplasmic reticulum ATPase	1.000	115.79	REDEEESLNEVGYDDIGGCR(1)K
Q01853	487	Transitional endoplasmic reticulum ATPase	1.000	88.43	R(1)ELQELVQYPVEHPDK
Q9CPW5	96	Translocon-associated protein subunit beta	1.000	82.61	IAPASNVSHTVVLR(1)PLK
Q62186	59	Translocon-associated protein subunit delta	1.000	98.68	NR(1)VQNMALYADVSGK
Q923K4	18	tRNA modification GTPase GTPBP3, mitochondrial	1.000	47.50	MWRGLSALVTQAAWAPLR(1)LCAR
Q9JKK7	159	Tropomodulin-2	1.000	147.10	FDEETTNGEGR(1)K
P58771	21	Tropomyosin alpha-1 chain	1.000	247.92	LDKENALDR(1)AEQAEADKK
P58771	91	Tropomyosin alpha-1 chain	1.000	180.05	R(1)IQLVEEELDRAQER
P21107	92	Tropomyosin alpha-3 chain	1.000	155.49	R(1)IQLVEEELDRAQER
P21107	102	Tropomyosin alpha-3 chain	1.000	194.37	IQLVEEELDR(1)AQER
P21107	106	Tropomyosin alpha-3 chain	0.901	208.13	IQLVEEELDR(0.099)AQER(0.901)
P19123	147	Troponin C, slow skeletal and cardiac muscles	1.000	202.46	DGDKNNDGR(1)IDYDEFLEFMK
P50752	243	Troponin T, cardiac muscle	1.000	220.82	ALAIDHLNEDQLR(1)EK
P50752	126	Troponin T, cardiac muscle	1.000	151.42	DLNELQTLIEAHFENR(1)K
P50752	97	Troponin T, cardiac muscle	0.992	52.36	LFMPNLVPPKIPDGER(0.992)VDFDDIHR(0.008)K
P05213	64	Tubulin alpha-1B chain	1.000	147.26	HVPR(1)AVFVDLEPTVIDEVR
P05213	221	Tubulin alpha-1B chain	1.000	147.09	NLDIER(1)PTYTNLNR
P05213	215	Tubulin alpha-1B chain	0.997	90.71	R(0.997)NLDIER(0.003)PTYTNLNR
P05213	339	Tubulin alpha-1B chain	1.000	132.79	R(1)SIQFVDWCPTGFK
P05213	422	Tubulin alpha-1B chain	0.999	42.92	R(0.001)AFVHWYVGEGMEEGEFSEAR(0.999)EDMAALEK
P05214	339	Tubulin alpha-3 chain	1.000	118.72	R(1)TIQFVDWCPTGFK
Q9CWF2	156	Tubulin beta-2B chain	0.840	53.63	IR(0.84)EEYPDR(0.16)IMNTFSVMPSPK
Q9CWF2	162	Tubulin beta-2B chain	0.799	72.30	IR(0.201)EEYPDR(0.799)IMNTFSVMPSPK
Q9ERD7	86	Tubulin beta-3 chain	1.000	91.64	SGAFGHLFR(1)PDNFIFGQSGAGNNWAK
P68372	2	Tubulin beta-4B chain	1.000	215.08	MR(1)EIVHLQAGQCGNQIGAK
P68372	241	Tubulin beta-4B chain	0.870	76.05	LTTPTYGDLNHLVSATMSGVTTCLR(0.87)FPGQLNADLR(0.13)K
P68372	86	Tubulin beta-4B chain	1.000	132.84	SGPFGQIFR(1)PDNFVFGQSGAGNNWAK
P68372	380	Tubulin beta-4B chain	1.000	128.01	R(1)ISEQFTAMFR
Q7TQD2	164	Tubulin polymerization-promoting protein	1.000	157.57	AVSSPTVSR(1)LTDTSK
Q7TQD2	29	Tubulin polymerization-promoting protein	1.000	46.86	R(1)LSLESEGANEGATAAPELSALEEAFRR
Q7TQD2	55	Tubulin polymerization-promoting protein	0.820	88.57	RLSLESEGANEGATAAPELSALEEAFR(0.82)R(0.18)
Q9R0P9	153	Ubiquitin carboxyl-terminal hydrolase isozyme L1	1.000	255.42	NEAIQAAHDSVAQEGQCR(1)VDDK
Q3TEI4	5	Uncharacterized protein C15orf39 homolog	1.000	47.08	AEKR(1)PLGPLGPMMYGK
Q3UHX9	48	Uncharacterized protein C9orf114 homolog	1.000	166.16	QR(1)AQEEEAK
P46735	159	Unconventional myosin-Ib	1.000	179.59	TVR(1)NDNSSR
Q64331	199	Unconventional myosin-VI	1.000	145.09	TVR(1)NNNSSR
O35488	408	Very long-chain acyl-CoA synthetase	1.000	87.65	DEPVR(1)DANGYCIKVPK
O35488	329	Very long-chain acyl-CoA synthetase	1.000	89.47	YLCNTPQKPNDR(1)DHK
P50544	616	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	1.000	67.83	IR(1)ENMASLQSSPQHQELFR
P63044	31	Vesicle-associated membrane protein 2	1.000	130.27	R(1)LQQTQAQVDEVVDIMR
P63044	30	Vesicle-associated membrane protein 2	0.769	44.56	SATAATVPPAAPAGEGGPPAPPPNLTSNR(0.769)R(0.231)
P63044	66	Vesicle-associated membrane protein 2	1.000	110.60	LSELDDR(1)ADALQAGASQFETSAAK
P46460	415	Vesicle-fusing ATPase	1.000	110.74	MR(1)GHQLLSADVDIK
P46460	533	Vesicle-fusing ATPase	1.000	109.42	NSDR(1)TPLVSVLLEGPPHSGK
P46460	337	Vesicle-fusing ATPase	1.000	91.10	QR(1)GSMAGSTGVHDTVVNQLLSK
P46460	67	Vesicle-fusing ATPase	1.000	124.19	THPSVVPGCIAFSLPQR(1)K
P46460	303	Vesicle-fusing ATPase	1.000	129.77	YVGESEANIR(1)K
Q8VDJ3	912	Vigilin	1.000	49.62	DYNVQIKFPDR(1)EENPVHSVEPSIQENGDEAGEGR
P20152	207	Vimentin	0.982	113.13	EEAESTLQSFR(0.982)QDVDNASLAR(0.018)
P20152	304	Vimentin	0.999	133.21	FADLSEAANR(0.999)NNDALR(0.001)
P20152	23	Vimentin	0.925	112.82	MFGGSGTSSR(0.925)PSSNR(0.075)
P20152	28	Vimentin	1.000	185.72	MFGGSGTSSRPSSNR(1)SYVTTSTR
P20152	159	Vimentin	1.000	116.19	R(1)QVDQLTNDK
P20152	450	Vimentin	1.000	205.24	TVETR(1)DGQVINETSQHHDDLE
Q64727	341	Vinculin	1.000	78.33	AR(1)GQGASPVAMQK
Q64727	547	Vinculin	1.000	121.25	CDR(1)VDQLTAQLADLAAR
Q64727	428	Vinculin	0.995	92.47	IAELCDDPKER(0.995)DDILR(0.005)

Q64727	684	Vinculin	1.000	79.68	ILLR(1)NPGNQAAYEHFETMK
Q64727	538	Vinculin	1.000	102.07	LANVMMGPYR(1)QDLLAK
P62761	150	Visinin-like protein 1	1.000	154.19	MNEDGLTPEQR(1)VDK
Q60932	133	Voltage-dependent anion-selective channel protein 1	1.000	121.94	R(1)EHINLGCDVDFDIAGPSIR
Q60932	176	Voltage-dependent anion-selective channel protein 1	1.000	181.72	SR(1)VTQSNFAVGYK
Q60930	264	Voltage-dependent anion-selective channel protein 2	1.000	172.74	VNNSSLIGVGYTQTLR(1)PGVK
Q60931	252	Voltage-dependent anion-selective channel protein 3	1.000	187.21	VNNASLIGLGYTQTLR(1)PGVK
P50516	215	V-type proton ATPase catalytic subunit A	1.000	173.36	QVR(1)PVTEKLPANHPLLTGQR
P50516	308	V-type proton ATPase catalytic subunit A	1.000	91.97	R(1)TALVANTSNMPVAAR
P62814	82	V-type proton ATPase subunit B, brain isoform	1.000	164.53	R(1)SGQVLEVSGSK
P50518	222	V-type proton ATPase subunit E 1	1.000	138.45	GALFGANANR(1)K
Q9WTT4	48	V-type proton ATPase subunit G 2	0.905	172.43	EEAQMEVEQYR(0.905)R(0.095)
Q9WTT4	107	V-type proton ATPase subunit G 2	1.000	165.85	VLAQLLGMVCEVR(1)PQVHPNYR
O88342	330	WD repeat-containing protein 1	1.000	77.75	SIQCLTVHR(1)NGGK
Q8BJH1	249	Zinc finger C2HC domain-containing protein 1A	1.000	82.09	ARNTTPPSLAR(1)NSVAGVLTNK
Q8BJH1	226	Zinc finger C2HC domain-containing protein 1A	1.000	41.89	ASSVNSPLGNKPQTLSPSHR(1)AIAAPQAGANTK
Q8K0C5	37	Zymogen granule membrane protein 16	1.000	64.45	R(1)FSHSGNQLDGPITAFR
P07146	95	Anionic trypsin-2	1.000	131.33	IIR(1)HPNYNSWTLDNDIMLIK
P10107	97	Annexin A1	1.000	104.84	AAYLQENGKPLDEVLR(1)K
Q9DBG3	127	AP-2 complex subunit beta	1.000	121.24	ITEYLCEPLR(1)K
P08226	214	Apolipoprotein E	0.877	108.71	TANLGAGAAQPLR(0.877)DR(0.123)
P55088	284	Aquaporin-4	1.000	44.53	GSYMEVEDNR(1)SQVETEDLILKPGVVHVIDIDRGEEK
Q61176	92	Arginase-1	1.000	139.22	NGR(1)VSVVLGGDHSLAVGSISGHAR

Table S2. List of all identified homocitrullinated proteins and sites from five brain regions and six body organs in mice. Homocitrullination localization probability, MaxQuant (Andromeda) searching score and peptide sequence for each site are also shown.

Protein	Position	Protein name	Localization probability	Andromeda score	Peptide sequence
P16330	254	2',3'-cyclic-nucleotide 3'-phosphodiesterase	1.000	184.20	RPPGVLHCTTK(1)FCDYGK
P16330	243	2',3'-cyclic-nucleotide 3'-phosphodiesterase	1.000	121.10	LELVSYFGK(1)RPPGVLHCTTK
Q60597	697	2-oxoglutarate dehydrogenase, mitochondrial	1.000	114.70	HHVLHDQNVDK(1)R
P63038	473	60 kDa heat shock protein, mitochondrial	1.000	113.40	RALK(1)IPAMTIAK
P20029	602	Endoplasmic reticulum chaperone BiP	1.000	147.41	AVEEK(1)IEWLESHQDADIEDFKAK
Q99KI0	138	Aconitate hydratase, mitochondrial	1.000	75.70	VAVPSTIHCDHLIEAQVGGEK(1)DLRR
P63260	50	Actin, cytoplasmic 2	1.000	106.28	HQGVMVGMGQK(1)DSYVGDEAQSK
P50712	64	Alpha-defensin 14	1.000	74.79	DLVCYCRTRGCK(1)R
Q8BVM7	213	Amyotrophic lateral sclerosis 2 chromosomal region candidate gene 12 protein homolog	0.953	55.71	ALAEEKLDKMSK(0.953)EYK
Q8BVM7	207	Amyotrophic lateral sclerosis 2 chromosomal region candidate gene 12 protein homolog	0.946	55.71	ALAEEK(0.946)LDKMSKEYK
Q8BVM7	210	Amyotrophic lateral sclerosis 2 chromosomal region candidate gene 12 protein homolog	0.946	55.71	ALAEEKLDK(0.946)MSKEYK
Q9D4H4	636	Angiomotin-like protein 1	1.000	54.96	LQQALTQLQSACEK(1)RGQMER
P05202	309	Aspartate aminotransferase, mitochondrial	1.000	129.80	RVESQLK(1)ILIRPLYSNPPLNGAR
Q9CQQ7	225	ATP synthase F(0) complex subunit B1, mitochondrial	1.000	187.37	HVVK(1)SISVQQEK
Q03265	305	ATP synthase subunit alpha, mitochondrial	1.000	124.69	DNGK(1)HALIIYDDLSK
Q03265	506	ATP synthase subunit alpha, mitochondrial	1.000	120.13	ITK(1)FENAFLSHVISQHQSLLGNIR
P56480	124	ATP synthase subunit beta, mitochondrial	1.000	104.14	GQK(1)VLDSGAPIKIPVGPETLGR
P56480	259	ATP synthase subunit beta, mitochondrial	0.994	131.84	TREGNDLYHEMIESGVINLK(0.994)DATSK(0.006)
Q76K27	202	Beta-galactoside alpha-2,6-sialyltransferase 2	1.000	95.20	AFLYRLWK(1)GAVSSK
Q9JI08	175	Bridging integrator 3	1.000	171.34	EDFEAK(1)NK
Q64338	588	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1C	1.000	85.45	KGDNPRGK(1)NSK
Q9D8B3	6	Charged multivesicular body protein 4b	1.000	44.97	SVFGK(1)LFGAGGGK
Q8K2X3	221	CST complex subunit STN1	1.000	42.31	DNLDLAGLTSLLSEK(1)IK
Q9CZ13	111	Cytochrome b-c1 complex subunit 1, mitochondrial	1.000	94.80	NRPGNALEK(1)EVESIGAHLNAYSTR
P99028	83	Cytochrome b-c1 complex subunit 6, mitochondrial	1.000	163.51	DHCVAHK(1)LFK
P97427	426	Dihydropyrimidinase-related protein 1	1.000	48.27	SHK(1)STVEYNIFEGMECHGSPLVVISQGK
O35493	403	Dual specificity protein kinase CLK4	1.000	42.12	EHLAMMERILGPIPAHMIQK(1)TRK
A3KGK3	767	Fer-1-like protein 4	1.000	53.57	VAERLDHGLQEVEK(1)MQRR
P97807	180	Fumarate hydratase, mitochondrial	0.989	126.14	K(0.011)PVHPNDHVNK(0.989)SQSSNDTFPTAMHIAAAVEVHK
Q69ZL1	1062	FYVE, RhoGEF and PH domain-containing protein 6	0.976	78.08	VFLK(0.976)EGTLMKLSRK
Q9JIB4	187	General transcription factor IIH subunit 2	1.000	43.31	EVLIIFSSLTTCDPSNIYDLIK(1)TLKTAKIR
Q61586	172	Glycerol-3-phosphate acyltransferase 1, mitochondrial	1.000	61.48	K(1)ILQEMVATVSPGMIR
P01942	41	Hemoglobin subunit alpha	1.000	138.09	MFASFPTTK(1)TYFPHFDVSHGSAQVK
P01942	57	Hemoglobin subunit alpha	0.999	190.24	TYFPHFDVSHGSAQVK(0.999)GHGK(0.001)
P02088	96	Hemoglobin subunit beta-1	1.000	192.01	GTFASLSELHCDK(1)LHVDPENFR
P02088	67	Hemoglobin subunit beta-1	1.000	83.40	K(1)VITAFNDGLNHLDSLK
Q8CGP2	86	Histone H2B type 1-P	1.000	127.56	LAHYNK(1)R
Q8CGP2	109	Histone H2B type 1-P	1.000	110.40	LLLPGELAK(1)HAVSEGTK
Q8CGP2	117	Histone H2B type 1-P	1.000	118.20	HAVSEGTK(1)AVTK
P84244	123	Histone H3.3	1.000	122.97	RVTIMPK(1)DIQLAR
P62806	92	Histone H4	1.000	140.22	TVTAMDVVYALK(1)R
P62806	78	Histone H4	0.830	181.58	DAVTYTEHAK(0.83)RK(0.17)
Q5U5V2	217	Hydroxylysine kinase	1.000	55.66	MFKEEVMTK(1)LSHFR
Q80V26	296	Inositol monophosphatase 3	1.000	119.87	K(1)WDICAGNAILK
Q9D9K8	40	IQ domain-containing protein F1	0.965	155.48	K(0.035)K(0.965)QELSEK
P54071	180	Isocitrate dehydrogenase [NADP], mitochondrial	1.000	145.77	HAHGDQYK(1)ATDFVVDR
Q99JN2	447	Kelch-like protein 22	1.000	54.52	KEVYAHAGTTLQGK(1)MYITCGR
P02468	1591	Laminin subunit gamma-1	0.802	111.12	DIHNLEDIK(0.802)K(0.198)
P51174	322	Long-chain specific acyl-CoA dehydrogenase, mitochondrial	1.000	91.41	AFGK(1)TVAHIQTVQHK
P08249	165	Malate dehydrogenase, mitochondrial	1.000	115.54	HGVYNPNK(1)IFGVTTLDIVR
P04370	145	Myelin basic protein	1.000	171.04	SK(1)YLATASTMDHAR
P04370	205	Myelin basic protein	1.000	181.40	TTHYGSLPQK(1)SQHGR
P04247	80	Myoglobin	1.000	150.06	K(1)GQHAAEIQPLAQSHATK
P04247	64	Myoglobin	1.000	162.32	K(1)HGCTVLTALGTILKK
P04247	43	Myoglobin	0.996	107.98	THPETLDK(0.996)FDK(0.001)FK(0.002)
Q02566	665	Myosin-6	1.000	139.88	LMTNLK(1)TTHPHFVR
Q02566	965	Myosin-6	0.987	111.65	KDIDDLELTLAK(0.987)VEK(0.01)EK(0.003)
Q02566	759	Myosin-6	0.977	84.33	LLGSLDIDHNQYK(0.977)FGHTK(0.023)
Q02566	1897	Myosin-6	0.815	171.29	RQAEEAEEQANTNLSK(0.815)FRK(0.185)

P70441	154	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	1.000	98.55	K(1)GPNGYGFNLHSDK
P08553	166	Neurofilament medium polypeptide	1.000	147.49	ATLEMVNHEK(1)AQVQLDSDHLEEDIHR
P08553	116	Neurofilament medium polypeptide	1.000	118.19	FAGYIEK(1)VHYLEQQNK
Q60974	2220	Nuclear receptor corepressor 1	1.000	87.52	LESTSPMVK(1)SK(1)K
Q60974	2222	Nuclear receptor corepressor 1	1.000	87.52	LESTSPMVK(1)SK(1)K
O70250	100	Phosphoglycerate mutase 2	1.000	92.56	HYGGLTGLNK(1)AETAAK
P97350	50	Plakophilin-1	1.000	57.48	TGTSGRQRVQEQVMMTVK(1)R
Q922K7	9	Probable 28S rRNA (cytosine-C(5))-methyltransferase	1.000	144.10	KLDPTK(1)KEK
Q06186	79	Proheparin-binding EGF-like growth factor	1.000	81.48	VAFSSK(1)PQGLATPSKERNGK
E9PVX6	1609	Proliferation marker protein Ki-67	1.000	58.08	IILRK(1)MDVTEEISGLWK
Q6PDI5	1659	Proteasome-associated protein ECM29 homolog	1.000	83.82	K(1)NSLESMGVRTTK
Q8C1D8	756	Protein IWS1 homolog	1.000	48.91	SAHAVK(1)ISIEGNK
Q91YM4	493	FAST kinase domain-containing protein 4	1.000	58.98	MTPLQK(1)ELQETLK
P13405	918	Retinoblastoma-associated protein	0.998	78.81	QRMNESK(0.002)DVSNK(0.998)EEK
Q8R527	11	Rho-related GTP-binding protein RhoQ	0.999	40.53	MAHGPGALMLK(0.999)CVVVGDGAVGK(0.001)
Q9CQF8	99	Ribosomal protein 63, mitochondrial	1.000	69.49	YIADQLDHLNISK(1)K
P0DP99	36	Short transmembrane mitochondrial protein 1	0.836	124.07	LEEIK(0.082)K(0.836)DLEAK(0.082)
P09671	68	Superoxide dismutase [Mn], mitochondrial	1.000	185.45	HHAAYVNNLNATEEK(1)YHEALAK
P70281	194	Synaptonemal complex protein 3	1.000	79.49	MFAMK(1)QIHEQFIK
P58771	152	Tropomyosin alpha-1 chain	1.000	103.60	EAK(1)HIAEDADRK
P05213	394	Tubulin alpha-1B chain	1.000	141.85	LDHK(1)FDLMYAK
Q8CE72	448	Ciliogenesis and planar polarity effector 1	1.000	65.95	LEK(1)AYQSMMLSEPK

Chapter 4

12-plex DiLeu isobaric labeling enabled high-throughput investigation of citrullination alterations in DNA damage response

Adapted from: **Li, Z.**; Wang, B.; Yu, Q.; Shi, Y.; Li, L., 12-plex DiLeu isobaric labeling enabled high-throughput investigation of citrullination alterations in DNA damage response. **2021.** Under review.

Abstract

Protein citrullination is a key post-translational modification (PTM) which leads to the loss of positive charge on arginine and consequent protein structural and functional changes. Though it has been indicated to play critical roles in various physiological and pathological processes, effective analytical tools are largely limited due to a few challenges such as the small mass shift induced by this PTM and its low-abundance nature. Recently, we developed a biotin thiol tag which enabled large-scale profiling of protein citrullination from complex biological samples via mass spectrometry. However, a high-throughput quantitative approach is still in great need to further improve the understanding of this PTM. In this study, we report an efficient pipeline using our custom developed *N*,*N*-dimethyl leucine isobaric tags to achieve a multiplexed quantitative analysis of citrullination from up to twelve samples for the first time. We then apply this strategy to investigating citrullination alterations in response to DNA damage stress using human cell lines. We unveil important biological functions regulated by protein citrullination and observe hypercitrullination on RNA binding proteins and DNA repair proteins, specifically. Our results reveal the involvement of citrullination in DNA damage pathways and may provide new insights into DNA damage-related disease pathogenesis.

Introduction

Protein citrullination is a post-translational modification (PTM) resulting from the conversion of peptidyl arginine to citrulline, which is catalyzed by calcium-regulated protein arginine deiminases (PADs) (**Figure S1a**).[1-3] Consequently, the loss of positive charges will significantly affect protein conformations, protein-protein interactions and protein functions.[1-2] Protein citrullination has been indicated to participate in various physiological functions such as gene regulation[4-8] and apoptosis[3, 9]. Dysregulation of this PTM was also associated with the development and progression of many diseases including rheumatoid arthritis[10-14], multiple sclerosis[15-17], Alzheimer's disease[18-20] and cancers[21-24].

Despite fast-growing interests in understanding the physiological and pathological involvement of protein citrullination, the development of analytical tools to detect this PTM lags far behind. On one hand, antibody-based methods such as western blotting and immunohistochemistry are prevalent in detecting citrullination with good sensitivity and specificity.[25-27] However, these methods are intrinsically not capable for large-scale analysis and confident localization of modification sites.[28-29] On the other hand, mass spectrometry (MS)-based strategies, which are emerging as powerful tools in numerous PTM-related studies, still face great challenges in profiling citrullination[30]. This is largely due to a very small mass increment induced by citrullination compared to its unmodified counterpart which can be easily confused with deamidation or ^{13}C isotopic peaks in a mass spectrum. In addition, the signals of low-abundance citrullinated peptides in the analytes are largely suppressed by other high-intensity molecules without effective enrichment methods.[29, 31] To deal with these difficulties, various directions have been explored such as carefully examining the spectra[32-33], applying more restrict criteria or novel algorithms during data searching[34-36], and developing chemical derivatization[37-38] and enrichment

methods[39-42]. Unfortunately, very few studies have provided easy yet effective means to characterize protein citrullination and even fewer achieved quantitative analysis. Most of these previous quantitative analyses utilized label-free approaches which usually suffers from large run-to-run variations and requires much longer instrument time.[43] Alternatively, stable isotope labeling especially isobaric labeling techniques, such as tandem mass tags (TMT)[44-45] and *N,N*-dimethyl leucine tags (DiLeu)[46-47], have become popular for comparative studies of different biological states. Although these strategies were shown to generate accurate quantitative results with great multiplexing capability and analytical throughput in many proteomics[48-49] and PTM-related work[50], they have not been successfully applied to citrullination analysis to date.

Recently we developed a novel biotin thiol tag which can specifically react with citrulline residues, not only inducing a larger mass increment for more confident characterizations but also enabling effective enrichment using biotin-avidin interactions (**Figure S1b, c**).[51] We further developed a robust pipeline using bottom-up proteomics approaches and achieved global profiling of protein citrullination. Herein, we extend the utility of this method by combing it with our 12-plex DiLeu labeling technique and achieved a high-throughput quantitative analysis of citrullination from different complex samples for the first time. We also apply it to investigating how protein citrullination changes in DNA damage response (DDR) using human cell lines, which may provide new insights into DDR-mediated cancer etiology and progression.

Methods

Full descriptions about chemicals and materials, synthesis of biotin thiol tag as well as the rest of the procedures can be found in the **Supplemental Information**.

Cell culture and treatment to induce DNA damage. Human MCF7 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 1 unit/mL penicillin and 1 mg/mL streptomycin according to the supplier's specifications at 37 °C with 5% CO_2. Different treatment methods to induce DNA damage are described as below. For ultraviolet (UV) radiation, cells were irradiated at 50 mJ/cm^2 using a UV light with a wavelength of 254 nm for 1 h and rested for 1 h before harvesting. For oxidative DNA damage, hydrogen peroxide (H_2O_2) was added into the medium at a final concentration of 0.6 mM and cells were incubated for 2 h before harvesting. For anthracycline-induced DNA damage, doxorubicin hydrochloride (ADR) was added into the medium at a final concentration of 2 μM and cells were incubated for 2 h before harvesting. For harvesting, the cultures were treated with 0.05% Trypsin-EDTA at 37 °C and 5% CO_2 for 3 min before transferred to 1× phosphate-buffered saline (PBS) and centrifuged at 300 ×g for 5 min. Cell pellets were washed with PBS buffer twice to avoid contamination from the medium. Each treated group was prepared as biological triplicates.

Protein extraction and digestion. Mouse tissues or cells were dissolved in extraction buffer solution (4 % sodium dodecyl sulfate, 50 mM Tris buffer, pH adjusted to 8 with hydrochloric acid) and sonicated using a probe sonicator (Thermo Scientific, San Jose, CA). 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 ℃) overnight. Samples were centrifuged at 14,000 ×g for 15 min after which supernatant 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 LysC/trypsin mixture in a 50:1 ratio

(protein:enzyme, w/w) at 37 ℃ overnight. The digestion was quenched with 1% trifluoracetic acid (TFA) and samples were desalted with Sep-Pak C18 cartridges. Concentrations of peptide mixture were measured by peptide assay. One hundred microgram of peptide was aliquoted for each sample, dried in vacuo and reconstituted in 0.5 M triethylammonium bicarbonate prior to DiLeu labeling.

12-plex DiLeu labeling. Synthesis of DiLeu tags and labeling process were performed according to protocols previously described. One milligram of each DiLeu tag was dissolved in 100 µL of anhydrous *N*,*N*-dimethylformamide and combined with 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium tetrafluoroborate and *N*-methylmorpholine at 0.7× molar ratios. The activation was performed by vortexing the mixture for 45 min at room temperature and supernatant was added to each sample for peptide labeling. After vortexing at room temperature for 2 h, the labeling reaction was quenched by addition of hydroxylamine to a concentration of 0.25%. The samples were then dried in vacuo, combined, and cleaned with strong cation exchange (SCX) SpinTips according to the manufacturer's protocols.

Derivatization of citrullinated peptides using biotin thiol tag. DiLeu-labeled samples and 300 µg of biotin thiol tag were dissolved with 40 µL 12.5% TFA. A solution of 2,3-butanedione was prepared by mixing 1 µL of 2,3-butanedione with 114 µL 12.5% TFA. Ten microliters of 2,3-butanedione solution was subsequently added to initiate the derivatization reaction and the mixture was vortexed in dark at 37 °C for 6 h. SCX clean-up was then performed to remove the excess tag and samples were dried in vacuo.

Enrichment of derivatized citrullinated peptides. The enrichment process was performed as previously described with slight modifications.[52] Briefly, 75 µL streptavidin agarose was washed with 1 mL 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% acetonitrile (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% formic acid (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.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Samples were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer coupled to a Dionex UltiMate 3000 UPLC system (Thermo Scientific, San Jose, CA). Each sample was dissolved in 8% ACN, 0.1% FA in water before loaded onto a 75 μm inner diameter homemade microcapillary column which was packed with 15 cm of Bridged Ethylene Hybrid C18 particles (1.7 μm, 130 Å) 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 8% to 37% 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 with an AGC target of 2×10^5 and maximum injection time of 100 ms. 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 Da, a resolving power of 60k, an AGC target of 5×10^4, a maximum injection time of 118 ms, and a lower mass limit of 110 *m/z*. Precursors were subject to dynamic exclusion for 45 s with a 10-p.p.m. tolerance. Each sample was acquired in technical triplicates.

Data analysis. Raw files were searched against the UniProt *Mus musculus* (December 2018) or *Homo sapiens* (July 2019) reviewed database using MaxQuant (version 1.5.2.8) with trypsin/P

selected as the enzyme and three missed cleavages allowed. Reporter ion MS2 was set as quantification type and DiLeu labeling on peptide N-termini and lysine residues were set as isobaric labels. Reporter mass tolerance was set as 0.003 Da. Carbamidomethylation of cysteine residues (+57.02146 Da) was chosen as the fixed modification and variable modifications included oxidation of methionine residues (+15.99492 Da), acetylation at protein N-termini (+42.01056 Da) and biotin tag-labeled citrullination of arginine (+354.10718 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. 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 site localization probability threshold was set to 0.75. All other parameters were set as default. C-terminal citrullination identifications were filtered out as well. Two-sample Student's *t* test with a two-tailed distribution for binary comparison and one-way ANOVA analysis were conducted using Perseus (version 1.6.0.7).[53] Bioinformatics analyses including hierarchical clustering, volcano plots, and box plots were achieved using R packages. Biological process network was generated using Metascape (version 3.5)[54] and exported using Cytoscape (version 3.7.1).

Results and discussion

Our DiLeu isobaric reagents feature compact structures (**Figure S2a**), higher intensity of generated reporter ions compared to commercial alternatives and improved collision-induced fragmentation of labeled peptides, all of which contribute to increased confidence in quantitative accuracy and peptide sequence identification. Therefore, we also anticipate the DiLeu labeling to be an attractive means for accurate quantitative citrullination analysis of multiple samples in parallel (**Figure S2b**).

Test of reaction efficiency and quantitation accuracy using citrullinated peptide standard. We first performed a proof-of-principle test using a citrullinated peptide standard (SAVRACitSSVPGVR) containing one citrullination site within the sequence (**Figure 1a**). After reacting with one channel of DiLeu tag for 2 h, we observed a clean peak corresponding to the addition of DiLeu to the peptide N-terminus, suggesting the high labeling efficiency as expected (**Figure 1b**). We then performed SCX clean-up to remove the excess tag and derivatized the DiLeu-labeled peptide standard using our recently developed biotin thiol tag. The prominent peak in **Figure 1c** corresponds to the addition of biotin thiol tag to the citrulline residue and the lower peak results from the loss of biotin moiety as discussed previously.[51] It is due to the in-source fragmentation when using a matrix-assisted laser desorption/ionization (MALDI) source instead of incomplete reaction, which can be further evidenced by the presence of this peak at a similar percentage after enrichment. Our results indicate that DiLeu labeling and biotin tag derivatization can be performed sequentially without observable interference to each other, which lays the foundation for any further applications.

To explore the fragmentation and quantitation performance, we differentially labeled the citrullinated peptide standard in two known ratios 1:1:1:1 and 1:1:2:5 using a select set of 4-plex DiLeu tags (115a, 116a, 117a and 118a). We then combined these labeled samples and derivatized them with biotin thiol tag. Upon high-energy collisional dissociation (HCD) fragmentation, the peptides generated a wealth of b- and y-ion series which enables annotation of the peptide sequence and localization of citrullination sites with high confidence (**Figure 2**). In the low-mass range, high-intensity DiLeu reporter ions were also observed for relative quantification. **Figure 2** also shows the example spectra of reporter ions from 1:1:1:1 and 1:1:2:5 samples, indicating great quantitation accuracy when using simple peptide standard.

Streamlined workflow of multiplexed quantitative citrullination analysis from complex biological samples. We then tried to streamline the workflow of quantitative citrullination analysis from different complex biological samples by the integration of bottom-up proteomics sample preparation method, DiLeu isobaric labeling and biotin thiol tag-enabled citrullination profiling strategies. Proteins were extracted from biological samples and enzymatically digested to peptides, typically using trypsin. Multiplexed DiLeu isobaric labeling was performed followed by necessary sample clean-up steps. Labeled peptides were then incubated with biotin thiol tag under acidic conditions and excess tag was removed by SCX. Derivatized citrullinated peptides were enriched by streptavidin agarose and then released for LC-MS/MS analysis and data processing using commercially available software (**Figure S3a**). Similarly, we differentially labeled mouse brain digest in two ratios 1:1:1:1 and 1:1:2:5, and conducted quantitative citrullination analysis on them. We then showed the measured ratios for all quantified citrullinated peptides as depicted by the boxplots (**Figure S3a, b**). Despite good precision, the quantitation accuracy is relatively poor, which is possibly due to a lack of intensity normalization. In typical comparative proteomics analyses, sample amount determined by peptide concentration is usually measured through colorimetric assays prior to the labeling step. Such measurements together with the sample handling can lead to systematic bias in the actual amount of each sample. Therefore, MS proteomics data are commonly normalized based on total intensity of all identified peptides or proteins. However, this method is not suitable for our citrullination analysis since the sum intensities of the enriched samples are uncertain. To deal with the issue, we proposed to add another step of analysis of DiLeu-labeled peptides and the intensities of non-enriched proteins could be used for normalizing the citrullination data (**Figure 3**). We first tested the normalization using total intensities of identified proteins but only observed slight improvement of quantitation

accuracy (data not shown). We reasoned that some proteins with extreme values might deviate the total intensity and compromise its effectiveness in good normalization. Alternatively, we chose to use median protein intensity and this strategy produced much improved accuracy and precision of the detected quantitative ratios for identified citrullinated peptides (**Figure 4**). Across all channels, the median ratios measure within 5% of the expected values with average coefficients of variation below 11%. Altogether, our results indicate a highly effective and accurate quantification method for protein citrullination analysis and this pipeline can be easily extended to process up to 12 samples in parallel.

High-throughput investigation of citrullination changes in response to DNA damage stress. DNA damage is well recognized as a critical factor in cancer development and progression,[55] and understanding the mechanisms of DDR can facilitate studies in cancer pathogenesis and targeted therapeutic approaches.[56] Previous studies have shown DNA damage-induced hypercitrullination but they only identified few specific proteins such as histone H4, lamin C and nucleophosmin[57-58] due to the limitations in analytical tools. Here, we adapt our method using 12-plex DiLeu tags to quantitatively investigate how protein citrullination changes in response to DNA damage in a larger scale and aim to provide deeper insights into the DDR pathways.

We used three different methods to introduce DNA damage to human MCF7 cells which included UV radiation, oxidative stress using H_2O_2 and anthracycline drug using ADR.[59] To ensure reproducible and unbiased analysis, we collected three biological replicates for each treated and control group, and the MS data were acquired in technical triplicates as well. In total, we identified and quantified 78 citrullination sites from 63 citrullinated proteins (**Table S1**). The tandem MS spectra of them contain a wealth of b- and y-ions, enabling confident peptide sequence identification and citrullination site localization. Twelve DiLeu reporter ion signals are fully

resolved in the low-mass range which allows for accurate quantification (**Figure S4**). We then performed hierarchical clustering of all quantified citrullination sites to explore their profiles in different groups (**Figure 5a**). This heatmap illustrates column-wise clustering of biological replicates in either treated or control group, suggesting larger intergroup differences than intragroup variations and thus demonstrating again the capability of our methods for reliable quantitative analysis of protein citrullination. Control group and H_2O_2-treated group show similar profiles, which is likely due to the less severe oxidative stress under the low concentration of H_2O_2 we used in the experiment. On the other hand, both UV and ADR-treated groups exhibit dramatic alterations compared to the control while obvious disparities are also observed between them, indicating different mechanisms between these two types of DNA damage. ANOVA analysis (FDR 0.05) reveals 31sites from 26 citrullinated proteins to be significantly changed among these groups (**Table S1**). To further decipher what functions these significantly changed proteins are involved in, we generated a network of enriched biological processes using Metascape[54] (**Figure. 5b**). Each node refers to an enriched term and different terms are grouped into clusters based on their similarities while the most statistically significant term represents the cluster name. We found that these proteins are associated with several important processes including regulation of DNA metabolic process and translational activities, modulation of protein structural changes and response to environmental stress, suggesting the potential critical roles of protein citrullination in DDR.

We then looked deeper into the difference between ADR-treated and control group which is of special interest since ADR is indeed one of the most effective chemotherapy drugs used against solid tumors in the treatment of several cancer types.[60] We found 24 citrullination sites from 20 proteins are significantly altered between them (two-sided *t* test, *p* value <0.05) **(Table**

S2) and many of them exhibit higher expression levels in the ADR-treated group (**Figure 5c**). For instance, citrullination on some DNA repair proteins (e.g., RAD50) is up regulated in the treated group. RAD50 is a major component of MRE11-RAD50-NBS1 complex and plays important roles in DNA double-strand break repair and telomere maintenance.[61-62] Higher citrullination level on these proteins may affect their structural integrity and thus impair their normal functions as DNA repair contributors. In addition, we also observed hypercitrullination on a few heterogeneous nuclear ribonucleoproteins which is a group of RNA-binding proteins that contribute to alternative splicing, mRNA stabilization, and transcriptional and translational regulation.[63-65] More interestingly, many of these citrullination sites are located in important domains such as RNA binding domain where positively charged amino acids, which form strong ionic hydrogen bonds (salt bridges), predominate these RNA-protein interactions.[66] Consequently, loss of positive charges due to excessive protein citrullination may negatively impact such interactions and the functions of this group of proteins. Collectively, our results indicate that protein citrullination actively participates in DDR mainly through interfering normal translational activities and DNA repair processes. One limitation of our approach is the relatively low identification rate of citrullination on highly basic histone proteins since the short peptides generated by trypsin digestion can easily get lost during sample cleanup or LC separation and therefore escape detection. To improve the analysis of these important proteins, an ongoing project in our lab aims to explore alternative digestion or top-down proteomics procedures.

Conclusions

Here, we demonstrate the compatibility of 12-plex DiLeu isobaric labeling with biotin thiol tag to achieve accurate quantitative analysis of protein citrullination, which represents the first high-throughput quantification method for this PTM. We apply this strategy to investigating

citrullination changes in response to DNA damage stress and find overexpressed protein citrullination on some DNA repair proteins and RNA binding proteins. Our results indicate an intimate involvement of this PTM in the regulation of translation and DNA repair processes, and provide new insights into DNA damage-associated disease pathogenesis. More importantly, this novel approach can be easily adapted by others and will greatly facilitate related studies to improve

Acknowledgements

This study was supported by grant funding from the NIH (R21AG060242). 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.

References

(1) Fuhrmann, J.; Clancy, K. W.; Thompson, P. R. *Chem. Rev.* **2015,** *115*, 5413-61.

(2) Fuhrmann, J.; Thompson, P. R. *ACS chemical biology* **2016,** *11*, 654-68.

(3) Gyorgy, B.; Toth, E.; Tarcsa, E.; Falus, A.; Buzas, E. I. *The international journal of biochemistry & cell biology* **2006,** *38*, 1662-77.

(4) Clancy, K. W.; Russell, A. M.; Subramanian, V.; Nguyen, H.; Qian, Y.; Campbell, R. M.; Thompson, P. R. *ACS chemical biology* **2017,** *12*, 1691-1702.

(5) Guo, Q.; Fast, W. *The Journal of biological chemistry* **2011,** *286*, 17069-78.

(6) Christophorou, M. A.; Castelo-Branco, G.; Halley-Stott, R. P.; Oliveira, C. S.; Loos, R.; Radzisheuskaya, A.; Mowen, K. A.; Bertone, P.; Silva, J. C.; Zernicka-Goetz, M.; Nielsen, M. L.; Gurdon, J. B.; Kouzarides, T. *Nature* **2014,** *507*, 104-8.

(7) Li, P.; Yao, H.; Zhang, Z.; Li, M.; Luo, Y.; Thompson, P. R.; Gilmour, D. S.; Wang, Y. *Mol Cell Biol* **2008,** *28*, 4745-58.

(8) Li, P.; Wang, D.; Yao, H.; Doret, P.; Hao, G.; Shen, Q.; Qiu, H.; Zhang, X.; Wang, Y.; Chen, G.; Wang, Y. *Oncogene* **2010,** *29*, 3153-62.

(9) Witalison, E.; Thompson, P.; Hofseth, L. *Current Drug Targets* **2015,** *16*, 700-710.

(10) Pruijn, G. J. *Front Immunol* **2015,** *6*, 192.

(11) Elkon, K. B. *Sci Transl Med* **2013,** *5*, 209fs39.

(12) Turunen, S.; Huhtakangas, J.; Nousiainen, T.; Valkealahti, M.; Melkko, J.; Risteli, J.; Lehenkari, P. *Arthritis Res Ther* **2016,** *18*, 239.

(13) Tilvawala, R.; Nguyen, S. H.; Maurais, A. J.; Nemmara, V. V.; Nagar, M.; Salinger, A. J.; Nagpal, S.; Weerapana, E.; Thompson, P. R. *Cell Chem Biol* **2018,** *25*, 691-704 e6.

(14) Fert-Bober, J.; Darrah, E.; Andrade, F. *Immunol Rev* **2020,** *294*, 133-147.

(15) Bradford, C. M.; Ramos, I.; Cross, A. K.; Haddock, G.; McQuaid, S.; Nicholas, A. P.; Woodroofe, M. N. *J Neuroimmunol* **2014,** *273*, 85-95.

(16) Moscarello, M. A.; Mastronardi, F. G.; Wood, D. D. *Neurochem Res* **2007,** *32*, 251-6.

(17) Yang, L.; Tan, D.; Piao, H. *Neurochem Res* **2016,** *41*, 1845-56.

(18) Ishigami, A.; Ohsawa, T.; Hiratsuka, M.; Taguchi, H.; Kobayashi, S.; Saito, Y.; Murayama, S.; Asaga, H.; Toda, T.; Kimura, N.; Maruyama, N. *Journal of neuroscience research* **2005,** *80*, 120-8.

(19) Acharya, N. K.; Nagele, E. P.; Han, M.; Coretti, N. J.; DeMarshall, C.; Kosciuk, M. C.; Boulos, P. A.; Nagele, R. G. *J Autoimmun* **2012,** *38*, 369-80.

(20) Ishigami, A.; Masutomi, H.; Handa, S.; Nakamura, M.; Nakaya, S.; Uchida, Y.; Saito, Y.; Murayama, S.; Jang, B.; Jeon, Y. C.; Choi, E. K.; Kim, Y. S.; Kasahara, Y.; Maruyama, N.; Toda, T. *Journal of neuroscience research* **2015,** *93*, 1664-74.

(21) Yuzhalin, A. E.; Gordon-Weeks, A. N.; Tognoli, M. L.; Jones, K.; Markelc, B.; Konietzny, R.; Fischer, R.; Muth, A.; O'Neill, E.; Thompson, P. R.; Venables, P. J.; Kessler, B. M.; Lim, S. Y.; Muschel, R. J. *Nat. Commun.* **2018,** *9*, 4783.

(22) Yuzhalin, A. E. *Cancer Res* **2019,** *79*, 1274-1284.

(23) Stadler, S. C.; Vincent, C. T.; Fedorov, V. D.; Patsialou, A.; Cherrington, B. D.; Wakshlag, J. J.; Mohanan, S.; Zee, B. M.; Zhang, X.; Garcia, B. A.; Condeelis, J. S.; Brown, A. M.; Coonrod, S. A.; Allis, C. D. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110*, 11851-6.

(24) Chang, X.; Han, J.; Pang, L.; Zhao, Y.; Yang, Y.; Shen, Z. *BMC Cancer* **2009,** *9*, 40.

(25) Senshu, T.; Sato, T.; Inoue, T.; Akiyama, K.; Asaga, H. *Anal. Biochem.* **1992,** *203*, 94-100.

(26) Nicholas, A. P.; King, J. L.; Sambandam, T.; Echols, J. D.; Gupta, K. B.; McInnis, C.; Whitaker, J. N. *The Journal of comparative neurology* **2003,** *459*, 251-66.

(27) Moelants, E. A.; Van Damme, J.; Proost, P. *PloS one* **2011,** *6*, e28976.

(28) Verheul, M. K.; van Veelen, P. A.; van Delft, M. A. M.; de Ru, A.; Janssen, G. M. C.; Rispens, T.; Toes, R. E. M.; Trouw, L. A. *Autoimmun Rev* **2018,** *17*, 136-141.

(29) Hensen, S. M.; Pruijn, G. J. *Mol Cell Proteomics* **2014,** *13*, 388-96.

(30) Vitorino, R.; Guedes, S.; Vitorino, C.; Ferreira, R.; Amado, F.; Van Eyk, J. E. *Journal of proteome research* **2020**.

(31) Clancy, K. W.; Weerapana, E.; Thompson, P. R. *Curr Opin Chem Biol* **2016,** *30*, 1-6.

(32) Lee, C. Y.; Wang, D.; Wilhelm, M.; Zolg, D. P.; Schmidt, T.; Schnatbaum, K.; Reimer, U.; Ponten, F.; Uhlen, M.; Hahne, H.; Kuster, B. *Mol Cell Proteomics* **2018,** *17*, 1378-1391.

(33) Raijmakers, R.; van Beers, J. J.; El-Azzouny, M.; Visser, N. F.; Bozic, B.; Pruijn, G. J.; Heck, A. J. *Arthritis Res Ther* **2012,** *14*, R114.

(34) Fert-Bober, J.; Venkatraman, V.; Hunter, C. L.; Liu, R.; Crowgey, E. L.; Pandey, R.; Holewinski, R. J.; Stotland, A.; Berman, B. P.; Van Eyk, J. E. *Journal of proteome research* **2019**.

(35) Wang, X.; Swensen, A. C.; Zhang, T.; Piehowski, P. D.; Gaffrey, M. J.; Monroe, M. E.; Zhu, Y.; Dong, H.; Qian, W. J. *Journal of proteome research* **2020,** *19*, 1863-1872.

(36) Villacres, C.; Spicer, V.; Krokhin, O. V. *Journal of proteome research* **2021**.

(37) Stensland, M.; Holm, A.; Kiehne, A.; Fleckenstein, B. *Rapid Commun Mass Spectrom* **2009,** *23*, 2754-62.

(38) De Ceuleneer, M.; De Wit, V.; Van Steendam, K.; Van Nieuwerburgh, F.; Tilleman, K.; Deforce, D. *Rapid Commun Mass Spectrom* **2011,** *25*, 1536-42.

(39) Tutturen, A. E.; Fleckenstein, B.; de Souza, G. A. *Journal of proteome research* **2014,** *13*, 2867-73.

(40) Tutturen, A. E.; Holm, A.; Fleckenstein, B. *Anal Bioanal Chem* **2013,** *405*, 9321-31.

(41) Choi, M.; Song, J. S.; Kim, H. J.; Cha, S.; Lee, E. Y. *Analytical biochemistry* **2013,** *437*, 62-7.

(42) Lewallen, D. M.; Bicker, K. L.; Subramanian, V.; Clancy, K. W.; Slade, D. J.; Martell, J.; Dreyton, C. J.; Sokolove, J.; Weerapana, E.; Thompson, P. R. *ACS chemical biology* **2015,** *10*, 2520-8.

(43) Megger, D. A.; Pott, L. L.; Ahrens, M.; Padden, J.; Bracht, T.; Kuhlmann, K.; Eisenacher, M.; Meyer, H. E.; Sitek, B. *Biochimica et biophysica acta* **2014,** *1844*, 967-76.

(44) Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. *Anal. Chem.* **2003,** *75*, 1895-904.

(45) Li, J.; Van Vranken, J. G.; Pontano Vaites, L.; Schweppe, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H.; Kuhn, K.; Pike, I.; Bomgarden, R. D.; Rogers, J. C.; Gygi, S. P.; Paulo, J. A. *Nature methods* **2020,** *17*, 399-404.

(46) Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L. *Anal. Chem.* **2010,** *82*, 2817-25.

(47) Frost, D. C.; Greer, T.; Li, L. *Anal. Chem.* **2015,** *87*, 1646-54.

(48) Li, Z.; Tremmel, D. M.; Ma, F.; Yu, Q.; Ma, M.; Delafield, D. G.; Shi, Y.; Wang, B.; Mitchell, S. A.; Feeney, A. K.; Jain, V. S.; Sackett, S. D.; Odorico, J. S.; Li, L. *Nat. Commun.* **2021,** *12*, 1020.

(49) Yu, Q.; Zhong, X.; Chen, B.; Feng, Y.; Ma, M.; Diamond, C. A.; Voeller, J. S.; Kim, M.; DeSantes, K. B.; Capitini, C. M.; Patel, N. J.; Hoover-Regan, M. L.; Burke, M. J.; Janko, K.; Puccetti, D. M.; Ikonomidou, C.; Li, L. *Journal of proteome research* **2020,** *19*, 2606-2616.

(50) Dieterich, I. A.; Cui, Y.; Braun, M. M.; Lawton, A. J.; Robinson, N. H.; Peotter, J. L.; Yu, Q.; Casler, J. C.; Glick, B. S.; Audhya, A.; Denu, J. M.; Li, L.; Puglielli, L. *Scientific reports* **2021,** *11*, 2013.

(51) Li, L.; Shi, Y.; Li, Z.; Wang, B.; Shi, X.; Ye, H.; Delafield, D.; Lv, L.; Ye, Z.; Chen, Z.; Ma, F. *Research Square* **2021**.

(52) Schiapparelli, L. M.; McClatchy, D. B.; Liu, H. H.; Sharma, P.; Yates, J. R., 3rd; Cline, H. T. *Journal of proteome research* **2014,** *13*, 3966-78.

(53) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. *Nature methods* **2016,** *13*, 731-40.

(54) Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. *Nat. Commun.* **2019,** *10*, 1523.

(55) Alhmoud, J. F.; Woolley, J. F.; Al Moustafa, A. E.; Malki, M. I. *Cancers (Basel)* **2020,** *12*.

(56) O'Connor, M. J. *Mol Cell* **2015,** *60*, 547-60.

(57) Tanikawa, C.; Espinosa, M.; Suzuki, A.; Masuda, K.; Yamamoto, K.; Tsuchiya, E.; Ueda, K.; Daigo, Y.; Nakamura, Y.; Matsuda, K. *Nat. Commun.* **2012,** *3*, 676.

(58) Tanikawa, C.; Ueda, K.; Nakagawa, H.; Yoshida, N.; Nakamura, Y.; Matsuda, K. *Cancer Res* **2009,** *69*, 8761-9.

(59) Chatterjee, N.; Walker, G. C. *Environ Mol Mutagen* **2017,** *58*, 235-263.

(60) Taymaz-Nikerel, H.; Karabekmez, M. E.; Eraslan, S.; Kirdar, B. *Scientific reports* **2018,** *8*, 13672.

(61) Rojowska, A.; Lammens, K.; Seifert, F. U.; Direnberger, C.; Feldmann, H.; Hopfner, K. P. *EMBO J* **2014,** *33*, 2847-59.

(62) Lamarche, B. J.; Orazio, N. I.; Weitzman, M. D. *FEBS Lett* **2010,** *584*, 3682-95.

(63) Geuens, T.; Bouhy, D.; Timmerman, V. *Hum Genet* **2016,** *135*, 851-67.

(64) Chen, Q.; Jin, M.; Zhu, J.; Xiao, Q.; Zhang, L. *BioMed research international* **2013,** *2013*, 623978.

(65) Lu, J.; Gao, F. H. *Biomed Rep* **2016,** *4*, 657-663.

(66) Corley, M.; Burns, M. C.; Yeo, G. W. *Mol Cell* **2020,** *78*, 9-29.

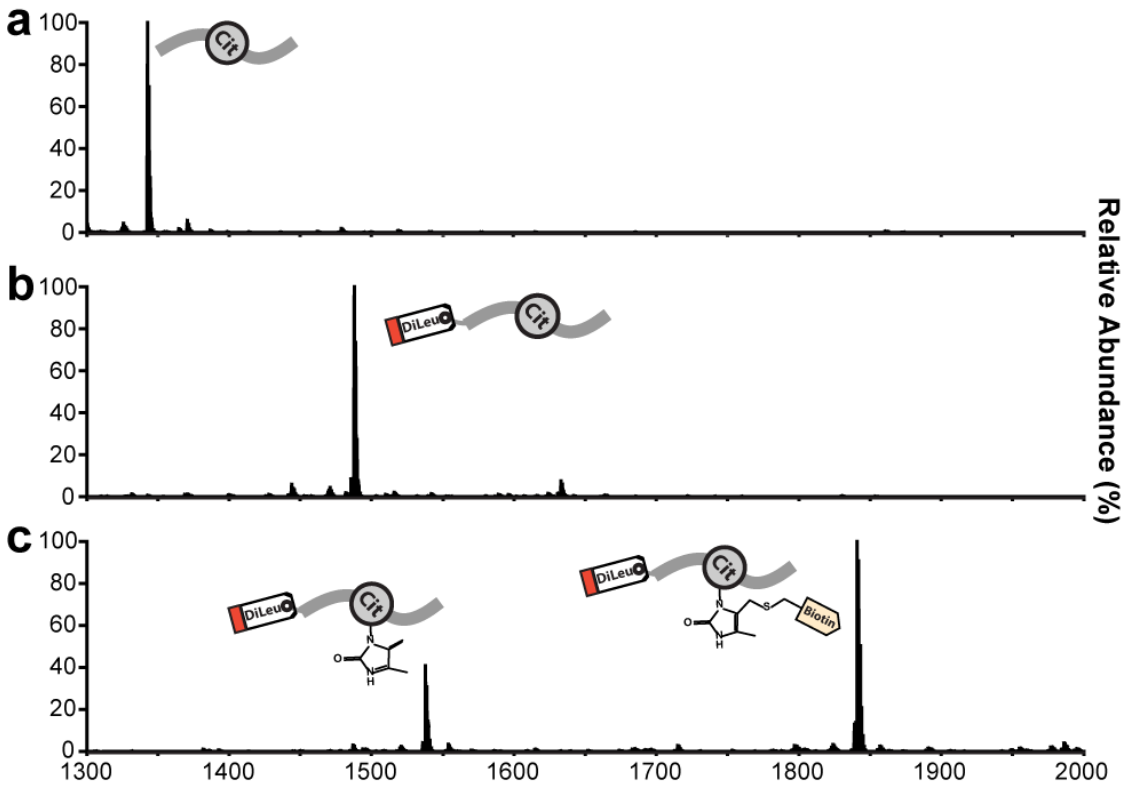

Figure 1. Test of reaction efficiency using citrullinated peptide standard. a, Spectrum of citrullinated peptide standard. **b,** Spectrum of DiLeu labeled peptide standard. **c,** Spectrum of DiLeu labeled and biotin thiol tag-derivatized peptide standard. The two-step reaction is complete without observable side products or interference. All spectra were acquired on a MALDI-LTQ Orbitrap XL mass spectrometer.

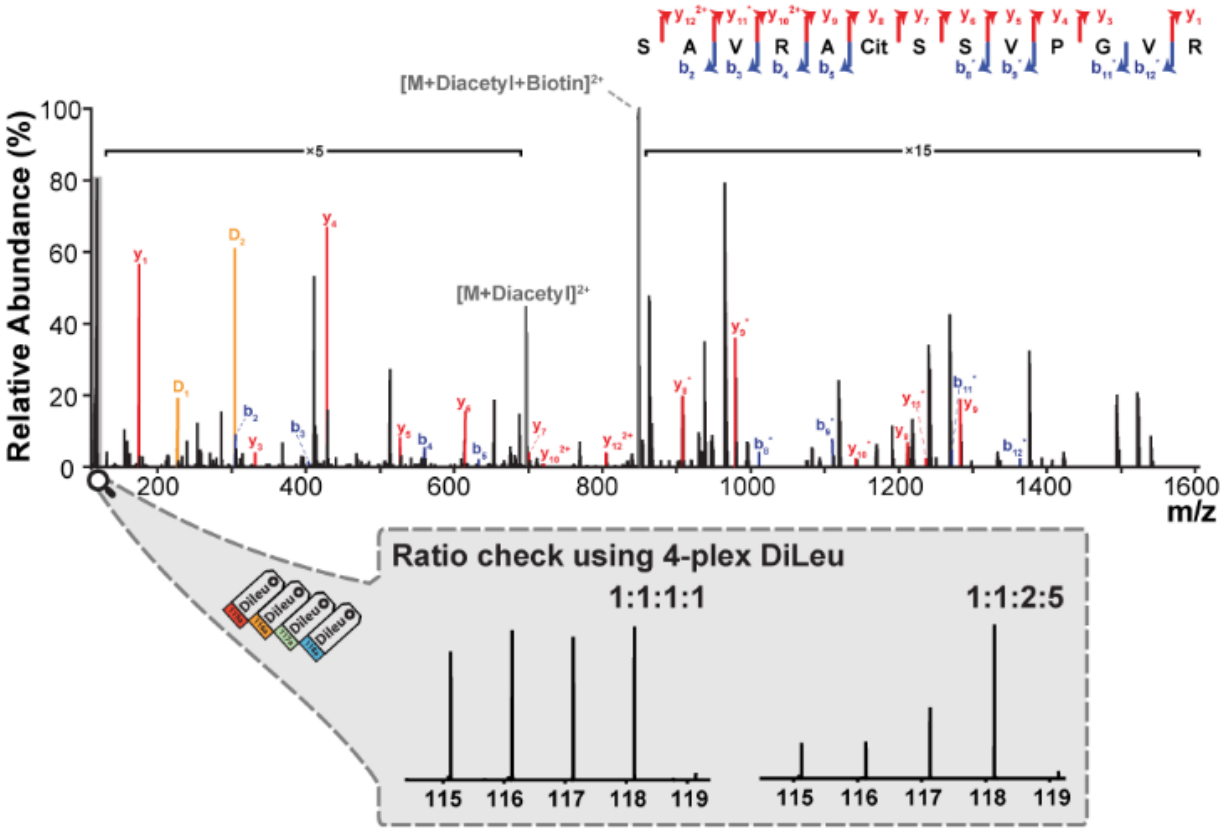

Figure 2. Tandem MS spectrum of the 4-plex DiLeu-labeled and biotin thiol tag-derivatized peptide standard. DiLeu reporter ion signals are fully resolved in the low-mass range, and a wealth of b- and y-ions are observed for confident peptide sequence identification and citrullination site localization. To test the quantification accuracy, the peptide standards were differentially labeled with 4-plex DiLeu in two known ratios 1:1:1:1 and 1:1:2:5, and subjected to citrullination analysis using biotin thiol tag. Example spectra of DiLeu reporter ions in the low-mass range are shown in the zoom-in figure, indicating great quantitative accuracy.

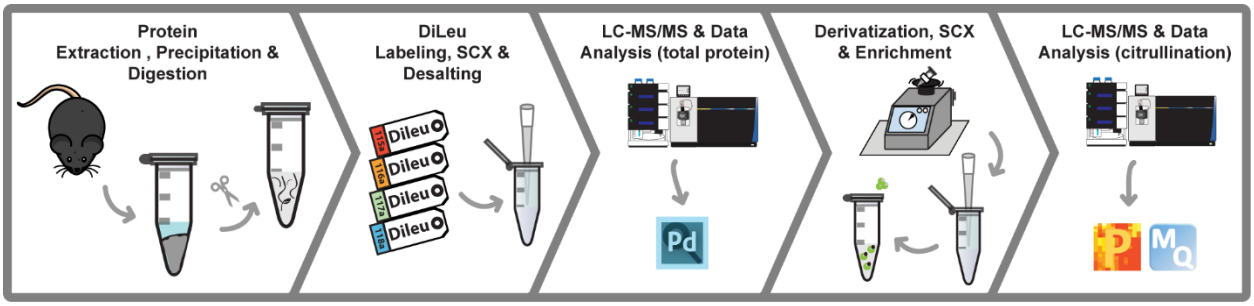

Figure 3. Experimental workflow of multiplexed quantitative citrullination analysis from complex biological samples using DiLeu isobaric tags. As an example, mouse brain samples were subjected to protein extraction, precipitation, and digestion according to a slightly modified SCAD method. DiLeu isobaric labeling was performed followed by a SCX clean-up step to remove excess DiLeu tags. Differentially labeled samples were then combined and citrullinated peptides were derivatized with biotin thiol tag. Target molecules were enriched using streptavidin agarose and released for LC-MS/MS analysis and data processing was done with commercially available software packages. An additional proteomics analysis was performed after DiLeu labeling for normalization purpose.

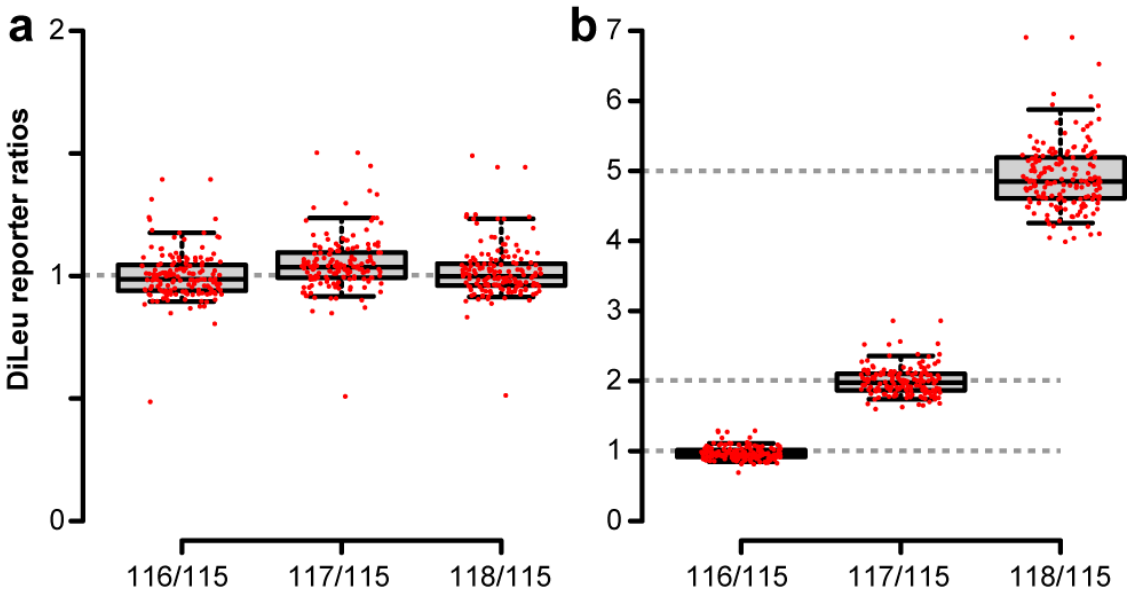

Figure 4. Test of quantitation accuracy and precision with 4-plex DiLeu labeling using mouse brain digest. Peptides were differentially labeled with 4-plex DiLeu in two known ratios 1:1:1:1 (**a**) and 1:1:2:5 (**b**), and combined for citrullination analysis. Red dots indicate the measured quantitative ratios for each detected citrullinated peptide. Top and bottom of boxes indicate 3rd and 1st quartile, respectively, and whiskers extend to 95th and 5th percentile. Horizontal lines within boxes denote median. Grey dash lines indicate theoretical ratios.

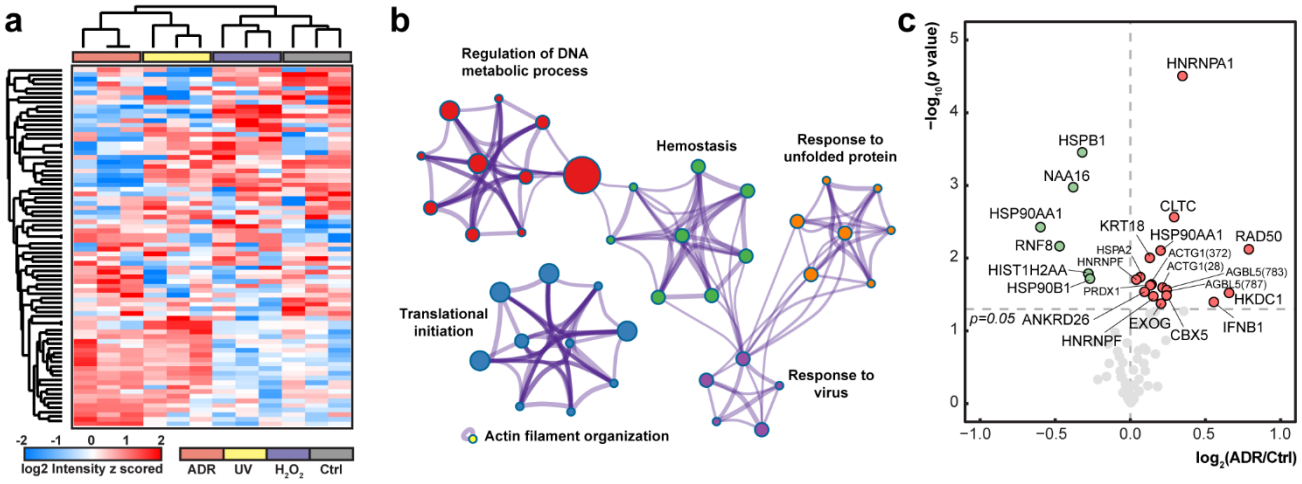

Figure 5. High-throughput investigation of citrullination changes in response to DNA damage stress. a, Hierarchical clustering of DiLeu reporter ion intensities of all quantified citrullinated peptides in control (Ctrl) and three treated groups. **b,** Network of biological processes enriched from significantly changed citrullinated peptides (one-way ANOVA, FDR 0.05) in different groups. Each node refers to an enriched term and different terms are grouped into clusters based on their similarities while the most statistically significant term represents the cluster name. The size of each node is proportional to the number of genes in that enriched term. **c,** Volcano plot showing pairwise comparison of citrullinated peptide expression levels between ADR treated and control groups. Points above horizontal dash line represent significantly altered peptides (two-sided *t* test, *p* value <0.05). Significantly down-regulated peptides are shown in green and up-regulated ones are shown in red. Gene names are annotated and citrullination sites are shown in the bracket if multiple sites are identified on the same protein. Detailed information can be found in **Table S2**.

Supplemental Information

Supplemental methods

Chemicals and materials. Dulbecco's modified Eagle's medium, fetal bovine serum, penicillin and streptomycin were purchased from Gibco (Gaithersburg, MD). MS grade Lys C/trypsin mixture and dithiothreitol (DTT) were purchased from Promega (Madison, WI). Urea, tris base, ACS grade acetone, ACS grade methanol (MeOH), ACS grade acetonitrile (ACN), ACS grade dichloromethane (DCM), Optima UPLC grade ACN, Optima UPLC grade water, and Optima LC/MS grade formic acid (FA) were purchased from Fisher Scientific (Pittsburgh, PA). Hydrogen peroxide (H_2O_2), doxorubicin hydrochloride (ADR), sodium dodecyl sulfate (SDS), guanidine hydrochloride (GuHCl), iodoacetamide (IAA), trifluoroacetic acid (TFA), triethylammonium bicarbonate (TEAB), *N*,*N*-dimethylformamide (DMF), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium tetrafluoroborate (DMTMM), *N*-methylmorpholine (NMM), 2,3-butanedione, streptavidin agarose, *N*,*N*-diisopropylethylamine, biotin-NHS ester, cysteamine, ammonium formate (NH_4HCO_2) and 2,5-dihydroxybenzoic acid were purchased from Sigma-Aldrich (St. Louis, MO). Hydroxylamine solution was purchased from Alfa Aesar (Ward Hill, MA). Sep-Pak C18 cartridges and Bridged Ethylene Hybrid C18 particles were purchased from Waters (Milford, MA). Quantitative colorimetric peptide assay was purchased from Thermo Scientific (San Jose, CA). Strong cation exchange (SCX) SpinTips containing PolySULFOETHYL A beads were purchased from PolyLC (Columbia, MD). Tris(2-carboxyethyl)phosphine (TCEP) were purchased from MilliporeSigma (Burlington, MA). 1× phosphate-buffered saline (PBS) buffer was purchased from Corning cellgro (Corning, NY). Citrullinated peptide standard SAVRACitSSVPGVR was purchased from GenScript (Piscataway, NJ).

Synthesis of biotin thiol tag. The synthesis of biotin thiol tag was performed accordingly to the previous publication. Briefly, *N*,*N*-diisopropylethylamine (0.88 mM) was added to a solution of biotin-NHS ester (0.29 mM) and cysteamine (0.44 mM) in DCM (5 mL) and stirred at 40 °C for 24 h. The crude product was purified using a CombiFlash system with a gradient of DCM from 0 to 20% in MeOH. Fractions containing pure product (as detected by UV) were collected (68% yield). TCEP 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.

DiLeu labeling of citrullinated peptide standard and derivatization using biotin thiol tag. Citrullinated peptide standard SAVRACitSSVPGVR was dissolved in water to a concentration of 1 mg/mL. Various amount of peptide standard was aliquoted, dried in vacuo and reconstituted in 0.5 M TEAB prior to DiLeu labeling which was the same as complex samples. Three hundred microgram of biotin thiol tag was dissolved with 40 µL 12.5% TFA. DiLeu labeled 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 following the manufacturer's protocols. Briefly, SCX tips were equilibrated with 100 µL loading buffer containing 50% ACN, 0.2% FA and 10 mM NH_4HCO_2 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 NH_4HCO_2 for 3 times. Flowthrough was collected and dried in vacuo. All centrifugation steps were performed at 400 ×g for 2 min.

MALDI-MS analysis of citrullinated peptide standard. 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% MeOH, 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 Scientific, San Jose, CA). Ionization was performed using a laser energy of 15 μJ. Spectra were acquired with a mass range of *m/z* 1300−2000 at a resolution of 30k.

LC-MS/MS analysis of citrullinated peptide standard. Samples were analyzed on a Q-Exactive quadrupole Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) coupled to a Waters nanoAcquity Ultra Performance LC. Each sample was dissolved in 10% ACN, 0.1% FA in water before loaded onto a 75 μm inner diameter homemade microcapillary column which is packed with 15 cm of Bridged Ethylene Hybrid C18 particles (1.7 μm, 130 Å) 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 90-min gradient elution of 10% to 40% mobile phase B at a flow rate of 300 nL/min. Survey scans of peptide precursors from 300 to 1500 *m/z* were performed at a resolving power of 70k with an AGC target of 1×10^6 and maximum injection time of 250 ms. The top 5 precursors were then selected for HCD fragmentation with a normalized collision energy of 30, an isolation width of 2.0 Da, a resolving power of 35k, an AGC target of 2×10^5, a maximum injection time of 120 ms, and lower mass limit of 110 *m/z*. Precursors were subject to dynamic exclusion for 10 s with a 10-p.p.m. tolerance. Fragment ions were manually annotated based on their accurate mass.

Supplemental figures and tables

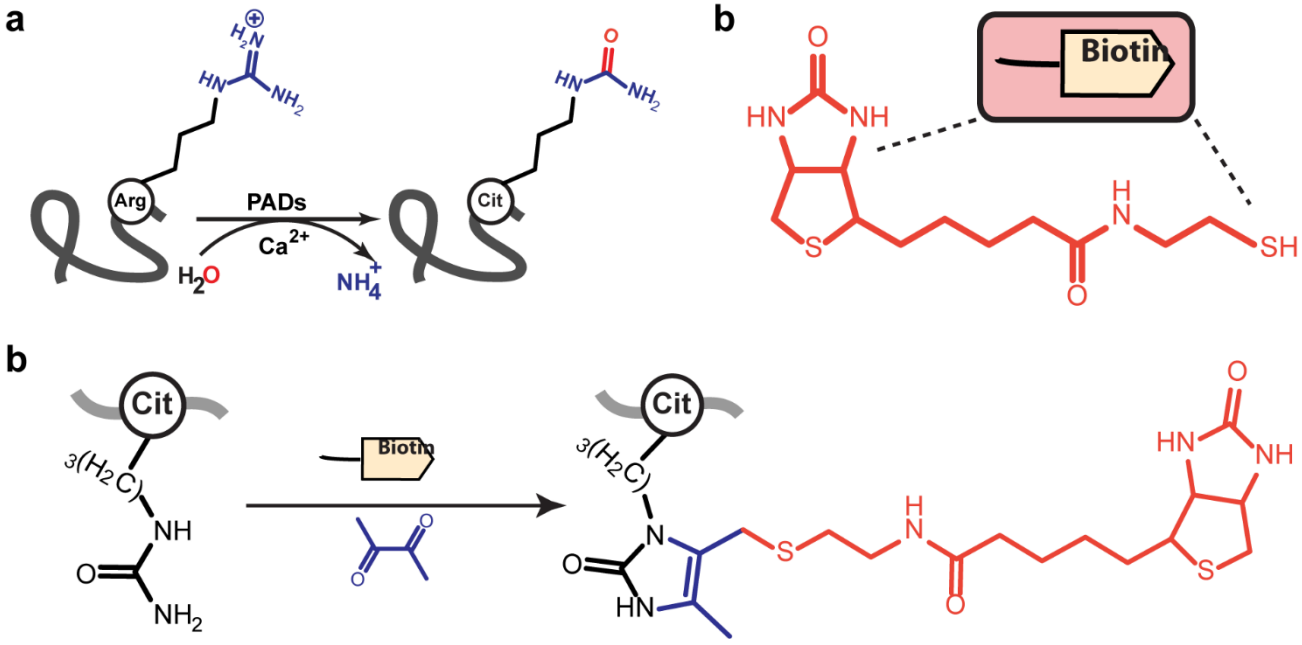

Figure S1. Protein citrullination and its analysis using biotin thiol tag derivatization. **a,** PAD enzymes catalyze the hydrolytic conversion of peptidyl arginine residues into citrulline. **b,** Structure of custom-developed biotin thiol tag. **c,** Derivatization of citrullinated peptides using biotin thiol tag and 2,3-butanadione under acidic conditions.

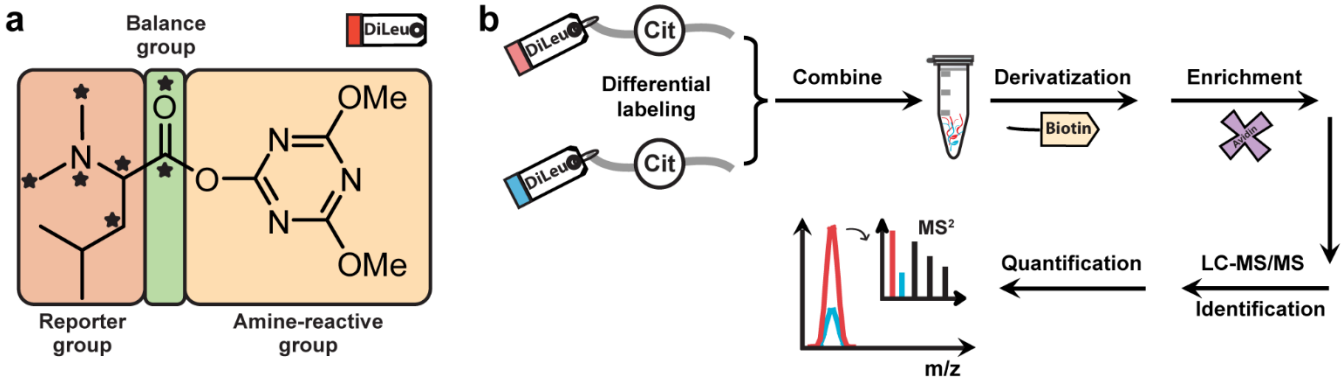

Figure S2. Quantitative citrullination analysis using DiLeu isobaric labeling strategy. **a,** Structure of DiLeu isobaric tags. Stars indicate the positions of isotopic substitution to achieve multiplexed analysis. **b,** Rationale of quantitative citrullination analysis. Samples will be differentially labeled using DiLeu isobaric tags and combined for biotin tag derivatization and enrichment of citrullinated peptides. Relative quantitation can be achieved using DiLeu reporter ions upon fragmentation.

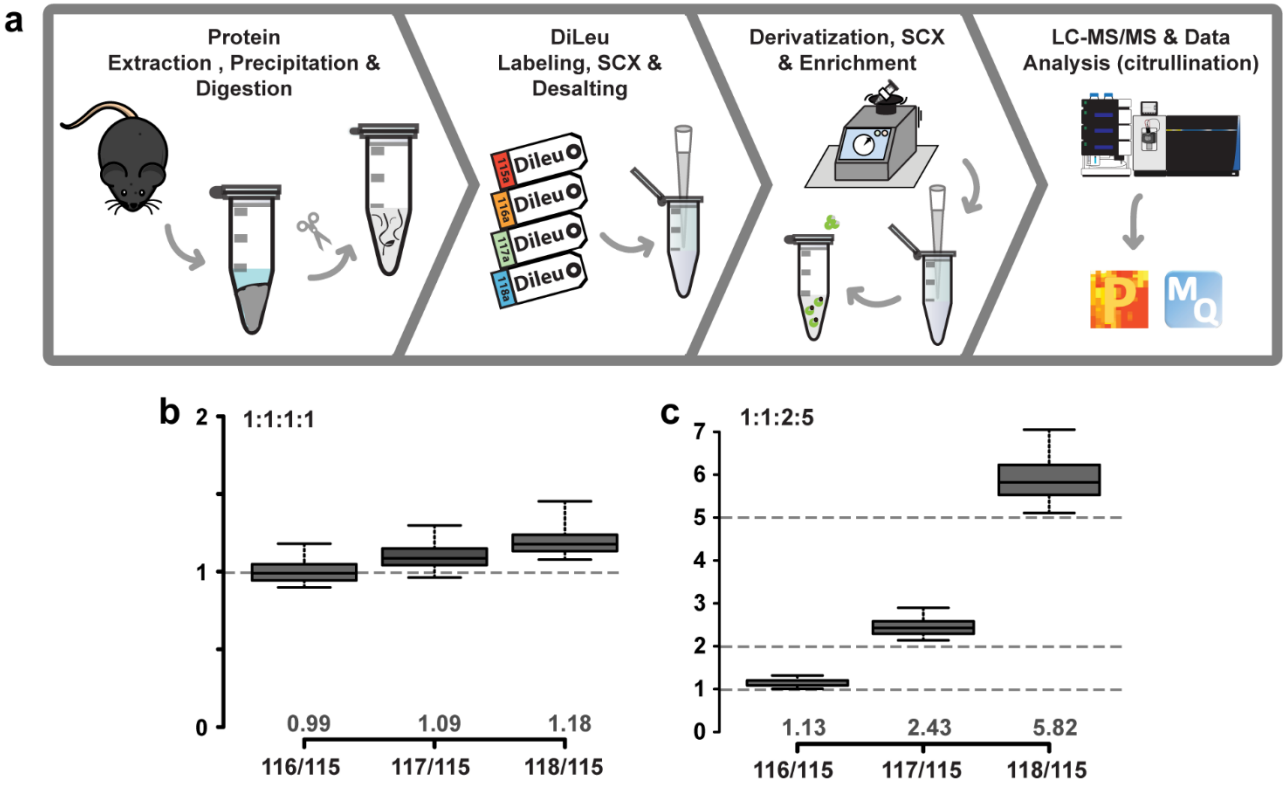

Figure S3. Test of quantitation accuracy and precision with 4-plex DiLeu labeling using mouse brain digest without normalization. **a,** Experimental workflow of quantitative citrullination analysis using mouse brain digest. Boxplots showing the detected ratios of all quantified citrullinated peptides when samples were differentially labeled with 4-plex DiLeu in two known ratios 1:1:1:1 (**b**) and 1:1:2:5 (**c**). Top and bottom of boxes indicate 3rd and 1st quartile, respectively, and whiskers extend to 95th and 5th percentile. Medians are shown as horizontal lines within boxes and the numbers below. Grey dash lines indicate theoretical ratios.

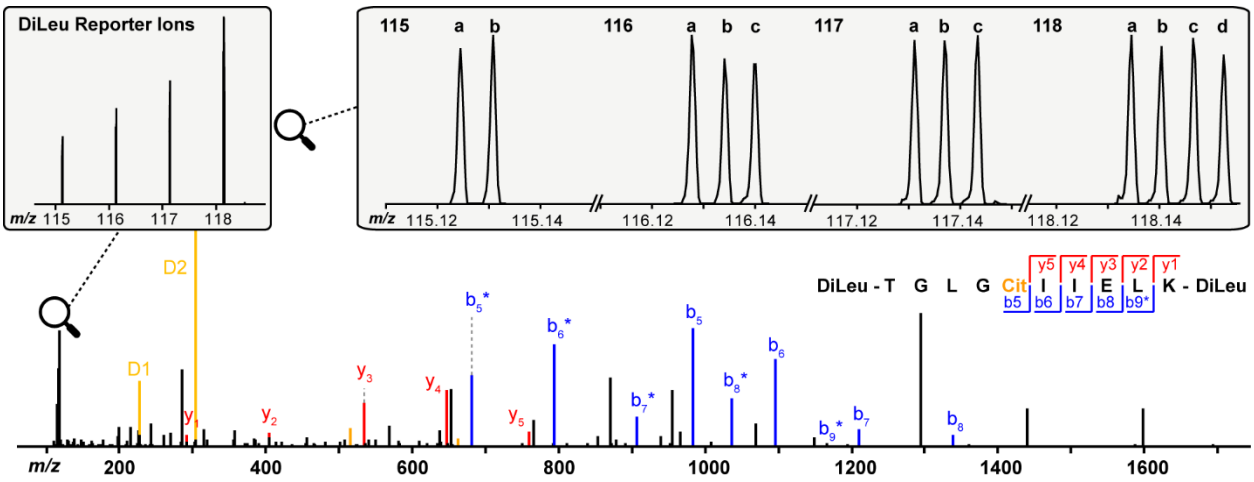

Figure S4. A tandem MS spectrum of a 12-plex DiLeu-labeled citrullinated peptide (R441Cit) on DNA repair protein RAD50 upon HCD fragmentation. A wealth of b- and y-ions are observed for confident peptide sequence identification and citrullination site localization. DiLeu reporter ion signals are fully resolved in the low-mass range which allows for accurate quantification.

Table S1. List of all identified citrullinated proteins and citrullination sites

UniProt	Gene	Cit site	*p* value[a]	UniProt	Gene	Cit site	*p* value[a]
P09651	HNRNPA1	178	0.00001	O95153	BZRAP1	1419	0.07781
O76064	RNF8	441	0.00001	P52597	HNRNPF	206	0.07790
Q92878	RAD50	441	0.00001	O95153	BZRAP1	1420	0.07971
Q9Y2C4	EXOG	109	0.00080	Q9Y3U8	RPL36	87	0.08084
P46782	RPS5	136	0.00100	P54652	HSPA2	419	0.10518
P46776	RPL27A	65	0.00133	P63261	ACTG1	28	0.12542
Q9BXB4	OSBPL11	709	0.00150	P38159	RBMX	172	0.13400
P01574	IFNB1	186	0.00160	P16949	STMN1	14	0.13430
Q9BXB4	OSBPL11	711	0.00171	P08865	RPSA	117	0.13433
P07900	HSP90AA1	620	0.00200	P67809	YBX1	231	0.14049
P07900	HSP90AA1	456	0.00233	P05783	KRT18	301	0.16720
P04075	ALDOA	201	0.00255	P68363	TUBA1B	64	0.16910
P04792	HSPB1	127	0.00462	Q5VTE0	EEF1A1P5	322	0.18054
Q6N069	NAA16	549	0.00486	P49327	FASN	1105	0.19713
P29692	EEF1D	136	0.00533	P32969	RPL9	129	0.20437
Q2TB90	HKDC1	42	0.00750	Q14194	CRMP1	467	0.23375
Q06830	PRDX1	110	0.01247	Q9HAU6	TPT1P8	107	0.23679
P14625	HSP90B1	587	0.01867	P68363	TUBA1B	339	0.29067
P09651	HNRNPA1	97	0.02084	P40429	RPL13A	94	0.31179
Q00610	CLTC	320	0.02260	P30084	ECHS1	54	0.34529
P52597	HNRNPF	212	0.02313	P78371	CCT2	156	0.36120
Q8NDL9	AGBL5	783	0.02345	P38159	RBMX	202	0.38275
P45973	CBX5	107	0.02352	P62241	RPS8	178	0.39116
Q96QV6	HIST1H2AA	89	0.02400	Q9HDC9	APMAP	151	0.40731
Q8NDL9	AGBL5	787	0.02457	P18124	RPL7	236	0.41371
P04075	ALDOA	331	0.02492	Q5VTE0	EEF1A1P5	240	0.43243
P62249	RPS16	27	0.02857	P11021	HSPA5	49	0.45661
Q68DQ2	CRYBG3	573	0.02889	P13010	XRCC5	400	0.51409
Q9H0U4	RAB1B	183	0.03779	Q9UPZ6	THSD7A	757	0.52312
P22626	HNRNPA2B1	185	0.04547	O95613	PCNT	2212	0.55588
P63261	ACTG1	372	0.04735	P04075	ALDOA	56	0.56583
P68363	TUBA1B	221	0.05075	Q9UII2	ATPIF1	50	0.60642
Q9H9D4	ZNF408	595	0.05847	Q9Y2K6	USP20	693	0.68194
P06744	GPI	417	0.05952	Q9Y365	STARD10	14	0.90665
Q9UPS8	ANKRD26	478	0.06389	Q07020	RPL18	38	0.90745
P38646	HSPA9	107	0.06389	P36578	RPL4	205	0.92487
Q9HC10	OTOF	787	0.06821	P04075	ALDOA	43	0.93668
P52597	HNRNPF	349	0.06865	P02545	LMNA	471	0.94679
Q15056	EIF4H	237	0.07077	P22314	UBA1	69	0.95579

[a] One-way ANOVA *p* values across four different groups

Table S2. Significantly changed citrullination sites between ADR-treated and control groups

UniProt	Gene	Cit site	p value[a]	log_2(ADR/Ctrl)
P09651	HNRNPA1	178	0.00003	0.347
P04792	HSPB1	127	0.00035	-0.320
Q6N069	NAA16	549	0.00106	-0.380
Q00610	CLTC	320	0.00272	0.292
P07900	HSP90AA1	620	0.00374	-0.598
O76064	RNF8	441	0.00685	-0.471
Q92878	RAD50	441	0.00755	0.789
P07900	HSP90AA1	456	0.00786	0.202
P05783	KRT18	301	0.00990	0.129
Q96QV6	HIST1H2AA	89	0.01632	-0.280
P54652	HSPA2	419	0.01823	0.066
P14625	HSP90B1	587	0.01912	-0.269
P52597	HNRNPF	206	0.01971	0.040
P63261	ACTG1	372	0.02294	0.141
Q06830	PRDX1	110	0.02369	0.132
P63261	ACTG1	28	0.02530	0.215
Q8NDL9	AGBL5	787	0.02742	0.243
Q8NDL9	AGBL5	783	0.02742	0.243
Q9UPS8	ANKRD26	478	0.02900	0.094
Q2TB90	HKDC1	42	0.02998	0.657
P45973	CBX5	107	0.03254	0.241
P52597	HNRNPF	212	0.03342	0.153
P01574	IFNB1	186	0.04006	0.556
Q9Y2C4	EXOG	109	0.04213	0.204

[a] Two-sided Student's *t* test *p* values between ADR treated and control groups

Chapter 5

LysargiNase complements trypsin for enhanced profiling of protein citrullination

Adapted from: **Li, Z.**; Ma, M.; Wang, B.; Yu, Q.; Shi, Y.; Li, L., LysargiNase complements trypsin for enhanced profiling of protein citrullination. **2021.** To be submitted.

Abstract

Citrullination is an important protein post-translational modification (PTM) occurred on arginine residues where the positively charged guanidium group is catalyzed to a neutral ureido group. It has profound influence on protein conformations and protein-protein interactions, and consequently, dysregulation of citrullination has been associated with various disease development and progression. However, detection of citrullination using mass spectrometry has long been challenging. Recently, we developed a novel biotin thiol tag which enabled large-scale and confident identification of this PTM. In this study, we further extend the utility of this pipeline by integrating LysargiNase-based digestion methods. We first explore different digestion protocols and observe the optimal results when using a combination of LysN and LysargiNase enzyme mixture. We then demonstrate the distinct fragmentation behavior of LysargiNase digested citrullinated peptides and its orthogonality in detecting protein citrullination compared with trypsin digestion. Benefiting from various digestion and fragmentation techniques, we identify 305 citrullination sites on 228 proteins from mouse brain, which represents the most in-depth analysis of site-specific protein citrullination analysis from a single tissue type. Our results indicate intimate involvement of this PTM in many important functions such as regulation of structural integrity, neuronal signal transmission and metabolic processes including respiration. More importantly, our novel method provides an attractive alternative strategy to enhance the profiling of protein citrullination and will greatly facilitate future investigations to understand the critical roles of this PTM in various biological systems and processes.

Introduction

Protein citrullination is a post-translational modification (PTM) resulting from the conversion of peptidyl arginine to citrulline, which is catalyzed by calcium-regulated protein arginine deiminases (PADs) (**Figure S1a**).[1-6] Under physiological pH, this conversion of positively charged guanidinium group to neutral ureido group is likely to exert profound effects on protein conformations, protein-protein interactions and protein functions.[3, 5, 7] Dysregulation of protein citrullination was therefore reported to be involved in many disease onset and progression including rheumatoid arthritis[8-13], multiple sclerosis[14-17], Alzheimer's disease[18] and cancers[19-21].

Currently, antibody-based methods are the most prevalent in detecting citrullination.[22-24] However, these methods are intrinsically incompatible with large-scale analysis, and it is difficult to pinpoint specific modification sites. On the other hand, mass spectrometry (MS)-based strategies, which have become powerful tools in many PTM-related studies, still face great challenges in citrullination profiling because of the nature of this PTM.[23, 25] First, citrullination only leads to a mass increment of 0.984 Da which is easily confused with deamidation or ^{13}C isotopic peaks in a mass spectrum and thus precludes confident identification and localization. Second, the signal of low-abundance citrullinated peptides is largely suppressed by high-abundance molecules without effective enrichment methods. To deal with these difficulties, tremendous effort has been made such as careful examination of the spectra[26-27], applying more restrict criteria or novel algorithms during data searching[28-33], and developing chemical derivatization[34-35] and enrichment methods.[8, 36-37] Unfortunately, few of these previous methods were able to provide satisfying results with reasonable time and resource investment. Recently we have developed a novel strategy for improved citrullination analysis by designing a biotin thiol tag which can specifically react with citrulline residues.[38] This method not only induces a larger mass

increment of citrullination compared to the non-modified counterparts for more confident characterizations, but also enables effective enrichment using biotin-avidin interactions (**Figure S1b**).

In our study as well as the majority of shotgun proteomics and PTM research, trypsin is commonly used to digest large proteins into smaller peptides. Trypsin cleaves after basic residues lysine and arginine (**Figure S1c**), generating relatively short peptides with a C-terminal positively charged residue and a N-terminal positively charged primary amine.[39] Consequently, trypsin digestion results in sequence-informative ions upon fragmentation and is considered as the gold standard in enzymatic digestion protocols. However, in recent years, novel proteases are being explored to facilitate more complete protein sequencing and characterization of its PTMs. Among them, a metalloproteinase called LysargiNase was developed for applications in proteomics[40-44] and a few other PTM studies such as phosphorylation, methylation and ubiquitination.[45-48] LysargiNase cleaves before lysine and arginine (**Figure S1c**) and produces peptides of same length and highly similar sequence as tryptic peptides, though with both positive charges swapped to peptide N-termini.[45, 49] LysargiNase, also known as mirror trypsin, can efficiently cleave modified residues in contrast to trypsin[45], which endows this protease with extra benefits especially in PTM investigations.

Herein, we explore the capability of LysargiNase in citrullination analysis from complex biological samples with the integration of our biotin thiol tag-based approach. We also demonstrate the complementarity of LysargiNase to trypsin for improved citrullination profiling, which provides new insights of this PTM in various biological processes and may benefit future effort towards a comprehensive understanding of its physiological and pathological involvement.

Methods

Chemicals and materials. LysargiNase and LysN enzymes are kind gifts from Dr. Yu Feng from Beijing Shengxia Proteins Scientific Ltd. MS grade Lys C/trypsin mixture and dithiothreitol (DTT) were purchased from Promega (Madison, WI). Urea, tris base, ACS grade acetone, ACS grade methanol (MeOH), ACS grade acetonitrile (ACN), ACS grade dichloromethane (DCM), Optima UPLC grade ACN, Optima UPLC grade water, and Optima LC/MS grade formic acid (FA) were purchased from Fisher Scientific (Pittsburgh, PA). Sodium dodecyl sulfate (SDS), iodoacetamide (IAA), trifluoroacetic acid (TFA), 2,3-butanedione, streptavidin agarose, *N*,*N*-diisopropylethylamine, biotin-NHS ester, cysteamine, ammonium formate (NH_4HCO_2) were purchased from Sigma-Aldrich (St. Louis, MO). Sep-Pak C18 cartridges and Bridged Ethylene Hybrid C18 particles were purchased from Waters (Milford, MA). Quantitative colorimetric peptide assay was purchased from Thermo Scientific (San Jose, CA). Strong cation exchange (SCX) SpinTips containing PolySULFOETHYL A beads were purchased from PolyLC (Columbia, MD). Tris(2-carboxyethyl)phosphine (TCEP) were purchased from MilliporeSigma (Burlington, MA). 1× phosphate-buffered saline (PBS) buffer was purchased from Corning cellgro (Corning, NY).

Synthesis of biotin thiol tag. The synthesis of biotin thiol tag was performed accordingly to the previous publication.[38] Briefly, *N*,*N*-diisopropylethylamine (0.88 mM) was added to a solution of biotin-NHS ester (0.29 mM) and cysteamine (0.44 mM) in DCM (5 mL) and stirred at 40 °C for 24 h. The crude product was purified using a CombiFlash system with a gradient of DCM from 0 to 20% in MeOH. Fractions containing pure product (as detected by UV) were collected (68% yield). TCEP 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.

Protein extraction and digestion. Mouse tissues were dissolved in extraction buffer solution (4 % SDS, 50 mM Tris buffer, pH adjusted to 8 with hydrochloric acid) and sonicated using a probe sonicator (Thermo Scientific, San Jose, CA). Protein extracts were reduced with 10 mM DTT for 30 min at room temperature and alkylated with 50 mM IAA for another 30 min in dark before quenched with DTT. Proteins were then precipitated with 80% (v/v) cold acetone (-20 ℃) overnight. Samples were centrifuged at 14,000 ×g for 15 min after which supernatant was discarded. Pellets were rinsed with cold acetone again and air-dried at room temperature.[50] Eight moles of urea were added to dissolve the pellets and 50 mM tris buffer was used to dilute the samples to a urea concentration <1 M. On-pellet digestion was performed with LysC/trypsin or LysN or LysargiNase mixture in a 50:1 ratio (protein:enzyme, w/w) at 37 ℃ overnight. The digestion was quenched with 1% TFA and samples were desalted with Sep-Pak C18 cartridges. Concentrations of peptide mixture were measured by peptide assay following the manufacturers' protocols. Four hundred microgram of each sample was dried in vacuo.

Derivatization of citrullinated peptides using biotin thiol tag. Each sample and 300 µg of biotin thiol tag were dissolved with 40 µL 12.5% TFA. A solution of 2,3-butanedione was prepared by mixing 1 µL of 2,3-butanedione with 114 µL 12.5% TFA. Ten microliters of 2,3-butanedione solution was subsequently added to initiate the derivatization reaction and the mixture was vortexed in dark at 37 °C for 6 h. SCX clean-up was then performed to remove the excess tag and samples were dried in vacuo.

Enrichment of derivatized citrullinated peptides. The enrichment process was performed as previously described with slight modifications.[51] Briefly, 75 µL streptavidin agarose was washed with 1 mL 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 while the other three release processes were conducted at 95 °C for 5 min with shaking. The eluents were combined and dried in vacuo.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Samples were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer coupled to a Dionex UltiMate 3000 UPLC system (Thermo Scientific, San Jose, CA). Each sample was dissolved in 8% ACN, 0.1% FA in water before loaded onto a 75 μm inner diameter homemade microcapillary column which was packed with 15 cm of Bridged Ethylene Hybrid C18 particles (1.7 μm, 130 Å) 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 8% to 37% mobile phase B at a flow rate of 300 nL/min. For HCD fragmentation method, survey scans of peptide precursors from 350 to 1500 *m/z* were performed at a resolving power of 60k with an AGC target of 2×10^5 and maximum injection time of 100 ms. 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-p.p.m. tolerance. For ETD fragmentation method, reaction time was set as 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. All other parameters were set the same as HCD method. Each sample was acquired in technical triplicates.

Data analysis. Raw files were searched against the UniProt *Mus musculus* (December 2018) reviewed database using MaxQuant (version 1.5.2.8) with trypsin/P or LysN or LysargiNase selected as the enzyme and three missed cleavages allowed. LysN was self-defined as protease that cleaves before lysine and LysargiNase cleaves before lysine and arginine. Carbamidomethylation of cysteine residues (+57.02146 Da) was chosen as the fixed modification and variable modifications included oxidation of methionine residues (+15.99492 Da), acetylation at protein N-termini (+42.01056 Da) and biotin tag-labeled citrullination of arginine (+354.10718 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. 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 site localization probability threshold was set to 0.75. All other parameters were set as default. Peptide C-terminal citrullination identifications in trypsin digestion or N-terminal identifications in LysargiNase digestion were filtered out as well. Box plots were generated using R packages. Gene ontology (GO) analysis was accomplished using DAVID bioinformatics resources[52] with a FDR cutoff of 0.05. Sequence motif analyses were done using WebLogo.[53]

Results and Discussion

Optimization of LysargiNase digestion protocols for citrullination analysis. A typical workflow we developed for large-scale citrullination profiling is shown in **Figure S2**. Proteins were extracted from complex biological samples and enzymatically digested into peptides. After sample clean-up and peptide assay which determines peptide amount, each sample was incubated

with our recently developed biotin thiol tag to react with citrulline residues followed by SCX to remove the excess tags. Derivatized citrullinated peptides were then enriched using streptavidin beads and released for LC-MS/MS analysis. We first compared LysargiNase digestion as opposed to the commonly used LysC/trypsin digestion in such analysis using mouse brain. We found LysargiNase has slightly better specificity of arginine than trypsin but is less specific towards lysine residues with more than 20% of the identified peptides containing missed cleavage of lysine (**Figure 1a**). Although this is sort of expected probably because arginine is more basic than lysine and therefore is better stabilized by the enzyme through ionic interactions during substrate recognition and binding processes. Indeed, trypsin itself is less specific towards lysine and many previous studies including ours utilized an enzyme mixture of trypsin and LysC to improve the digestion efficiency.[54] We reason that a combined use of LysN, which cleaves before lysine, will be similarly beneficial to LysargiNase-based digestion applications. We tested different orders of adding LysN and LysargiNase in a serial digestion and observed the lowest missed cleavage rate when using LysN to pre-digest the samples for four hours followed by overnight digestion with LysargiNase (**Figure 1a**). Lower missed cleavage rate may produce more shorter-length peptides and contribute to higher ionization and fragmentation efficiency during MS acquisition. Consequently, LysN/LysargiNase digestion protocol results in the highest number of identified citrullination sites and citrullinated proteins compared to the others using HCD fragmentation technique (**Figure 1b** and **Figure S3**), and accordant trend is observed using ETD fragmentation as well (**Figure 1c**). In general, HCD fragmentation outperforms ETD in terms of identification numbers likely due to the longer duty cycle in ETD. We also found LysargiNase digestion leads to overall lower identification numbers compared to LysC/trypsin digestion regardless of the fragmentation type. This is in part caused by two positive charges predominantly locating at

peptide N-termini which hampers efficient fragmentation.[49] Despite this, LysargiNase digestion method provides decent number of identifications since it still produces peptides of desired length, which makes it a potentially good complement to trypsin in citrullination analysis.

LysargiNase was shown to also cleave before modified lysine and arginine in contrast to trypsin, which renders its superiority in PTM studies.[45] However, similar to trypsin, it also relies on a negatively charged aspartate residue in the binding pocket to capture and stabilize positively charged amino acids during the hydrolysis process.[41] This property of the enzyme raises the questions whether the enzyme is capable to cleave neutral citrullinated arginine and it is indeed still controversial on whether trypsin can cleave citrulline residues.[5, 26, 29, 55] In this study, we treated a few citrullinated peptides of varying length and sequences with LysargiNase and found none of them could be cleaved *in vitro* (data now shown). On the contrary, N-terminal citrullination sites were identified with high confidence in complex biological samples, similar to our previous observations in trypsin digestion.[38] These results may indicate the possibility of LysargiNase to cleave before citrulline *in vivo* but in a protein-specific manner and slower rate. To ensure the reliability of our citrullination identifications, we manually removed those N-terminal hits in the results, but we believe that a systematic and careful study into the mechanisms is required to fully understand the enzyme specificity and cleavage product.

Complementarity of citrullination identifications using different digestion methods and fragmentation techniques. We then sought to see if LysargiNase could provide complementary results to trypsin digestion. **Figure 2a** shows example spectra of a same citrullination site (R228) on myelin basic protein identified by either trypsin or LysargiNase digestion using HCD fragmentation. We found more y ions were generated in tryptic peptides while b ions were more frequently present in the spectra of LysargiNase digested samples, which is consistent with the

fact that both protons are located at the N-terminal in the latter.[49] It is an obvious unique characteristic of LysargiNase peptides that b ions are predominantly formed upon HCD fragmentation. **Figure 2b** shows the number of assigned b ions/y ions of each identified citrullinated peptide with two different digestion methods and we observe significantly higher ratios in LysargiNase peptides, indicating again its preference in b ion generation. In accordance with this, sequence coverage of b ions is more than 20% higher in LysargiNase peptides and vice versa in tryptic peptides (**Figure S4a**). In ETD fragmentation, c or z ions are more evenly distributed in tryptic peptides while c ions consist of a slightly higher proportion in LysargiNase peptides (**Figure 2c** and **Figure S4b**). The less significant difference between the two enzymes upon ETD fragmentation may explain that identification numbers in LysargiNase digestion are 70% of those in trypsin digestion (**Figure 1c**) while the percentage is only 60% in HCD fragmentation (**Figure 1b**). Since the two enzymes exhibit distinct fragmentation features, there is a large degree of orthogonality in citrullination coverage using LysargiNase and trypsin. With a combined use of both enzymes, we almost doubled the identification rate compared with using only one of them (**Figure 3a** and **Figure S5a**). In addition, nearly 30% more citrullination sites and proteins could be identified with ETD fragmentation technique (**Figure 3b** and **Figure S5b**), which further improves the depth of coverage. Collectively, our results reveal clear complementarity of LysargiNase versus trypsin, which makes it an attractive supplement in citrullination analysis, and ETD fragmentation method may provide another dimension of benefits to push the boundary in related studies.

Enhanced profiling of protein citrullination in mouse brain. With a combination of various digestion and fragmentation methods, we were able to achieve an in-depth profiling of protein citrullination in mouse brain. In total, we identified 305 citrullination sites from 228 citrullinated

proteins, which represents the most comprehensive detection of this PTM from a single tissue type (**Table S1**). To better discern the localization and functions of these citrullinated proteins, we performed GO analysis and the significantly enriched terms are shown in **Table S2**. We found that these proteins are mainly involved in regulation of protein structure and binding, neuronal signal transduction and metabolic processes. Concordantly, citrullinated proteins are observed to concentrate in cytoskeleton and microtubule, myelin sheath, axon and synapse, and cytoplasm and mitochondria (**Figure 4a, Table S2**). Our results indicate widespread distribution of protein citrullination and its active participation in various critical biological processes *in vivo*, which greatly improves the current understanding of this PTM in epigenetic regulation. We then performed a motif analysis of all identified citrullinated peptides and found glutamic acid residues are present slightly more frequently in the proximity of citrullination positions especially on the N-terminal side, which is consistent with a few previous studies.[28] And such motif preserves regardless of what enzymes (**Figure S6**) or fragmentation methods (**Figure S7**) we used. Interestingly, we also noticed that basic (lysine, arginine) or acidic amino acids (aspartate, glutamate) and proline also surround citrullination sites more often, which coincide with the sequence patterns that promote missed cleavage in either trypsin or LysargiNase digestion.[45, 56-57] Since we manually removed enzyme-cleaved citrullination site identifications (peptide terminal sites), this coincidence might indicate that citrulline residues are enzymatically cleavable. Again, a meticulous examination and verification on this perspective will be highly merited. **Figure 4c** captures the prevalence of singly- and multiply-citrullinated proteins where nearly 80% of the identified proteins were observed with only one citrullination site. Some of the proteins that carry multiple modification sites are shown in the table and they are all associated with the important biological functions mentioned before. For instance, citrullination on spectrin and microtubule-

associated proteins may indicate previously unknown relationship of this PTM with structural integrity of membranes and cells[58-59], which is also consistent with previous findings that dysregulation of citrullination led to structural collapse and facilitated cell death.[7] In another example, synapsin functions importantly in the regulation of neurotransmitter release in synaptic vesicles[60-61] and our study indicates the possible involvement of its citrullinated form. In our previous work, we identified nine citrullination sites on myelin basic protein[38], which is the major component of myelin sheath that secures neuronal signal transduction as an insulator.[62] Here, we identify two additional sites on this important protein, which provides a richer reservoir and can help further conjecture the regulatory mechanisms of this PTM.

Conclusions

Herein we demonstrate the compatibility of LysargiNase with our citrullination analysis pipeline and further optimize the method for improved detection. We then reveal the unique fragmentation characteristics of different digestion methods and demonstrate the complementarity of LysargiNase to trypsin in citrullination analysis. We also perform an in-depth profiling of citrullination in mouse brain using a combination of various digestion and fragmentation approaches. Our results unveil the previously unknown distribution and functions of citrullinated proteins *in vivo*, which greatly improves the understanding of this PTM under physiological conditions and will also benefit future explorations of its involvement in different disease states.

Acknowledgements

This study was supported by grant funding from the NIH (R21 AG060242). 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. The authors would like to acknowledge Dr. Yu Feng from Beijing Shengxia Proteins Scientific Ltd. for kindly providing the LysargiNase and LysN enzymes. 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.

References

(1) Fuhrmann, J.; Clancy, K. W.; Thompson, P. R. *Chem. Rev.* **2015,** *115*, 5413-61.

(2) Tilvawala, R.; Thompson, P. R. *Curr Opin Struct Biol* **2019,** *59*, 205-215.

(3) Witalison, E.; Thompson, P.; Hofseth, L. *Current Drug Targets* **2015,** *16*, 700-710.

(4) Fuhrmann, J.; Thompson, P. R. *ACS chemical biology* **2016,** *11*, 654-68.

(5) Vitorino, R.; Guedes, S.; Vitorino, C.; Ferreira, R.; Amado, F.; Van Eyk, J. E. *Journal of proteome research* **2020**.

(6) Mondal, S.; Thompson, P. R. *Acc Chem Res* **2019,** *52*, 818-832.

(7) Gyorgy, B.; Toth, E.; Tarcsa, E.; Falus, A.; Buzas, E. I. *The international journal of biochemistry & cell biology* **2006,** *38*, 1662-77.

(8) Tutturen, A. E.; Fleckenstein, B.; de Souza, G. A. *Journal of proteome research* **2014,** *13*, 2867-73.

(9) Schellekens, G. A.; de Jong, B. A.; van den Hoogen, F. H.; van de Putte, L. B.; van Venrooij, W. J. *The Journal of clinical investigation* **1998,** *101*, 273-81.

(10) Raijmakers, R.; van Beers, J. J.; El-Azzouny, M.; Visser, N. F.; Bozic, B.; Pruijn, G. J.; Heck, A. J. *Arthritis Res Ther* **2012,** *14*, R114.

(11) Pruijn, G. J. *Front Immunol* **2015,** *6*, 192.

(12) Fert-Bober, J.; Darrah, E.; Andrade, F. *Immunol Rev* **2020,** *294*, 133-147.

(13) Elkon, K. B. *Sci Transl Med* **2013,** *5*, 209fs39.

(14) Moscarello, M. A.; Mastronardi, F. G.; Wood, D. D. *Neurochem Res* **2007,** *32*, 251-6.

(15) Yang, L.; Tan, D.; Piao, H. *Neurochem Res* **2016,** *41*, 1845-56.

(16) Gs Chirivi, R. *J. Clin. Cell. Immunol.* **2013,** *04*.

(17) Bradford, C. M.; Ramos, I.; Cross, A. K.; Haddock, G.; McQuaid, S.; Nicholas, A. P.; Woodroofe, M. N. *J Neuroimmunol* **2014,** *273*, 85-95.

(18) Ishigami, A.; Ohsawa, T.; Hiratsuka, M.; Taguchi, H.; Kobayashi, S.; Saito, Y.; Murayama, S.; Asaga, H.; Toda, T.; Kimura, N.; Maruyama, N. *Journal of neuroscience research* **2005,** *80*, 120-8.

(19) Yuzhalin, A. E.; Gordon-Weeks, A. N.; Tognoli, M. L.; Jones, K.; Markelc, B.; Konietzny, R.; Fischer, R.; Muth, A.; O'Neill, E.; Thompson, P. R.; Venables, P. J.; Kessler, B. M.; Lim, S. Y.; Muschel, R. J. *Nat. Commun.* **2018,** *9*, 4783.

(20) Wang, L.; Song, G.; Zhang, X.; Feng, T.; Pan, J.; Chen, W.; Yang, M.; Bai, X.; Pang, Y.; Yu, J.; Han, J.; Han, B. *Cancer Res* **2017,** *77*, 5755-5768.

(21) Thalin, C.; Lundstrom, S.; Seignez, C.; Daleskog, M.; Lundstrom, A.; Henriksson, P.; Helleday, T.; Phillipson, M.; Wallen, H.; Demers, M. *PloS one* **2018,** *13*, e0191231.

(22) Hensen, S. M.; Pruijn, G. J. *Mol Cell Proteomics* **2014,** *13*, 388-96.

(23) Clancy, K. W.; Weerapana, E.; Thompson, P. R. *Curr Opin Chem Biol* **2016,** *30*, 1-6.

(24) Moelants, E. A.; Van Damme, J.; Proost, P. *PloS one* **2011,** *6*, e28976.

(25) Verheul, M. K.; van Veelen, P. A.; van Delft, M. A. M.; de Ru, A.; Janssen, G. M. C.; Rispens, T.; Toes, R. E. M.; Trouw, L. A. *Autoimmun Rev* **2018,** *17*, 136-141.

(26) Lee, C. Y.; Wang, D.; Wilhelm, M.; Zolg, D. P.; Schmidt, T.; Schnatbaum, K.; Reimer, U.; Ponten, F.; Uhlen, M.; Hahne, H.; Kuster, B. *Mol Cell Proteomics* **2018,** *17*, 1378-1391.

(27) Jin, Z.; Fu, Z.; Yang, J.; Troncosco, J.; Everett, A. D.; Van Eyk, J. E. *Proteomics* **2013,** *13*, 2682-91.

(28) Fert-Bober, J.; Venkatraman, V.; Hunter, C. L.; Liu, R.; Crowgey, E. L.; Pandey, R.; Holewinski, R. J.; Stotland, A.; Berman, B. P.; Van Eyk, J. E. *Journal of proteome research* **2019**.

(29) Wang, X.; Swensen, A. C.; Zhang, T.; Piehowski, P. D.; Gaffrey, M. J.; Monroe, M. E.; Zhu, Y.; Dong, H.; Qian, W. J. *Journal of proteome research* **2020,** *19*, 1863-1872.

(30) Villacres, C.; Spicer, V.; Krokhin, O. V. *Journal of proteome research* **2021**.

(31) Maurais, A. J.; Salinger, A. J.; Tobin, M.; Shaffer, S. A.; Weerapana, E.; Thompson, P. R. *Biochemistry* **2021**.

(32) Huh, S.; Hwang, D.; Kim, M. S. *Anal. Chem.* **2020,** *92*, 12975-12986.

(33) De Ceuleneer, M.; Van Steendam, K.; Dhaenens, M.; Elewaut, D.; Deforce, D. *Journal of proteome research* **2012,** *11*, 5245-51.

(34) Stensland, M.; Holm, A.; Kiehne, A.; Fleckenstein, B. *Rapid Commun Mass Spectrom* **2009,** *23*, 2754-62.

(35) De Ceuleneer, M.; De Wit, V.; Van Steendam, K.; Van Nieuwerburgh, F.; Tilleman, K.; Deforce, D. *Rapid Commun Mass Spectrom* **2011,** *25*, 1536-42.

(36) Tutturen, A. E.; Holm, A.; Fleckenstein, B. *Anal Bioanal Chem* **2013,** *405*, 9321-31.

(37) Lewallen, D. M.; Bicker, K. L.; Subramanian, V.; Clancy, K. W.; Slade, D. J.; Martell, J.; Dreyton, C. J.; Sokolove, J.; Weerapana, E.; Thompson, P. R. *ACS chemical biology* **2015,** *10*, 2520-8.

(38) Li, L.; Shi, Y.; Li, Z.; Wang, B.; Shi, X.; Ye, H.; Delafield, D.; Lv, L.; Ye, Z.; Chen, Z.; Ma, F. *Research Square* **2021**.

(39) Olsen, J. V.; Ong, S. E.; Mann, M. *Mol Cell Proteomics* **2004,** *3*, 608-14.

(40) Yang, H.; Li, Y. C.; Zhao, M. Z.; Wu, F. L.; Wang, X.; Xiao, W. D.; Wang, Y. H.; Zhang, J. L.; Wang, F. Q.; Xu, F.; Zeng, W. F.; Overall, C. M.; He, S. M.; Chi, H.; Xu, P. *Mol Cell Proteomics* **2019,** *18*, 773-785.

(41) Tallant, C.; Garcia-Castellanos, R.; Seco, J.; Baumann, U.; Gomis-Ruth, F. X. *The Journal of biological chemistry* **2006,** *281*, 17920-8.

(42) Liu, S.; Xu, F.; Yin, Y.; Zhang, J.; Wang, F.; Li, Y.; Xu, P. *Rapid Commun Mass Spectrom* **2019,** *33*, 1381-1389.

(43) Hu, H.; Zhao, W.; Zhu, M.; Zhao, L.; Zhai, L.; Xu, J. Y.; Liu, P.; Tan, M. *Anal. Chem.* **2019,** *91*, 14522-14529.

(44) Li, Q.; Zhang, Y.; Huang, J.; Wu, Z.; Tang, L.; Huang, L.; Zhang, X. *Anal. Chem.* **2020,** *92*, 4742-4748.

(45) Huesgen, P. F.; Lange, P. F.; Rogers, L. D.; Solis, N.; Eckhard, U.; Kleifeld, O.; Goulas, T.; Gomis-Ruth, F. X.; Overall, C. M. *Nature methods* **2015,** *12*, 55-8.

(46) Xiao, W.; Zhang, J.; Wang, Y.; Liu, Z.; Wang, F.; Sun, J.; Chang, L.; Xia, Z.; Li, Y.; Xu, P. *Anal. Chem.* **2019,** *91*, 15890-15898.

(47) Ma, M.; Zhao, X.; Chen, S.; Zhao, Y.; Yang, L.; Feng, Y.; Qin, W.; Li, L.; Jia, C. *Anal. Chem.* **2017,** *89*, 12909-12917.

(48) Xu, F.; Yu, L.; Peng, X.; Zhang, J.; Li, S.; Liu, S.; Yin, Y.; An, Z.; Wang, F.; Fu, Y.; Xu, P. *Journal of proteome research* **2020,** *19*, 2185-2194.

(49) Tsiatsiani, L.; Giansanti, P.; Scheltema, R. A.; van den Toorn, H.; Overall, C. M.; Altelaar, A. F.; Heck, A. J. *Journal of proteome research* **2017,** *16*, 852-861.

(50) Ma, F.; Liu, F.; Xu, W.; Li, L. *Journal of proteome research* **2018,** *17*, 2744-2754.

(51) Schiapparelli, L. M.; McClatchy, D. B.; Liu, H. H.; Sharma, P.; Yates, J. R., 3rd; Cline, H. T. *Journal of proteome research* **2014,** *13*, 3966-78.

(52) Huang da, W.; Sherman, B. T.; Lempicki, R. A. *Nat Protoc* **2009,** *4*, 44-57.

(53) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res* **2004,** *14*, 1188-90.

(54) Saveliev, S.; Bratz, M.; Zubarev, R.; Szapacs, M.; Budamgunta, H.; Urh, M. *Nature methods* **2013,** *10*, i-ii.

(55) Bennike, T.; Lauridsen, K. B.; Olesen, M. K.; Andersen, V.; Birkelund, S.; Stensballe, A. *Journal of Proteomics & Bioinformatics* **2013,** *6*.

(56) Siepen, J. A.; Keevil, E. J.; Knight, D.; Hubbard, S. J. *Journal of proteome research* **2007,** *6*, 399-408.

(57) Rodriguez, J.; Gupta, N.; Smith, R. D.; Pevzner, P. A. *Journal of proteome research* **2008,** *7*, 300-5.

(58) Unsain, N.; Stefani, F. D.; Caceres, A. *Front. Synaptic Neurosci.* **2018,** *10*, 10.

(59) Zhang, R.; Zhang, C.; Zhao, Q.; Li, D. *Sci China Life Sci* **2013,** *56*, 1076-85.

(60) Gitler, D.; Xu, Y.; Kao, H. T.; Lin, D.; Lim, S.; Feng, J.; Greengard, P.; Augustine, G. J. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2004,** *24*, 3711-20.

(61) Lugara, E.; De Fusco, A.; Lignani, G.; Benfenati, F.; Humeau, Y. *Front. Cell. Neurosci.* **2019,** *13*, 220.

(62) Boggs, J. M. *Cellular and molecular life sciences : CMLS* **2006,** *63*, 1945-61.

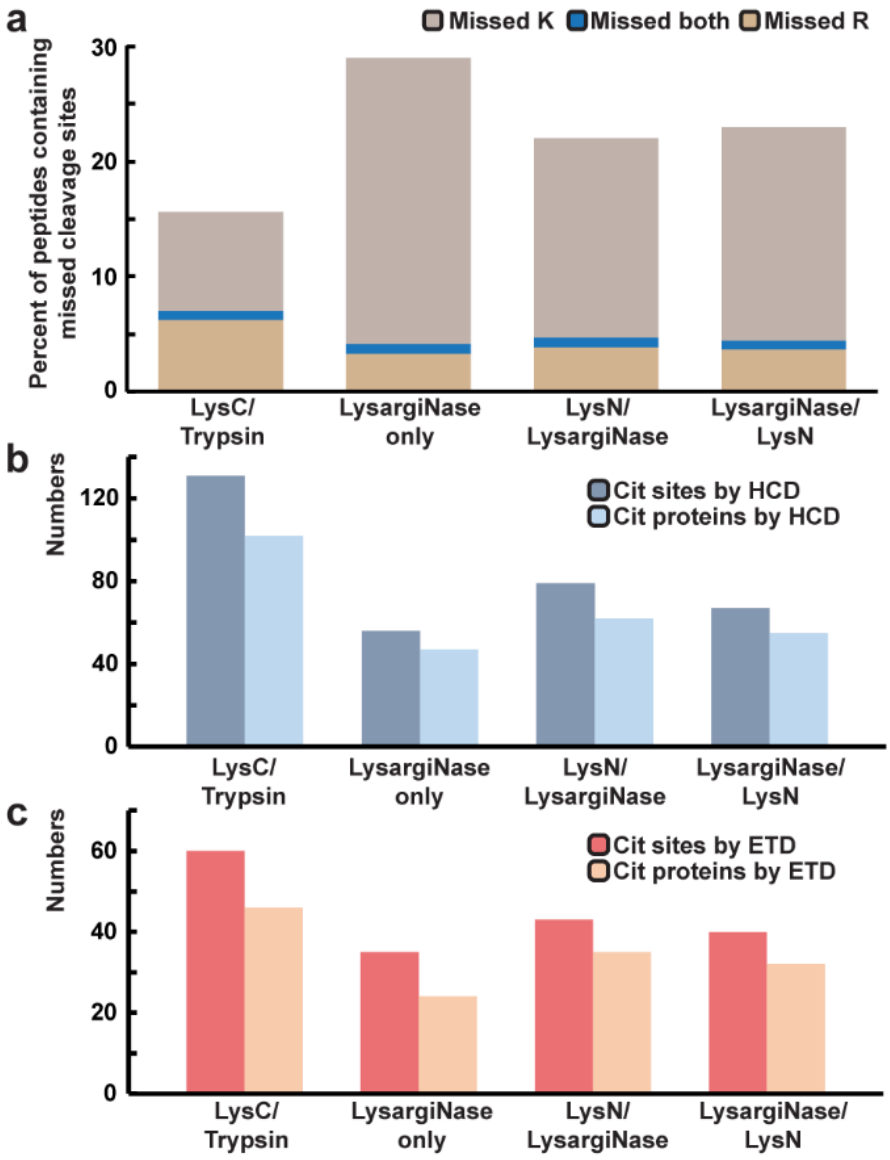

Figure 1. Optimization of LysargiNase digestion protocols for citrullination analysis. a, Missed cleavage rate of lysine and arginine residues using different digestion methods. **b,** Number of identified citrullination sites and citrullinated proteins using different digestion methods and HCD fragmentation technique. **c,** Number of identified citrullination sites and citrullinated proteins using different digestion methods and ETD fragmentation technique.

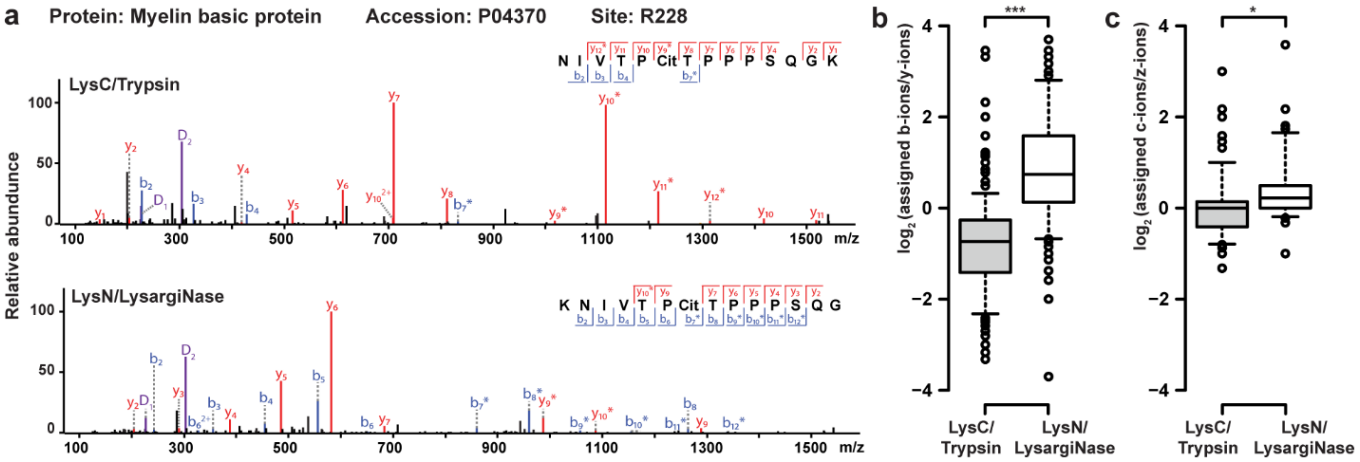

Figure 2. Opposite fragmentation characteristics of tryptic or LysargiNase peptides. a, Example spectra of the same citrullination site on myelin basic protein (R228Cit) identified with two different digestion methods. Boxplots showing the ratios of assigned b-ions/y-ions by HCD fragmentation (**b**) or c-ions/z-ions by ETD fragmentation (**c**) in all identified citrullinated peptides using two different digestion methods. Top and bottom of boxes indicate 3rd and 1st quartile, respectively, and whiskers extend to 5th and 95th percentile. Horizontal lines within boxes denote median. Significance level is marked with an asterisk (*p value <0.05, ***p value <0.001).

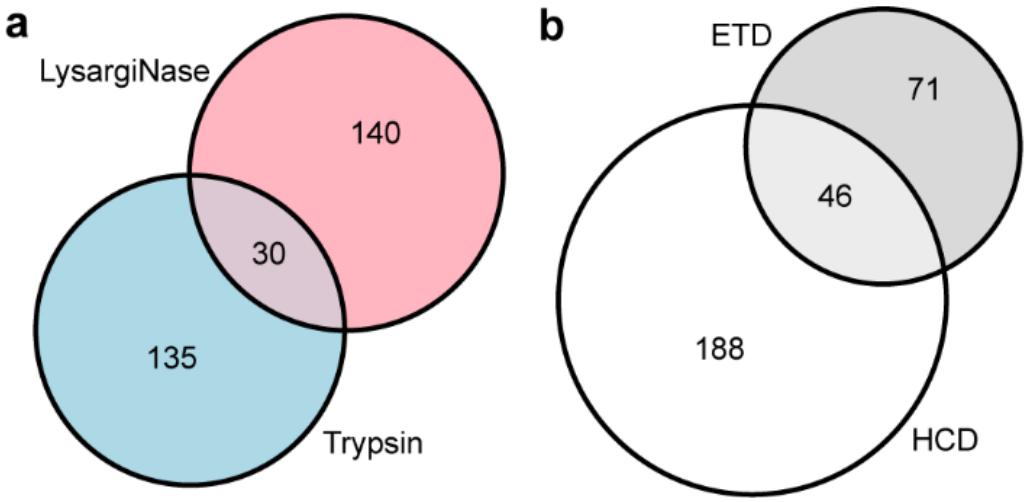

Figure 3. Complementarity of citrullination identifications using different digestion methods and fragmentation techniques. a, Venn diagram showing the overlap of identified citrullination sites using different digestion methods. LysargiNase digestion results are based on combination of three digestion protocols. **b,** Venn diagram showing the overlap of identified citrullination sites using different fragmentation techniques.

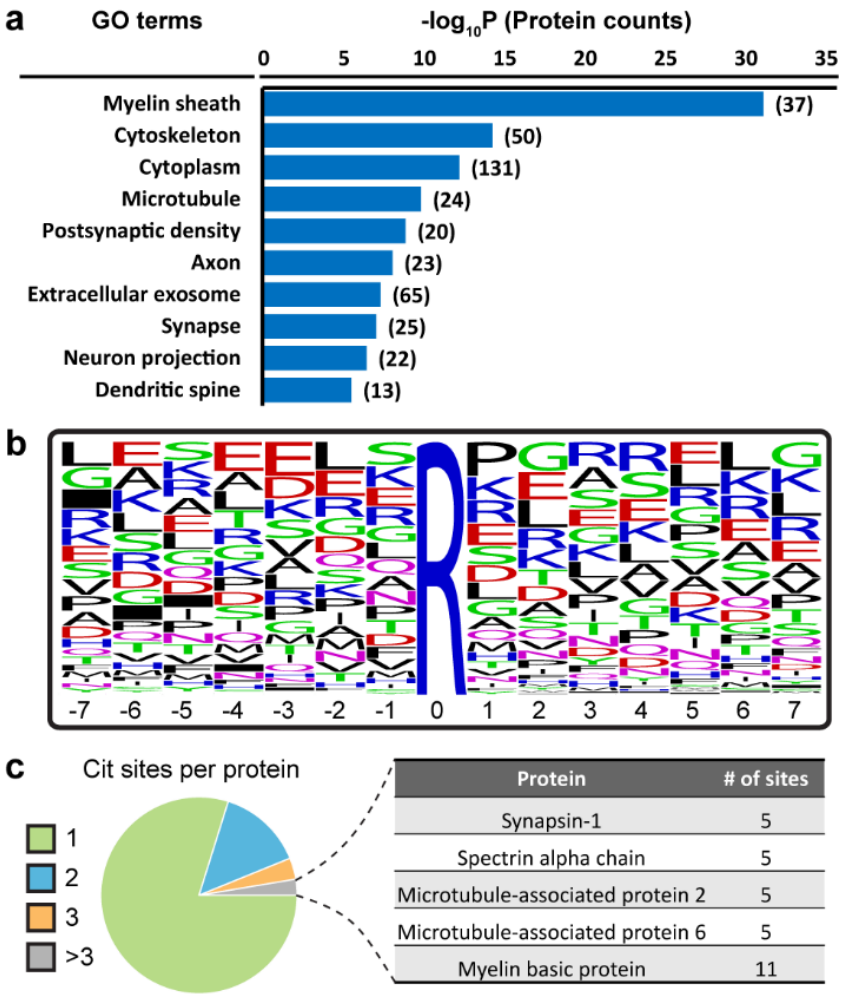

Figure 4. Enhanced profiling of protein citrullination in mouse brain. a, Gene ontology analysis of all identified citrullinated proteins showing the significantly enriched cellular components (Fisher's exact test, *p* value <0.05, *p* values were adjusted with Benjamini-Hochberg correction). Ten most significant terms are shown. **b,** Sequence motif of all identified citrullinated peptides. Citrullination sites are centered in the middle as "0" position. The height of letters indicates the relative frequency of each amino acid at certain positions. Black rectangles within sequence indicate vacancy due to peptides being identified at protein termini. **c,** Distribution of the number of citrullination sites per citrullinated proteins identified. Example proteins with multiple identified sites are shown in the table.

Supplemental Information

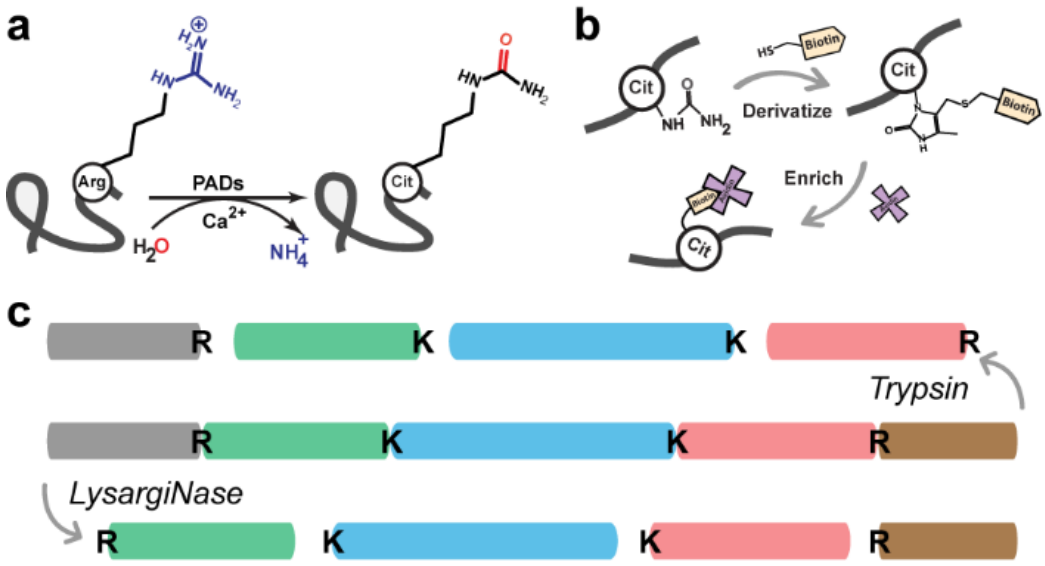

Figure S1. a, Conversion of peptidyl arginine residues into citrulline catalyzed by PAD enzymes. **b,** Derivatization and enrichment strategy of citrullinated peptides using biotin thiol tag. **c,** Cleavage of a protein by either trypsin or LysargiNase at C- or N-terminal sides, respectively.

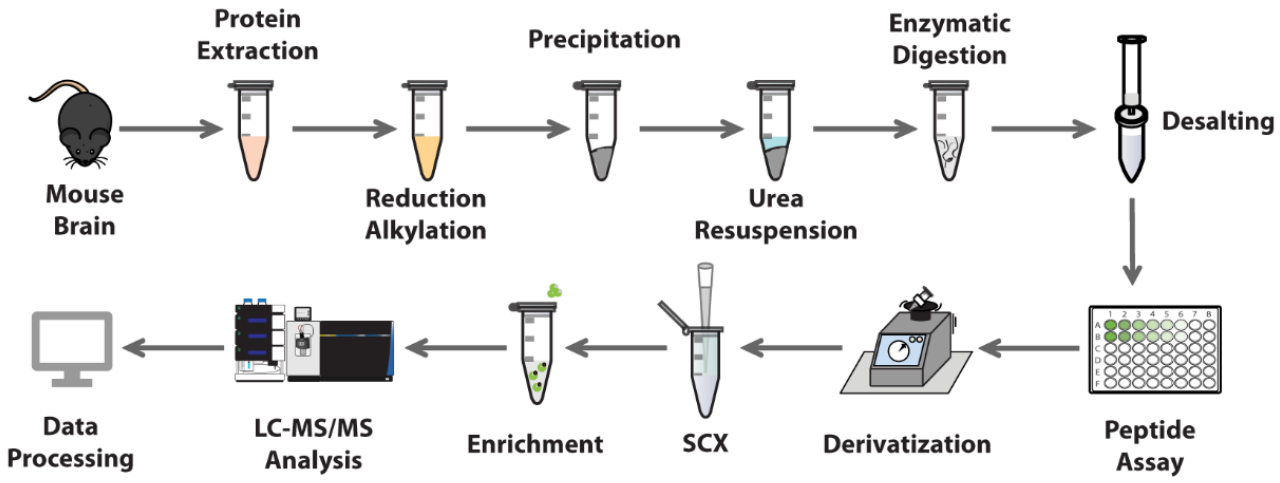

Figure S2. Workflow of citrullination analysis from complex biological samples using bottom-up proteomics and biotin-thiol tag assisted mass spectrometry approaches. SCX, strong cation exchange.

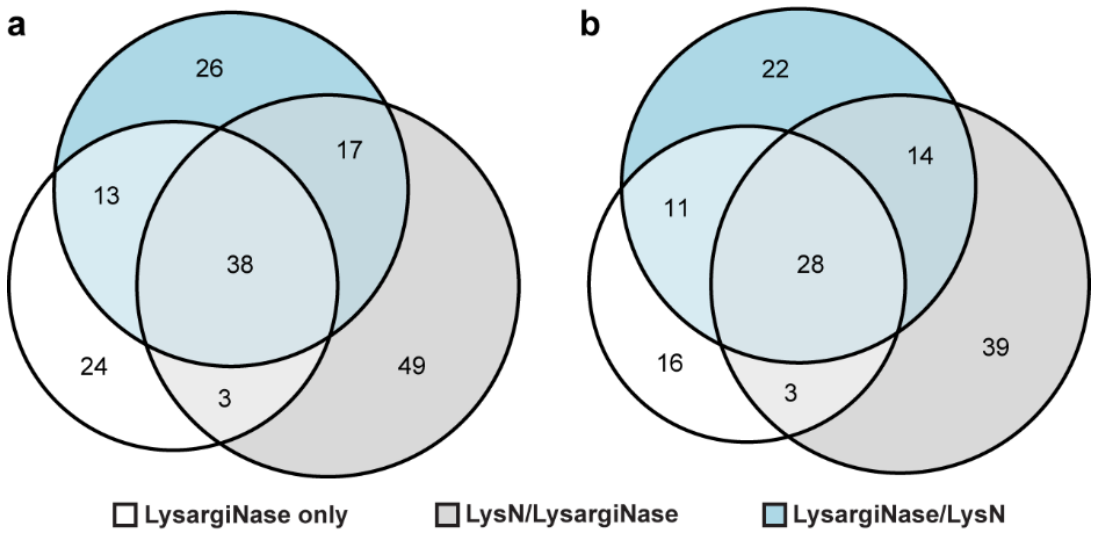

Figure S3. a, Venn diagram showing the overlap of identified citrullination sites using three different LysargiNase digestion protocols. **b,** Venn diagram showing the overlap of identified citrullinated proteins using three different LysargiNase digestion protocols.

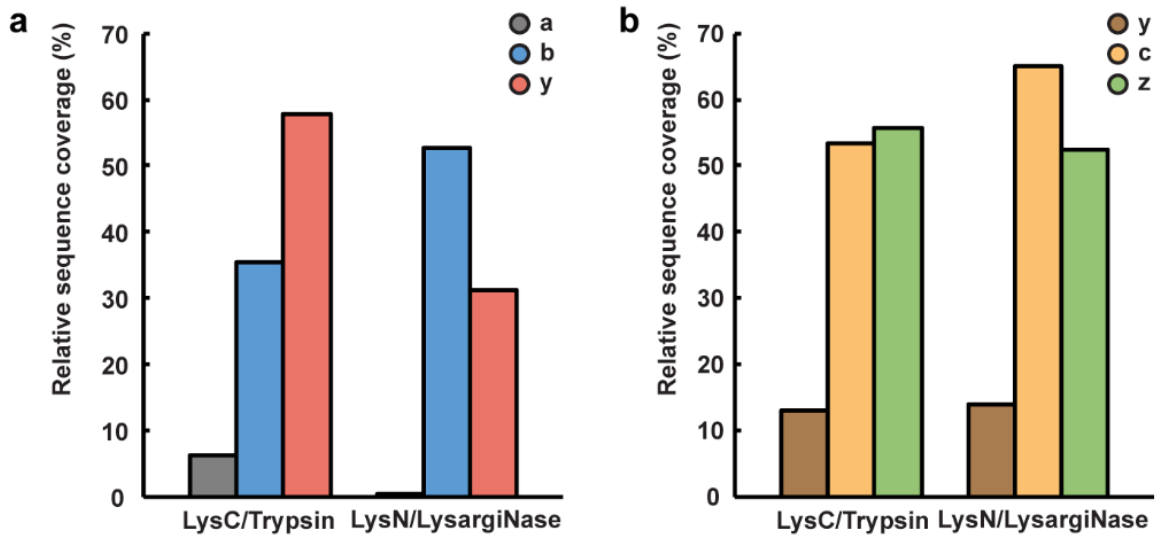

Figure S4. Citrullinated peptide sequence coverage of fragment ions using two different digestion methods. **a,** Sequence coverage of a-, b- and y-ions upon HCD fragmentation. **b,** Sequence coverage of y-, c- and z-ions upon ETD fragmentation.

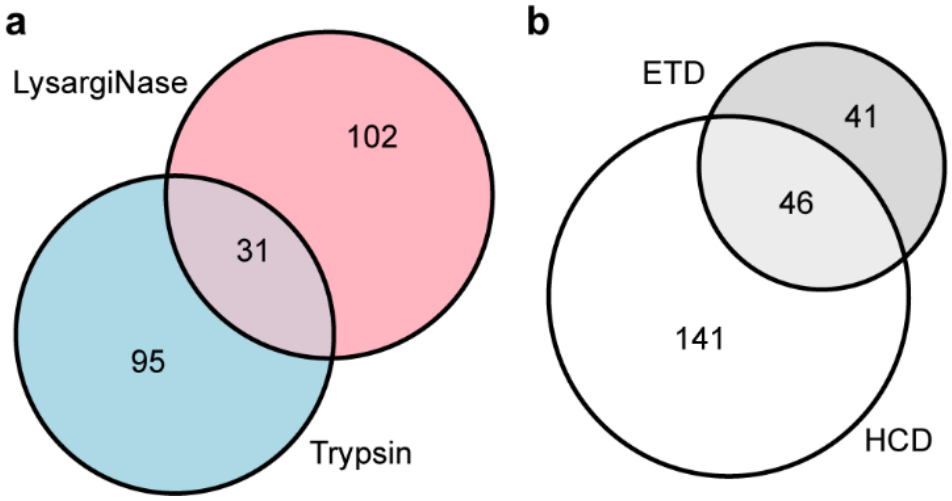

Figure S5. a, Venn diagram showing the overlap of identified citrullinated proteins using different digestion methods. LysargiNase digestion results are based on combination of three digestion protocols. **b,** Venn diagram showing the overlap of identified citrullinated proteins using different fragmentation techniques.

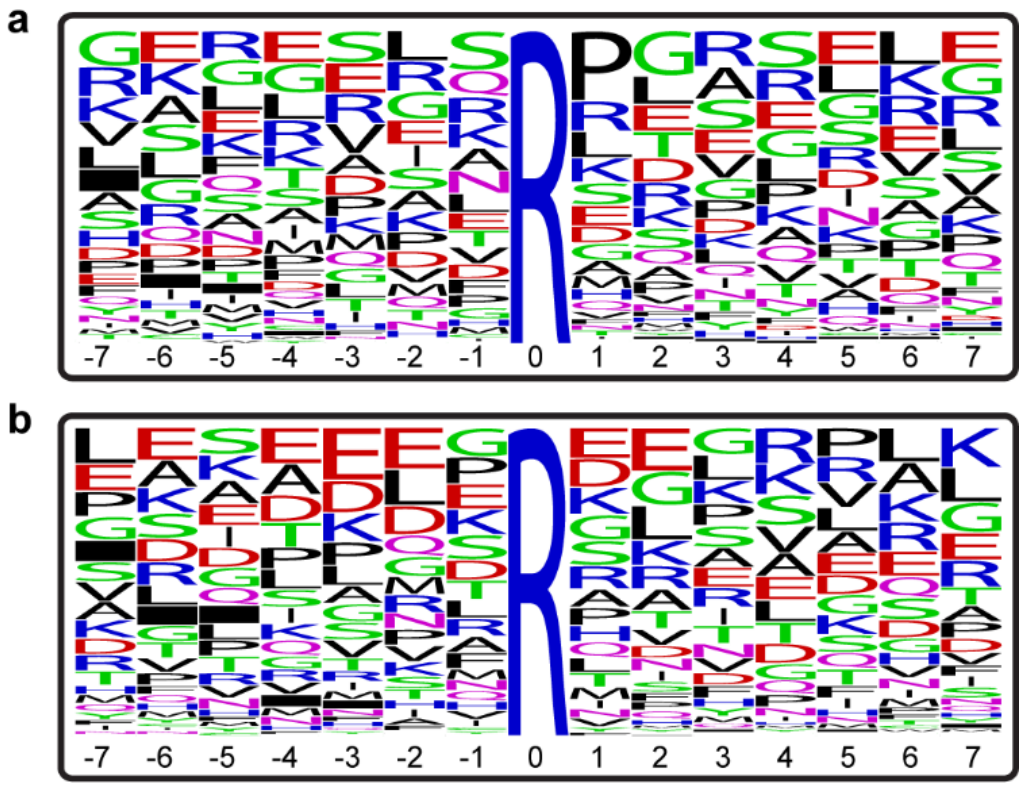

Figure S6. Sequence motif of identified citrullinated peptides using trypsin (**a**) or LysargiNase-based (**b**) digestion methods. Citrullination sites are centered in the middle as "0" position. The height of letters indicates the relative frequency of each amino acid at certain positions. Black rectangles within sequence indicate vacancy due to peptides being identified at protein termini.

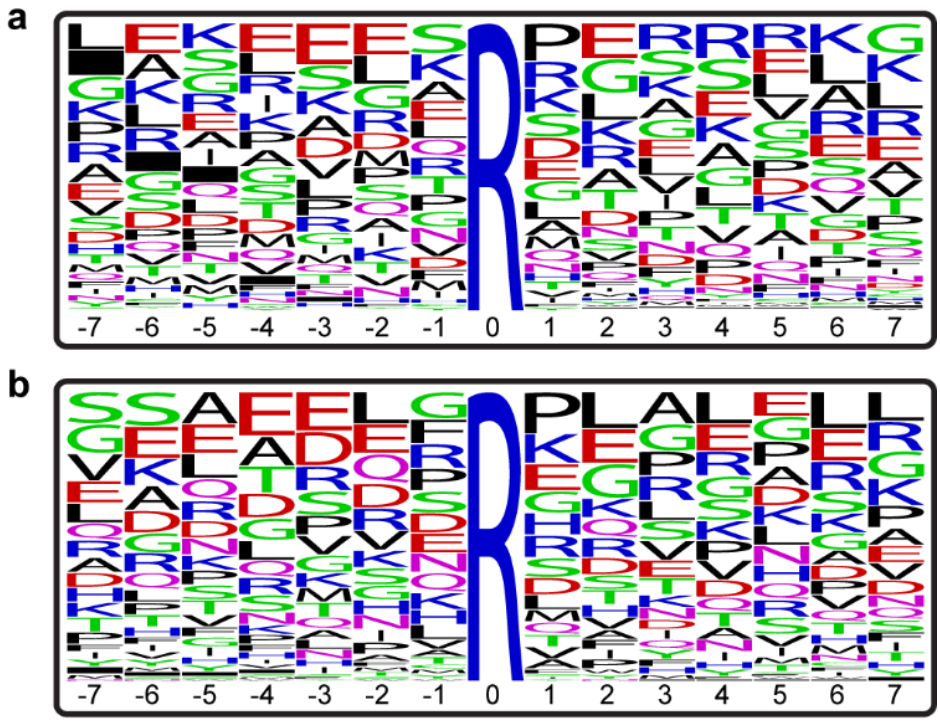

Figure S7. Sequence motif of identified citrullinated peptides using HCD (**a**) or ETD (**b**) fragmentation techniques. Citrullination sites are centered in the middle as "0" position. The height of letters indicates the relative frequency of each amino acid at certain positions. Black rectangles within sequence indicate vacancy due to peptides being identified at protein termini.

Table S1. List of all identified citrullination sites

Accession	Description	Position
A2AJI0	MAP7 domain-containing protein 1	147
A2AJI0	MAP7 domain-containing protein 1	148
A2AJI0	MAP7 domain-containing protein 1	188
A2APY7	Arginine-hydroxylase NDUFAF5, mitochondrial	49
A2AQ25	Sickle tail protein	61
B1B1A0	Lethal(3)malignant brain tumor-like protein 4	6
B9EJ80	PDZ domain-containing protein 8	996
B9EKI3	TATA element modulatory factor	620
D3YZP9	Coiled-coil domain-containing protein 6	303
D3YZP9	Coiled-coil domain-containing protein 6	307
E9Q5M6	Cilia- and flagella-associated protein 44	679
E9Q8I9	Protein furry homolog	1937
E9QAM5	Helicase with zinc finger domain 2	2750
O08528	Hexokinase-2	546
O08553	Dihydropyrimidinase-related protein 2	227
O08553	Dihydropyrimidinase-related protein 2	565
O09107	Insulin-like 3	59
O35594	Intraflagellar transport protein 81 homolog	188
O35738	Krueppel-like factor 12	7
O54782	Epididymis-specific alpha-mannosidase	715
O88398	Advillin	195
O88737	Protein bassoon	3605
O88737	Protein bassoon	3654
O88737	Protein bassoon	3661
O88737	Protein bassoon	3685
O88876	Short-chain dehydrogenase/reductase 3	52
O88935	Synapsin-1	534
O88935	Synapsin-1	547
O88935	Synapsin-1	587
O88935	Synapsin-1	602
O88935	Synapsin-1	680
P00158	Cytochrome b	5
P03995	Glial fibrillary acidic protein	20
P03995	Glial fibrillary acidic protein	413
P04370	Myelin basic protein	143
P04370	Myelin basic protein	157
P04370	Myelin basic protein	163
P04370	Myelin basic protein	165
P04370	Myelin basic protein	175
P04370	Myelin basic protein	181

P04370	Myelin basic protein	195
P04370	Myelin basic protein	210
P04370	Myelin basic protein	228
P04370	Myelin basic protein	242
P04370	Myelin basic protein	249
P05063	Fructose-bisphosphate aldolase C	56
P05064	Fructose-bisphosphate aldolase A	56
P05064	Fructose-bisphosphate aldolase A	92
P05064	Fructose-bisphosphate aldolase A	201
P05213	Tubulin alpha-1B chain	64
P05213	Tubulin alpha-1B chain	221
P05214	Tubulin alpha-3 chain	221
P06745	Glucose-6-phosphate isomerase	472
P07901	Heat shock protein HSP 90-alpha	457
P10126	Elongation factor 1-alpha 1	240
P10637	Microtubule-associated protein tau	5
P10637	Microtubule-associated protein tau	358
P11798	Calcium/calmodulin-dependent protein kinase type II subunit alpha	311
P14094	Sodium/potassium-transporting ATPase subunit beta-1	91
P14094	Sodium/potassium-transporting ATPase subunit beta-1	136
P14094	Sodium/potassium-transporting ATPase subunit beta-1	143
P14115	60S ribosomal protein L27a	132
P14211	Calreticulin	366
P14873	Microtubule-associated protein 1B	2395
P15105	Glutamine synthetase	341
P15919	V(D)J recombination-activating protein 1	703
P16330	2',3'-cyclic-nucleotide 3'-phosphodiesterase	93
P16330	2',3'-cyclic-nucleotide 3'-phosphodiesterase	127
P16546	Spectrin alpha chain, non-erythrocytic 1	285
P16546	Spectrin alpha chain, non-erythrocytic 1	439
P16546	Spectrin alpha chain, non-erythrocytic 1	1427
P16546	Spectrin alpha chain, non-erythrocytic 1	2006
P16546	Spectrin alpha chain, non-erythrocytic 1	2199
P17182	Alpha-enolase	50
P17183	Gamma-enolase	15
P19157	Glutathione S-transferase P 1	101
P20152	Vimentin	28
P20152	Vimentin	450
P20357	Microtubule-associated protein 2	1460
P20357	Microtubule-associated protein 2	1488
P20357	Microtubule-associated protein 2	1582
P20357	Microtubule-associated protein 2	1662

P20357	Microtubule-associated protein 2	1766
P20917	Myelin-associated glycoprotein	563
P20917	Myelin-associated glycoprotein	588
P21271	Unconventional myosin-Vb	1198
P23242	Gap junction alpha-1 protein	362
P30275	Creatine kinase U-type, mitochondrial	310
P31316	GS homeobox 2	3
P31786	Acyl-CoA-binding protein	44
P31938	Dual specificity mitogen-activated protein kinase kinase 1	47
P35802	Neuronal membrane glycoprotein M6-a	269
P39053	Dynamin-1	343
P39053	Dynamin-1	818
P39053	Dynamin-1	835
P39749	Flap endonuclease 1	98
P43006	Excitatory amino acid transporter 2	570
P45376	Aldo-keto reductase family 1 member B1	70
P46425	Glutathione S-transferase P 2	202
P46460	Vesicle-fusing ATPase	314
P48962	ADP/ATP translocase 1	259
P49312	Heterogeneous nuclear ribonucleoprotein A1	97
P55937	Golgin subfamily A member 3	1098
P56564	Excitatory amino acid transporter 1	10
P60761	Neurogranin	38
P60761	Neurogranin	68
P60879	Synaptosomal-associated protein 25	17
P62069	Ubiquitin carboxyl-terminal hydrolase 46	241
P62069	Ubiquitin carboxyl-terminal hydrolase 46	243
P62259	14-3-3 protein epsilon	141
P62631	Elongation factor 1-alpha 2	240
P62754	40S ribosomal protein S6	131
P63017	Heat shock cognate 71 kDa protein	77
P63024	Vesicle-associated membrane protein 3	53
P63328	Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	42
P68368	Tubulin alpha-4A chain	79
P68372	Tubulin beta-4B chain	251
P68372	Tubulin beta-4B chain	359
P70275	Semaphorin-3E	434
P81069	GA-binding protein subunit beta-2	15
P84086	Complexin-2	96
P84099	60S ribosomal protein L19	81
P84099	60S ribosomal protein L19	88
P97315	Cysteine and glycine-rich protein 1	99

P97433	Rho guanine nucleotide exchange factor 28	1515
P97858	Solute carrier family 35 member B1	5
P99024	Tubulin beta-5 chain	86
P99029	Peroxiredoxin-5, mitochondrial	172
Q00547	Hyaluronan mediated motility receptor	329
Q03265	ATP synthase subunit alpha, mitochondrial	231
Q03265	ATP synthase subunit alpha, mitochondrial	405
Q04447	Creatine kinase B-type	13
Q0VFX2	Cilia- and flagella-associated protein 157	227
Q11011	Puromycin-sensitive aminopeptidase	806
Q3U2S8	Voltage-gated hydrogen channel 1	11
Q3U2S8	Voltage-gated hydrogen channel 1	12
Q3U319	E3 ubiquitin-protein ligase BRE1B	713
Q3UGF1	WD repeat-containing protein 19	1174
Q3UMQ8	H/ACA ribonucleoprotein complex non-core subunit NAF1	319
Q3URK3	Methylcytosine dioxygenase TET1	5
Q3USZ2	Upstream-binding factor 1-like protein 1	149
Q3UZP0	MARVEL domain-containing protein 2	7
Q3V0F0	RIMS-binding protein 3	513
Q3V300	Kinesin-like protein KIF22	237
Q5DTW2	Scm-like with four MBT domains protein 2	777
Q5DU31	Interactor protein for cytohesin exchange factors 1	361
Q5DU37	Zinc finger FYVE domain-containing protein 26	989
Q5HZG4	Transcription initiation factor TFIID subunit 3	561
Q5HZG4	Transcription initiation factor TFIID subunit 3	563
Q5RJI5	Serine/threonine-protein kinase BRSK1	334
Q5RJI5	Serine/threonine-protein kinase BRSK1	337
Q5SNZ0	Girdin	261
Q5SNZ0	Girdin	262
Q5SNZ0	Girdin	264
Q5SUS0	F-box/WD repeat-containing protein 10	4
Q5SV06	Spermatogenesis-associated protein 22	3
Q60598	Src substrate cortactin	97
Q60674	Nuclear receptor subfamily 1 group D member 2	150
Q60675	Laminin subunit alpha-2	1880
Q60737	Casein kinase II subunit alpha	8
Q60862	Origin recognition complex subunit 2	166
Q60930	Voltage-dependent anion-selective channel protein 2	264
Q60931	Voltage-dependent anion-selective channel protein 3	252
Q61097	Kinase suppressor of Ras 1	305
Q61097	Kinase suppressor of Ras 1	326
Q61241	Testis-specific serine/threonine-protein kinase 1	163

Q61361	Brevican core protein	866
Q61655	ATP-dependent RNA helicase DDX19A	425
Q61792	LIM and SH3 domain protein 1	130
Q62108	Disks large homolog 4	110
Q62167	ATP-dependent RNA helicase DDX3X	622
Q62261	Spectrin beta chain, non-erythrocytic 1	1834
Q62261	Spectrin beta chain, non-erythrocytic 1	1903
Q62433	Protein NDRG1	3
Q64332	Synapsin-2	429
Q64521	Glycerol-3-phosphate dehydrogenase, mitochondrial	627
Q68FD5	Clathrin heavy chain 1	437
Q68FD5	Clathrin heavy chain 1	574
Q69ZK7	Protein FAM214A	72
Q6GYP7	Ral GTPase-activating protein subunit alpha-1	847
Q6KAQ7	ZZ-type zinc finger-containing protein 3	12
Q6NXY1	TBC1 domain family member 31	734
Q6NXY1	TBC1 domain family member 31	735
Q6NZP1	DNA annealing helicase and endonuclease ZRANB3	181
Q6PB70	Anoctamin-8	219
Q6PCX7	Repulsive guidance molecule A	7
Q6PDC8	Major facilitator superfamily domain-containing protein 4A	368
Q6PDH0	Pleckstrin homology-like domain family B member 1	897
Q6PHS6	Sorting nexin-13	859
Q6PHS6	Sorting nexin-13	861
Q6PHS6	Sorting nexin-13	863
Q6PIC6	Sodium/potassium-transporting ATPase subunit alpha-3	43
Q6PIC6	Sodium/potassium-transporting ATPase subunit alpha-3	648
Q6X6Z7	Tektin-3	205
Q6X6Z7	Tektin-3	399
Q6ZPR1	Cilia- and flagella-associated protein 97	169
Q6ZQA0	Neurobeachin-like protein 2	2572
Q6ZWY9	Histone H2B type 1-C/E/G	80
Q70FJ1	A-kinase anchor protein 9	8
Q7TNU6	Zinc finger protein 250	133
Q7TPR4	Alpha-actinin-1	614
Q7TQD2	Tubulin polymerization-promoting protein	164
Q7TQF7	Amphiphysin	48
Q7TSJ2	Microtubule-associated protein 6	159
Q7TSJ2	Microtubule-associated protein 6	195
Q7TSJ2	Microtubule-associated protein 6	385
Q7TSJ2	Microtubule-associated protein 6	431
Q7TSJ2	Microtubule-associated protein 6	435

Q80T11	Usher syndrome type-1G protein homolog	141
Q80TE0	RNA polymerase II-associated protein 1	5
Q80TM9	Nischarin	938
Q80U49	Centrosomal protein of 170 kDa protein B	443
Q80VC9	Calmodulin-regulated spectrin-associated protein 3	794
Q8BGV5	Zinc finger protein 8	505
Q8BKA3	Serine/arginine repetitive matrix protein 4	216
Q8BKI2	Trinucleotide repeat-containing gene 6B protein	1142
Q8BNX7	Nutritionally-regulated adipose and cardiac-enriched protein	6
Q8BPM0	Disheveled-associated activator of morphogenesis 1	752
Q8BPP1	Protein mab-21-like 2	32
Q8BQP8	Rab11 family-interacting protein 4	503
Q8BSS9	Liprin-alpha-2	159
Q8BVD5	MAGUK p55 subfamily member 7	453
Q8BZ98	Dynamin-3	217
Q8BZH4	Pogo transposable element with ZNF domain	853
Q8C4J0	Coiled-coil domain-containing protein 60	56
Q8C624	Spermatogenesis-associated protein 33	9
Q8C6M1	Ubiquitin carboxyl-terminal hydrolase 20	695
Q8C8R3	Ankyrin-2	1849
Q8CDN1	Uncharacterized protein C3orf20 homolog	640
Q8CDN1	Uncharacterized protein C3orf20 homolog	643
Q8CE64	Zinc finger protein 276	480
Q8CHB8	Tubulin polyglutamylase TTLL5	565
Q8CHB8	Tubulin polyglutamylase TTLL5	567
Q8CHH5	BRD4-interacting chromatin-remodeling complex-associated protein-like	763
Q8CHH9	Septin-8	84
Q8CJ19	[F-actin]-monooxygenase MICAL3	1472
Q8K0T0	Reticulon-1	375
Q8K0X8	Fasciculation and elongation protein zeta-1	257
Q8K135	Dyslexia-associated protein KIAA0319-like protein	4
Q8K1N2	Pleckstrin homology-like domain family B member 2	1036
Q8R089	Probable fibrosin-1	258
Q8R0A5	Transcription elongation factor A protein-like 3	77
Q8R0F9	SEC14-like protein 4	161
Q8R331	Zinc finger HIT domain-containing protein 1	26
Q8VC56	E3 ubiquitin-protein ligase RNF8	290
Q91V79	Fat storage-inducing transmembrane protein 1	3
Q91YE8	Synaptopodin-2	256
Q91YP2	Neurolysin, mitochondrial	119
Q921U8	Smoothelin	312
Q921U8	Smoothelin	330

Q923T9	Calcium/calmodulin-dependent protein kinase type II su	494
Q99J87	Probable ATP-dependent RNA helicase DHX58	492
Q99KI0	Aconitate hydratase, mitochondrial	58
Q99KI0	Aconitate hydratase, mitochondrial	61
Q99KI0	Aconitate hydratase, mitochondrial	534
Q99M71	Mammalian ependymin-related protein 1	4
Q99M71	Mammalian ependymin-related protein 1	7
Q99PJ0	Neurotrimin	62
Q99PJ1	Protocadherin-15	1579
Q99PV0	Pre-mRNA-processing-splicing factor 8	1418
Q9CQF0	39S ribosomal protein L11, mitochondrial	6
Q9CQT5	Proteasome maturation protein	4
Q9CUN6	E3 ubiquitin-protein ligase SMURF1	7
Q9CZB0	Succinate dehydrogenase cytochrome b560 subunit, mitochondrial	50
Q9CZC8	Secernin-1	366
Q9D051	Pyruvate dehydrogenase E1 component subunit beta, mito	49
Q9D0C4	tRNA (guanine(37)-N1)-methyltransferase	113
Q9D1Z3	ATP synthase subunit C lysine N-methyltransferase	3
Q9D3E6	Cohesin subunit SA-1	1167
Q9D439	Cilia- and flagella-associated protein 53	250
Q9D6R2	Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	125
Q9D6R2	Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	360
Q9D8B7	Junctional adhesion molecule C	303
Q9D975	Sulfiredoxin-1	4
Q9D9K3	Cell death regulator Aven	13
Q9D9K3	Cell death regulator Aven	14
Q9D9K3	Cell death regulator Aven	17
Q9DBG3	AP-2 complex subunit beta	371
Q9DBJ1	Phosphoglycerate mutase 1	240
Q9DCT2	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	230
Q9EPB4	Apoptosis-associated speck-like protein containing a CARD	5
Q9ERD7	Tubulin beta-3 chain	86
Q9JI08	Bridging integrator 3	91
Q9JI91	Alpha-actinin-2	67
Q9JIS5	Synaptic vesicle glycoprotein 2A	136
Q9JIS5	Synaptic vesicle glycoprotein 2A	145
Q9JKK7	Tropomodulin-2	159
Q9JLM8	Serine/threonine-protein kinase DCLK1	161
Q9JM52	Misshapen-like kinase 1	761
Q9JM96	Cdc42 effector protein 4	35
Q9QUI0	Transforming protein RhoA	145
Q9QXS6	Drebrin	151

Q9QXS6	Drebrin	286
Q9QYC0	Alpha-adducin	594
Q9QYR6	Microtubule-associated protein 1A	2390
Q9R0K7	Plasma membrane calcium-transporting ATPase 2	249
Q9R0U0	Serine/arginine-rich splicing factor 10	180
Q9R0U0	Serine/arginine-rich splicing factor 10	182
Q9R269	Periplakin	1390
Q9WTT4	V-type proton ATPase subunit G 2	48
Q9WUK2	Eukaryotic translation initiation factor 4H	22
Q9WUK2	Eukaryotic translation initiation factor 4H	136
Q9WUT3	Ribosomal protein S6 kinase alpha-2	727
Q9WUT3	Ribosomal protein S6 kinase alpha-2	732
Q9Z2Q2	Lysine-rich nucleolar protein 1	295

Table S2. List of enriched Gene Ontology terms

Category	Term	Count	*p* value[a]
Cellular Component	myelin sheath	37	2.90E-31
	cytoskeleton	50	1.00E-14
	cytoplasm	131	1.00E-12
	microtubule	24	2.20E-10
	postsynaptic density	20	1.90E-09
	axon	23	1.20E-08
	extracellular exosome	65	6.40E-08
	synapse	25	1.20E-07
	neuron projection	22	4.50E-07
	dendritic spine	13	3.80E-06
	cell projection	27	3.80E-06
	neuronal cell body	23	4.50E-06
	synaptic vesicle	12	4.70E-06
	cell-cell adherens junction	17	1.40E-05
	cytosol	44	2.60E-05
	focal adhesion	18	4.50E-05
	dendrite	20	5.70E-05
	perinuclear region of cytoplasm	24	6.40E-05
	cell junction	24	1.10E-04
	mitochondrion	40	2.80E-04
	terminal bouton	9	6.60E-04
	nucleus	96	8.00E-04
	extracellular vesicle	6	3.20E-03
	protein complex	19	3.70E-03
	membrane	104	5.00E-03
	cortical cytoskeleton	5	5.00E-03
	Schmidt-Lanterman incisure	4	7.20E-03
	tubulin complex	3	9.00E-03
	cell body	7	1.20E-02
	paranodal junction	3	1.40E-02
	photoreceptor inner segment	5	2.10E-02
	Z disc	7	2.90E-02
	microtubule associated complex	4	3.40E-02
	axonal growth cone	4	3.60E-02
	cytoplasmic ribonucleoprotein granule	4	3.60E-02
	dendritic spine head	3	4.00E-02
	mitochondrial inner membrane	12	4.00E-02
	dendritic shaft	5	4.40E-02
	melanosome	6	4.90E-02

Biological Process	glycolytic process	7	3.60E-03
	microtubule-based process	7	4.00E-03
	microtubule cytoskeleton organization	9	5.70E-03
	regulation of neuron projection development	6	6.30E-03
	positive regulation of potassium ion transmembrane transporter activity	4	1.20E-02
	sensory perception of sound	9	3.80E-02
Molecular Function	protein binding	82	3.40E-05
	cadherin binding involved in cell-cell adhesion	17	3.40E-05
	microtubule binding	14	8.20E-05
	protein kinase binding	19	2.70E-04
	structural constituent of cytoskeleton	9	2.70E-04
	calmodulin binding	12	5.60E-04
	actin binding	16	5.60E-04
	nucleotide binding	44	9.60E-04
	GTPase activity	12	1.50E-03
	syntaxin-1 binding	5	5.30E-03
	protein complex binding	14	1.00E-02
	spectrin binding	5	1.50E-02
	poly(A) RNA binding	27	1.60E-02
	MHC class II protein complex binding	4	1.80E-02
	protein C-terminus binding	10	2.00E-02
	ATP binding	32	3.00E-02
	double-stranded RNA binding	6	3.10E-02
	GTP binding	13	3.90E-02
KEGG Pathway	Carbon metabolism	11	4.10E-04
	Synaptic vesicle cycle	8	1.30E-03
	Glycolysis / Gluconeogenesis	8	1.30E-03
	Biosynthesis of amino acids	8	2.40E-03
	Gap junction	8	4.30E-03
	HIF-1 signaling pathway	8	1.00E-02
	Biosynthesis of antibiotics	11	1.10E-02
	Endocrine and other factor-regulated calcium reabsorption	6	1.20E-02
	Huntington's disease	10	2.10E-02
	cGMP-PKG signaling pathway	9	2.20E-02
	Phagosome	9	2.70E-02

[a]*p* values are adjusted by Benjamini–Hochberg correction

Chapter 6

Mass spectrometric profiling of citrullination alterations in cerebrospinal fluid of patients with Alzheimer's disease

Abstract

Citrullination is an important protein post-translational modification (PTM) which has profound influence on protein conformations and protein-protein interactions. Dysregulation of citrullination has been associated with various disease development and progression including Alzheimer's disease (AD). However, previous studies only identified few citrullinated proteins mainly due to the limitations of effective analytical tools. Therefore, a systematic and comprehensive investigation is lacking yet in great need. Recently, we developed a novel biotin thiol tag which enabled large-scale and confident identification of this PTM. In this work, we utilize this approach to investigate the citrullination changes with the progression of AD using human cerebrospinal fluid (CSF) samples. In total, we identify 135 citrullination sites from 68 proteins with high confidence, which represents the first report of this PTM in CSF and a pilot study to explore the association of citrullination with AD at the site-specific level. We demonstrate previously unknown biological functions of these citrullinated proteins and find an overall lower expression level of citrullination in AD compared to healthy individuals, which may provide new insights into the disease pathogenesis. Specifically, we identify novel citrullination sites on a few important proteins such as apolipoprotein E, clusterin and complement C3, which will greatly benefit future studies for diagnosis and drug development of AD.

Introduction

Alzheimer's disease (AD) is one of the leading causes of death in the world. In past decades, tremendous effort has been put into the examination of AD pathogenesis and the biomarker discovery for early diagnosis and clinical treatment. Among many proposed theories, the amyloid hypothesis and tau hypothesis are two mainstream theories regarding AD pathology, and therefore amyloid β and tau protein are two well-accepted biomarkers for the diagnosis and monitoring of AD progression as well as potential therapeutic targets for pharmaceutical drug design.[1-3] However, the development of therapeutic approaches targeting these two biomarkers is far from being successful, suggesting that amyloid and tau hypothesis cannot fully explain for the complex AD pathogenesis. Therefore, exploration of synergic mechanisms of this disease is still in great need and the discovery of additional biomarkers is highly desired. Alternatively, numerous protein post-translational modifications (PTMs) have been indicated to play important roles in the development and progression of AD such as phosphorylation, glycosylation, methylation and acetylation.[4-6]

Protein citrullination is a critical PTM leading to the conversion of positively charged arginine residues to neutral citrulline residues,[7-12] which greatly influences on protein conformations, protein-protein interactions and protein functions.[9, 11, 13] Dysregulation of protein citrullination was therefore indicated to participate in many disease development and progression including rheumatoid arthritis[14-19], multiple sclerosis[20-23], and cancers[24-26]. Interestingly, several previous studies also indicated the possible association of protein citrullination with AD progression, though further investigation was impeded primarily due to a lack of methods for global citrullination analysis.[27-31] For example, Ishigami et al. found increased PAD2 expression in AD patients and identified citrullinated vimentin and glial fibrillary acidic protein (GFAP) in human brain hippocampus and cortex.[28-29] Similarly, Acharya et al. observed colocalization of

PAD with citrullinated proteins in both human AD and healthy brains.[31] However, these studies only incorporated limited number of specimens and identified very few proteins with limited modification site information.

Currently, antibody-based methods are the most prevalent in detecting citrullination.[32-34] However, these methods are intrinsically incapable for large-scale analysis, and it is difficult to pinpoint specific modification sites. On the other hand, mass spectrometry (MS)-based strategies also face great challenges in citrullination profiling due to the tiny mass shift and low-abundance nature of this PTM.[33, 35] To combat these difficulties, tremendous effort has been made such as careful examination of the spectra[36-37], applying more restrict criteria or novel algorithms during data searching[38-43], and developing chemical derivatization[44-45] and enrichment methods,[14, 46-47] although few of these approaches provided satisfying results with ease. Recently we have developed a novel strategy for improved citrullination analysis by utilizing a novel biotin thiol tag which can specifically react with citrulline residues.[48] This method not only induces a larger mass addition of citrullination compared to the non-modified counterparts for more confident identification, but also enables effective enrichment capability.

Cerebrospinal fluid (CSF) is a clear, colorless body fluid found within the tissue that surrounds the brain and spinal cord of all vertebrates. Since the composition of CSF is directly influenced by biochemical processes in the brain, it represents an appealing source for biomarker discovery in neurodegenerative diseases.[49-50] In this study, we investigate expression level changes of protein citrullination during the development of AD using human CSF samples. We expect to identify some citrullinated proteins that exhibit changes in relative abundance or sites of location in response to the progression of AD, and therefore could be employed as novel biomarkers for the early diagnosis of AD and to facilitate the molecular understanding of AD pathogenesis.

Methods

Chemicals and materials. MS grade Lys C/trypsin mixture and dithiothreitol (DTT) were purchased from Promega (Madison, WI). Urea, tris base, ACS grade acetone, ACS grade methanol (MeOH), ACS grade acetonitrile (ACN), ACS grade dichloromethane (DCM), Optima UPLC grade ACN, Optima UPLC grade water, and Optima LC/MS grade formic acid (FA) were purchased from Fisher Scientific (Pittsburgh, PA). Iodoacetamide (IAA), trifluoroacetic acid (TFA), 2,3-butanedione, streptavidin agarose, *N,N*-diisopropylethylamine, biotin-NHS ester, cysteamine, ammonium formate (NH_4HCO_2) were purchased from Sigma-Aldrich (St. Louis, MO). Sep-Pak C18 cartridges and Bridged Ethylene Hybrid C18 particles were purchased from Waters (Milford, MA). Quantitative colorimetric peptide assay was purchased from Thermo Scientific (San Jose, CA). Strong cation exchange (SCX) SpinTips containing PolySULFOETHYL A beads were purchased from PolyLC (Columbia, MD). Tris(2-carboxyethyl)phosphine (TCEP) were purchased from MilliporeSigma (Burlington, MA). 1× phosphate-buffered saline (PBS) buffer was purchased from Corning cellgro (Corning, NY).

Synthesis of biotin thiol tag. The synthesis of biotin thiol tag was performed accordingly to the previous publication.[48] Briefly, *N,N*-diisopropylethylamine (0.88 mM) was added to a solution of biotin-NHS ester (0.29 mM) and cysteamine (0.44 mM) in DCM (5 mL) and stirred at 40 °C for 24 h. The crude product was purified using a CombiFlash system with a gradient of DCM from 0 to 20% in MeOH. Fractions containing pure product (as detected by UV) were collected (68% yield). TCEP 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.

CSF sample preparation and digestion. Human CSF samples (N=4 for either healthy or dementia group) were dried in vacuo and resuspended using 8 M urea. Samples were reduced with

10 mM DTT for 30 min at room temperature and alkylated with 50 mM IAA for another 30 min in dark before quenched with DTT. Samples were then diluted with 50 mM tris buffer to a urea concentration <1 M. On-pellet digestion was performed with LysC/trypsin mixture in a 50:1 ratio (protein:enzyme, w/w) at 37 ℃ overnight. The digestion was quenched with 1% TFA and samples were desalted with Sep-Pak C18 cartridges. Concentrations of peptide mixture were measured by peptide assay following the manufacturers' protocols. Four hundred microgram of each sample was dried in vacuo.

Derivatization of citrullinated peptides using biotin thiol tag. Each sample and 300 μg of biotin thiol tag were dissolved with 40 μL 12.5% TFA. A solution of 2,3-butanedione was prepared by mixing 1 μL of 2,3-butanedione with 114 μL 12.5% TFA. Ten microliters of 2,3-butanedione solution was subsequently added to initiate the derivatization reaction and the mixture was vortexed in dark at 37 °C for 6 h. SCX clean-up was then performed to remove the excess tag and samples were dried in vacuo.

Enrichment of derivatized citrullinated peptides. The enrichment process was performed as previously described with slight modifications.[51] Briefly, 75 μL streptavidin agarose was washed with 1 mL 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 while the other three release processes were conducted at 95 °C for 5 min with shaking. The eluents were combined and dried in vacuo.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Samples were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer coupled to a Dionex UltiMate 3000 UPLC system (Thermo Scientific, San Jose, CA). Each sample was dissolved in 8% ACN, 0.1% FA in water before loaded onto a 75 μm inner diameter homemade microcapillary column which was packed with 15 cm of Bridged Ethylene Hybrid C18 particles (1.7 μm, 130 Å) 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 8% to 37% 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 with an AGC target of 2×10^5 and maximum injection time of 100 ms. 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-p.p.m. tolerance. Each sample was acquired in technical triplicates.

Data analysis. Raw files were searched against the UniProt *Hopo sapiens* (February 2020) reviewed database using MaxQuant (version 1.5.2.8) with trypsin/P selected as the enzyme and three missed cleavages allowed. Carbamidomethylation of cysteine residues (+57.02146 Da) was chosen as the fixed modification and variable modifications included oxidation of methionine residues (+15.99492 Da), acetylation at protein N-termini (+42.01056 Da) and biotin tag-labeled citrullination of arginine (+354.10718 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. 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 site localization probability threshold was set to 0.75. “LFQ quantification” feature was enabled and all other parameters were set as default. Two-sample Student’s *t* test with a two-tailed distribution for binary comparison was conducted using Perseus (version 1.6.0.7).[52] Bioinformatics analyses including hierarchical clustering, volcano plots, and box plots were achieved using R packages. Gene ontology (GO) analysis was accomplished using DAVID bioinformatics resources[53] with a FDR cutoff of 0.05.

Results and Discussion

Benefiting from the efficacy of our citrullination analysis pipeline, we are able to perform an in-depth citrullinome profiling in human CSF samples from healthy individuals and AD patients (N=4 for each group). In total, we identify 135 citrullination sites from 68 proteins with high confidence (**Table S1**), which represents the first report of the presence of protein citrullination in human CSF samples. We then mapped the distribution of the number of citrullination sites per citrullinated proteins identified and found that a higher percentage of the proteins carry multiple citrullination sites (**Figure 1**) compared to our previous observation in mouse tissues or human cell lines[48], which may suggest that protein conformations and functions are more likely to be affected or regulated through citrullination in human CSF. To better discern the cellular localization and biological functions of these citrullinated proteins, we performed GO enrichment analysis and the results are shown in **Table S2**. As expected, these proteins are mainly localized in the extracellular space since CSF is biofluid outside of the cells. Citrullinated proteins in CSF are actively involved in various important biological processes such as regulation of protein binding, activation of complement system and other immune response.

We performed hierarchical clustering of all identified citrullinated peptides to explore their profile changes in AD patients and age-matched healthy individuals (**Figure 2**). This heatmap illustrates column-wise clustering of biological replicates in either of these two groups, suggesting larger intergroup differences than intragroup variations. The results indicate that our approach is capable of confident identification and accurate quantification of protein citrullination, and this PTM is possibly involved in the disease development. Interestingly, we find that AD samples exhibit an overall lower expression level of citrullination compared to healthy samples. Previously, most of the citrullination-related diseases are caused or intensified by excessive citrullination which often resulted in structural collapse and functional loss of modified proteins.[13, 22, 54] Inflammation pathway is another important driver during these processes which leads to higher concentration of calcium ions and hypercitrullination in turn. The only documented pathology associated with decreased citrullination is psoriasis where the lack of citrullinated keratin K1 causes excessive cornification and dry patches of skin samples due to strengthened inter- and intra-molecular interactions.[55-56] Similarly, decreased protein citrullination in AD CSF may be involved in the formation of aggregated proteins, plaques or tangles and therefore contributes to the disease pathogenesis. In accordance with this, 46 citrullination sites are exclusively identified in healthy samples while only eight are identified exclusively in AD samples (**Table S3**). In addition, 71 citrullination sites are identified in both groups and we performed a pairwise comparison of them using student's *t* test (**Table S4**). Thirty-two of them are significantly altered between these two groups and the majority are downregulated in AD patients (**Figure 3**), which is consistent with the above-mentioned observations.

Many identified citrullinated proteins have been proven to be closely associated with the pathogenesis of AD and our results suggest the possible roles of their citrullinated forms during

the development of this neurodegenerative disease. For example, we identify nine citrullination sites on apolipoprotein E, which is a cholesterol carrier that supports lipid transport and injury repair in the brain and is a well-known protein related to neurodegeneration in AD (**Figure 4**).[57-58] Three of these citrullination sites are only identified in healthy individuals and four of them show significantly decreased expression in AD patients. Emerging data suggest that clusterin also contributes to AD through various pathways, including amyloid-β aggregation and clearance, lipid metabolism, neuroinflammation, and neuronal cell cycle control and apoptosis,[59] while we identify six citrullination sites on this important protein for the first time. Complement C3 plays an important role in the immune system and was found to be activated in human AD brain and required for neurodegeneration development.[60] We identify seven citrullination sites on this protein, and the overall down-regulated expression in AD samples may provide additional evidence of its activation during AD progression. Collectively, our results reveal the possible involvement of protein citrullination in the pathogenesis of AD, and raise the possibility of using citrullinated proteins as novel biomarkers for diagnosis and prognosis of this devastating disease.

Conclusions

In this study, we explore the association of protein citrullination with the progression of AD and reveal the presence and alteration of citrullinated proteins in human CSF samples for the first time. Our results demonstrate an overall decreased expression of this PTM in AD patients, which may help further explain the disease mechanisms. We also provide a list of candidate citrullination biomarkers that may inspire future investigations for diagnosis and treatment of AD. An ongoing study which incorporates a large sample cohort including 16 individuals in each group will help generate more conclusive results.

Acknowledgements

This study was supported by grant funding from the NIH (R21 AG060242). 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.

References

(1) Blennow, K.; de Leon, M. J.; Zetterberg, H. *The Lancet* **2006,** *368*, 387-403.

(2) Roychaudhuri, R.; Yang, M.; Hoshi, M. M.; Teplow, D. B. *The Journal of biological chemistry* **2009,** *284*, 4749-53.

(3) Ballatore, C.; Lee, V. M.; Trojanowski, J. Q. *Nat. Rev. Neurosci.* **2007,** *8*, 663-72.

(4) Ramesh, M.; Gopinath, P.; Govindaraju, T. *Chembiochem* **2020,** *21*, 1052-1079.

(5) Wesseling, H.; Mair, W.; Kumar, M.; Schlaffner, C. N.; Tang, S.; Beerepoot, P.; Fatou, B.; Guise, A. J.; Cheng, L.; Takeda, S.; Muntel, J.; Rotunno, M. S.; Dujardin, S.; Davies, P.; Kosik, K. S.; Miller, B. L.; Berretta, S.; Hedreen, J. C.; Grinberg, L. T.; Seeley, W. W.; Hyman, B. T.; Steen, H.; Steen, J. A. *Cell* **2020,** *183*, 1699-1713 e13.

(6) Kelley, A. R.; Bach, S. B. H.; Perry, G. *Biochim Biophys Acta Mol Basis Dis* **2019,** *1865*, 2040-2047.

(7) Fuhrmann, J.; Clancy, K. W.; Thompson, P. R. *Chem. Rev.* **2015,** *115*, 5413-61.

(8) Tilvawala, R.; Thompson, P. R. *Curr Opin Struct Biol* **2019,** *59*, 205-215.

(9) Witalison, E.; Thompson, P.; Hofseth, L. *Current Drug Targets* **2015,** *16*, 700-710.

(10) Fuhrmann, J.; Thompson, P. R. *ACS chemical biology* **2016,** *11*, 654-68.

(11) Vitorino, R.; Guedes, S.; Vitorino, C.; Ferreira, R.; Amado, F.; Van Eyk, J. E. *Journal of proteome research* **2020**.

(12) Mondal, S.; Thompson, P. R. *Acc Chem Res* **2019,** *52*, 818-832.

(13) Gyorgy, B.; Toth, E.; Tarcsa, E.; Falus, A.; Buzas, E. I. *The international journal of biochemistry & cell biology* **2006,** *38*, 1662-77.

(14) Tutturen, A. E.; Fleckenstein, B.; de Souza, G. A. *Journal of proteome research* **2014,** *13*, 2867-73.

(15) Schellekens, G. A.; de Jong, B. A.; van den Hoogen, F. H.; van de Putte, L. B.; van Venrooij, W. J. *The Journal of clinical investigation* **1998,** *101*, 273-81.

(16) Raijmakers, R.; van Beers, J. J.; El-Azzouny, M.; Visser, N. F.; Bozic, B.; Pruijn, G. J.; Heck, A. J. *Arthritis Res Ther* **2012,** *14*, R114.

(17) Pruijn, G. J. *Front Immunol* **2015,** *6*, 192.

(18) Fert-Bober, J.; Darrah, E.; Andrade, F. *Immunol Rev* **2020,** *294*, 133-147.

(19) Elkon, K. B. *Sci Transl Med* **2013,** *5*, 209fs39.

(20) Moscarello, M. A.; Mastronardi, F. G.; Wood, D. D. *Neurochem Res* **2007,** *32*, 251-6.

(21) Yang, L.; Tan, D.; Piao, H. *Neurochem Res* **2016,** *41*, 1845-56.

(22) Gs Chirivi, R. *J. Clin. Cell. Immunol.* **2013,** *04*.

(23) Bradford, C. M.; Ramos, I.; Cross, A. K.; Haddock, G.; McQuaid, S.; Nicholas, A. P.; Woodroofe, M. N. *J Neuroimmunol* **2014,** *273*, 85-95.

(24) Yuzhalin, A. E.; Gordon-Weeks, A. N.; Tognoli, M. L.; Jones, K.; Markelc, B.; Konietzny, R.; Fischer, R.; Muth, A.; O'Neill, E.; Thompson, P. R.; Venables, P. J.; Kessler, B. M.; Lim, S. Y.; Muschel, R. J. *Nat. Commun.* **2018,** *9*, 4783.

(25) Wang, L.; Song, G.; Zhang, X.; Feng, T.; Pan, J.; Chen, W.; Yang, M.; Bai, X.; Pang, Y.; Yu, J.; Han, J.; Han, B. *Cancer Res* **2017,** *77*, 5755-5768.

(26) Thalin, C.; Lundstrom, S.; Seignez, C.; Daleskog, M.; Lundstrom, A.; Henriksson, P.; Helleday, T.; Phillipson, M.; Wallen, H.; Demers, M. *PloS one* **2018,** *13*, e0191231.

(27) Olsen, I.; Singhrao, S. K.; Potempa, J. *J Oral Microbiol* **2018,** *10*, 1487742.

(28) Ishigami, A.; Ohsawa, T.; Hiratsuka, M.; Taguchi, H.; Kobayashi, S.; Saito, Y.; Murayama, S.; Asaga, H.; Toda, T.; Kimura, N.; Maruyama, N. *Journal of neuroscience research* **2005,** *80*, 120-8.

(29) Ishigami, A.; Masutomi, H.; Handa, S.; Nakamura, M.; Nakaya, S.; Uchida, Y.; Saito, Y.; Murayama, S.; Jang, B.; Jeon, Y. C.; Choi, E. K.; Kim, Y. S.; Kasahara, Y.; Maruyama, N.; Toda, T. *Journal of neuroscience research* **2015,** *93*, 1664-74.

(30) Nicholas, A. P. *Neuroscience letters* **2013,** *545*, 107-11.

(31) Acharya, N. K.; Nagele, E. P.; Han, M.; Coretti, N. J.; DeMarshall, C.; Kosciuk, M. C.; Boulos, P. A.; Nagele, R. G. *J Autoimmun* **2012,** *38*, 369-80.

(32) Hensen, S. M.; Pruijn, G. J. *Mol Cell Proteomics* **2014,** *13*, 388-96.

(33) Clancy, K. W.; Weerapana, E.; Thompson, P. R. *Curr Opin Chem Biol* **2016,** *30*, 1-6.

(34) Moelants, E. A.; Van Damme, J.; Proost, P. *PloS one* **2011,** *6*, e28976.

(35) Verheul, M. K.; van Veelen, P. A.; van Delft, M. A. M.; de Ru, A.; Janssen, G. M. C.; Rispens, T.; Toes, R. E. M.; Trouw, L. A. *Autoimmun Rev* **2018,** *17*, 136-141.

(36) Lee, C. Y.; Wang, D.; Wilhelm, M.; Zolg, D. P.; Schmidt, T.; Schnatbaum, K.; Reimer, U.; Ponten, F.; Uhlen, M.; Hahne, H.; Kuster, B. *Mol Cell Proteomics* **2018,** *17*, 1378-1391.

(37) Jin, Z.; Fu, Z.; Yang, J.; Troncosco, J.; Everett, A. D.; Van Eyk, J. E. *Proteomics* **2013,** *13*, 2682-91.

(38) Fert-Bober, J.; Venkatraman, V.; Hunter, C. L.; Liu, R.; Crowgey, E. L.; Pandey, R.; Holewinski, R. J.; Stotland, A.; Berman, B. P.; Van Eyk, J. E. *Journal of proteome research* **2019**.

(39) Wang, X.; Swensen, A. C.; Zhang, T.; Piehowski, P. D.; Gaffrey, M. J.; Monroe, M. E.; Zhu, Y.; Dong, H.; Qian, W. J. *Journal of proteome research* **2020,** *19*, 1863-1872.

(40) Villacres, C.; Spicer, V.; Krokhin, O. V. *Journal of proteome research* **2021**.

(41) Maurais, A. J.; Salinger, A. J.; Tobin, M.; Shaffer, S. A.; Weerapana, E.; Thompson, P. R. *Biochemistry* **2021**.

(42) Huh, S.; Hwang, D.; Kim, M. S. *Anal. Chem.* **2020,** *92*, 12975-12986.

(43) De Ceuleneer, M.; Van Steendam, K.; Dhaenens, M.; Elewaut, D.; Deforce, D. *Journal of proteome research* **2012,** *11*, 5245-51.

(44) Stensland, M.; Holm, A.; Kiehne, A.; Fleckenstein, B. *Rapid Commun Mass Spectrom* **2009,** *23*, 2754-62.

(45) De Ceuleneer, M.; De Wit, V.; Van Steendam, K.; Van Nieuwerburgh, F.; Tilleman, K.; Deforce, D. *Rapid Commun Mass Spectrom* **2011,** *25*, 1536-42.

(46) Tutturen, A. E.; Holm, A.; Fleckenstein, B. *Anal Bioanal Chem* **2013,** *405*, 9321-31.

(47) Lewallen, D. M.; Bicker, K. L.; Subramanian, V.; Clancy, K. W.; Slade, D. J.; Martell, J.; Dreyton, C. J.; Sokolove, J.; Weerapana, E.; Thompson, P. R. *ACS chemical biology* **2015,** *10*, 2520-8.

(48) Li, L.; Shi, Y.; Li, Z.; Wang, B.; Shi, X.; Ye, H.; Delafield, D.; Lv, L.; Ye, Z.; Chen, Z.; Ma, F. *Research Square* **2021**.

(49) Zetterberg, H. *Curr Opin Psychiatry* **2015,** *28*, 402-9.

(50) Rosen, C.; Hansson, O.; Blennow, K.; Zetterberg, H. *Mol. Neurodegener.* **2013,** *8*, 20.

(51) Schiapparelli, L. M.; McClatchy, D. B.; Liu, H. H.; Sharma, P.; Yates, J. R., 3rd; Cline, H. T. *Journal of proteome research* **2014,** *13*, 3966-78.

(52) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. *Nature methods* **2016,** *13*, 731-40.

(53) Huang da, W.; Sherman, B. T.; Lempicki, R. A. *Nat Protoc* **2009,** *4*, 44-57.

(54) Yuzhalin, A. E. *Cancer Res* **2019,** *79*, 1274-1284.

(55) Ishida-Yamamoto, A.; Senshu, T.; Takahashi, H.; Akiyama, K.; Nomura, K.; Iizuka, H. *J Invest Dermatol* **2000,** *114*, 701-5.

(56) Mechin, M. C.; Sebbag, M.; Arnaud, J.; Nachat, R.; Foulquier, C.; Adoue, V.; Coudane, F.; Duplan, H.; Schmitt, A. M.; Chavanas, S.; Guerrin, M.; Serre, G.; Simon, M. *Int. J. Cosmet. Sci.* **2007,** *29*, 147-68.

(57) Liu, C. C.; Liu, C. C.; Kanekiyo, T.; Xu, H.; Bu, G. *Nat Rev Neurol* **2013,** *9*, 106-18.

(58) Yamazaki, Y.; Zhao, N.; Caulfield, T. R.; Liu, C.-C.; Bu, G. *Nature Reviews Neurology* **2019,** *15*, 501-518.

(59) Yu, J. T.; Tan, L. *Mol Neurobiol* **2012,** *45*, 314-26.

(60) Wu, T.; Dejanovic, B.; Gandham, V. D.; Gogineni, A.; Edmonds, R.; Schauer, S.; Srinivasan, K.; Huntley, M. A.; Wang, Y.; Wang, T. M.; Hedehus, M.; Barck, K. H.; Stark, M.; Ngu, H.; Foreman, O.; Meilandt, W. J.; Elstrott, J.; Chang, M. C.; Hansen, D. V.; Carano, R. A. D.; Sheng, M.; Hanson, J. E. *Cell Rep* **2019,** *28*, 2111-2123 e6.

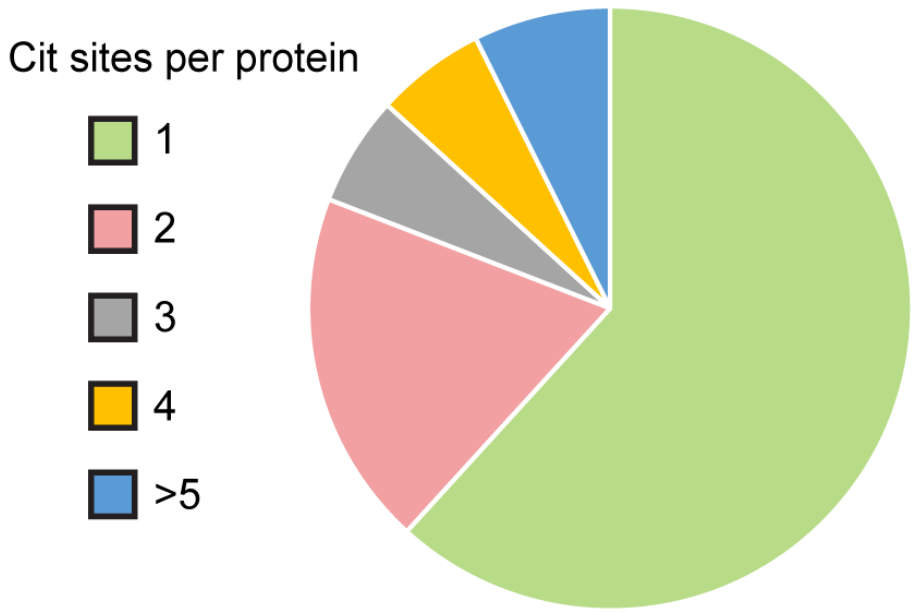

Figure 1. Distribution of the number of citrullination sites per citrullinated proteins identified.

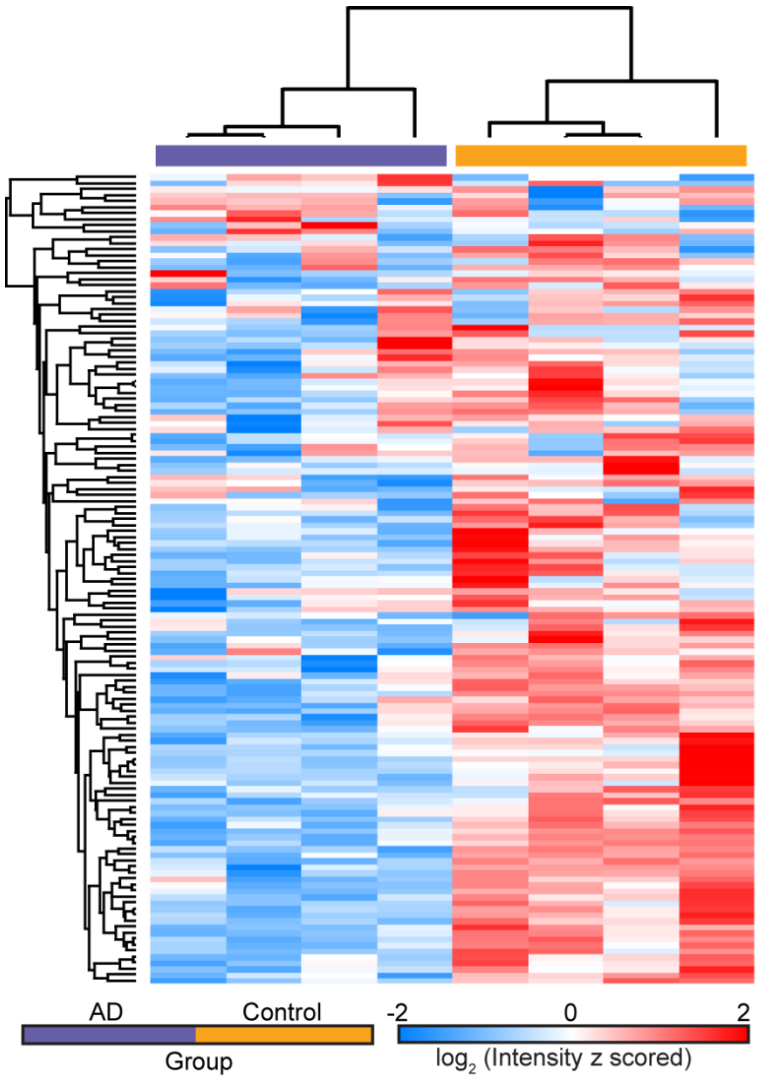

Figure 2. Hierarchical clustering of LFQ ion intensities showing the differential expression of citrullinated peptides in human CSF from healthy individuals and age-matched AD patients.

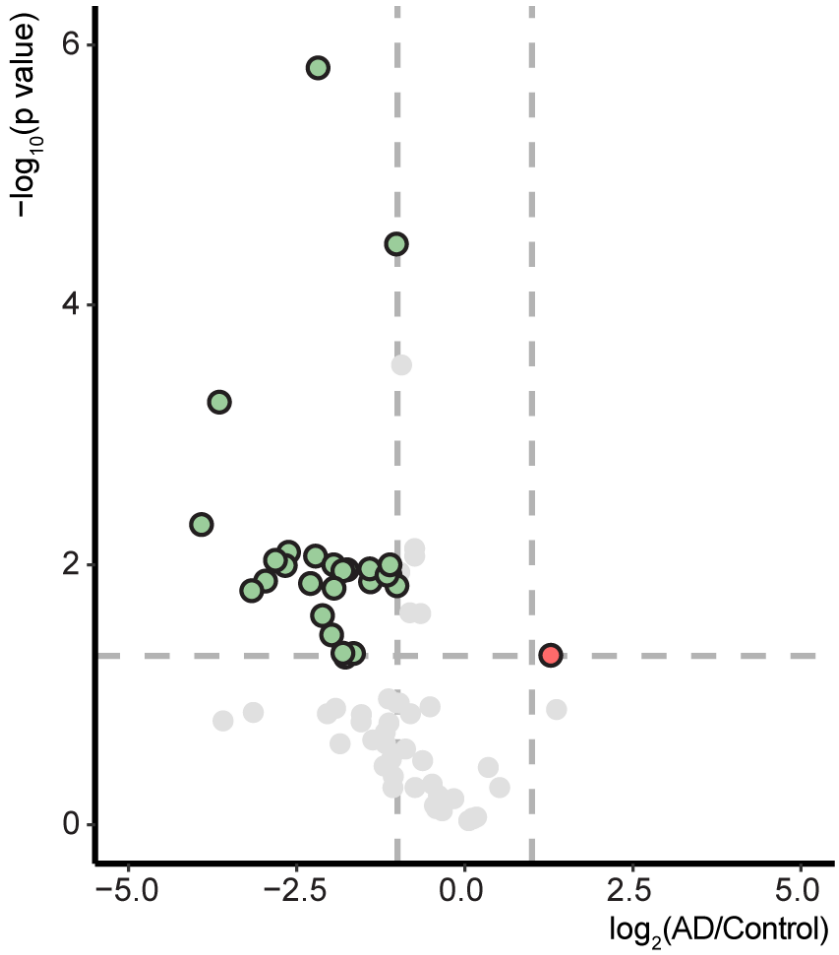

Figure 3. Volcano plot showing pairwise comparison of citrullinated peptide expression levels between AD and healthy samples. Points above horizontal dash lines represent significantly altered proteins (two-sided *t* test, *p* value <0.05). Significantly down-regulated proteins are shown in green (protein fold change <0.5) and up-regulated ones are shown in red (protein fold change > 2).

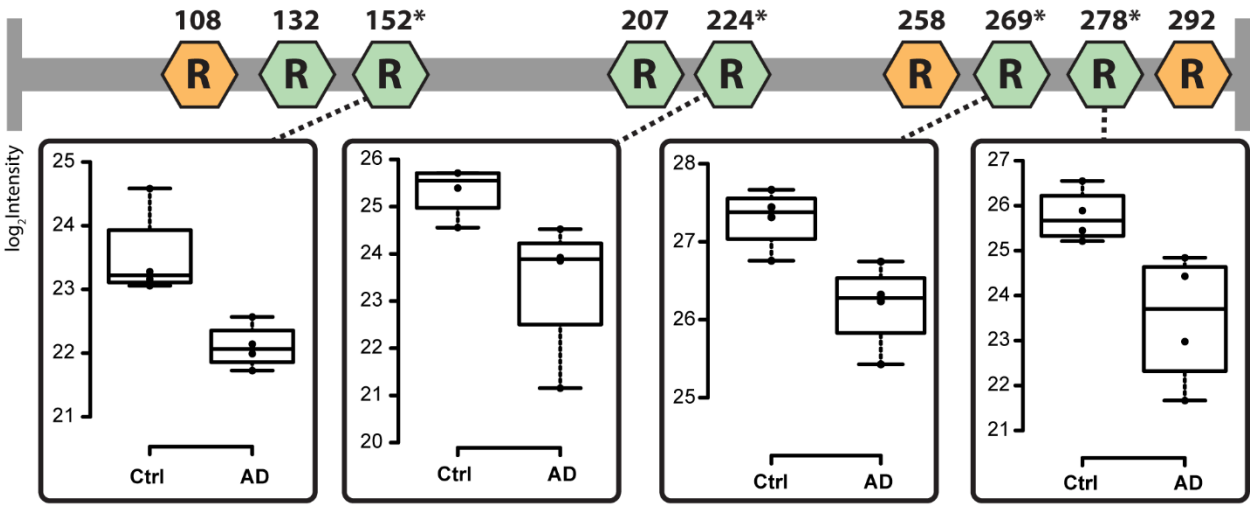

Figure 4. Citrullination profile on protein apolipoprotein E. Yellow sites are only identified in healthy (Ctrl) samples while green sites are identified in both AD and healthy samples. Sites with an asterisk indicate their expression levels are significantly different between AD and healthy and the connecting box plots show the expression levels of these sites. Black dots within boxes indicate replicate data points. All boxplots indicate median (center line), 25th and 75th percentiles (bounds of box), and minimum and maximum (whiskers).

Supplemental Information

Table S1. List of all identified citrullination sites

Accession	Description	Position
O00533	Neural cell adhesion molecule L1-like protein	713
O15240	Neurosecretory protein VGF	371
O15240	Neurosecretory protein VGF	176
O15484	Calpain-5	438
O43583	Density-regulated protein	18
P00450	Ceruloplasmin	62
P00450	Ceruloplasmin	258
P00734	Prothrombin	263
P00734	Prothrombin	486
P00751	Complement factor B	193
P00751	Complement factor B	593
P01008	Antithrombin-III	438
P01009	Alpha-1-antitrypsin	247
P01023	Alpha-2-macroglobulin	1014
P01023	Alpha-2-macroglobulin	598
P01023	Alpha-2-macroglobulin	853
P01023	Alpha-2-macroglobulin	586
P01024	Complement C3	1512
P01024	Complement C3	573
P01024	Complement C3	439
P01024	Complement C3	425
P01024	Complement C3	1582
P01024	Complement C3	1042
P01024	Complement C3	1376
P01701	Immunoglobulin lambda variable 1-51	74
P01834	Immunoglobulin kappa constant	35
P01859	Immunoglobulin heavy constant gamma 2	16
P01860	Immunoglobulin heavy constant gamma 3	113
P01876	Immunoglobulin heavy constant alpha 1	273
P02647	Apolipoprotein A-I	51
P02647	Apolipoprotein A-I	212
P02647	Apolipoprotein A-I	107
P02647	Apolipoprotein A-I	147
P02649	Apolipoprotein E	258
P02649	Apolipoprotein E	108
P02649	Apolipoprotein E	292
P02649	Apolipoprotein E	224
P02649	Apolipoprotein E	152
P02649	Apolipoprotein E	269
P02649	Apolipoprotein E	132

P02649	Apolipoprotein E	207
P02649	Apolipoprotein E	278
P02671	Fibrinogen alpha chain	573
P02686	Myelin basic protein	199
P02751	Fibronectin	2447
P02751	Fibronectin	953
P02753	Retinol-binding protein 4	195
P02753	Retinol-binding protein 4	173
P02763	Alpha-1-acid glycoprotein 1	167
P02766	Transthyretin	123
P02774	Vitamin D-binding protein	218
P02774	Vitamin D-binding protein	149
P02787	Serotransferrin	346
P02787	Serotransferrin	69
P02787	Serotransferrin	696
P02787	Serotransferrin	273
P02787	Serotransferrin	132
P02787	Serotransferrin	600
P02787	Serotransferrin	239
P02787	Serotransferrin	251
P02787	Serotransferrin	651
P02787	Serotransferrin	587
P02790	Hemopexin	151
P02790	Hemopexin	208
P02790	Hemopexin	371
P02794	Ferritin heavy chain	64
P04217	Alpha-1B-glycoprotein	90
P04217	Alpha-1B-glycoprotein	402
P05060	Secretogranin-1	202
P05060	Secretogranin-1	101
P05060	Secretogranin-1	162
P05060	Secretogranin-1	642
P05090	Apolipoprotein D	60
P05090	Apolipoprotein D	82
P05155	Plasma protease C1 inhibitor	366
P06396	Gelsolin	43
P06396	Gelsolin	188
P06396	Gelsolin	397
P06396	Gelsolin	48
P06727	Apolipoprotein A-IV	180
P06727	Apolipoprotein A-IV	304
P06727	Apolipoprotein A-IV	154

P07998	Ribonuclease pancreatic	113
P08603	Complement factor H	175
P08603	Complement factor H	166
P0C0L5	Complement C4-B	337
P0C0L5	Complement C4-B	138
P0C0L5	Complement C4-B	954
P0C0L5	Complement C4-B	185
P0C0L5	Complement C4-B	155
P0C0L5	Complement C4-B	1349
P0C0L5	Complement C4-B	729
P10645	Chromogranin-A	455
P10645	Chromogranin-A	150
P10909	Clusterin	286
P10909	Clusterin	167
P10909	Clusterin	194
P10909	Clusterin	336
P10909	Clusterin	182
P10909	Clusterin	306
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	259
P22680	Cytochrome P450 7A1	112
P23142	Fibulin-1	497
P23142	Fibulin-1	607
P23142	Fibulin-1	405
P31323	cAMP-dependent protein kinase type II-beta regulatory subunit	383
P41222	Prostaglandin-H2 D-isomerase	85
P51693	Amyloid-like protein 1	124
P51693	Amyloid-like protein 1	388
P55010	Eukaryotic translation initiation factor 5	73
P98171	Rho GTPase-activating protein 4	371
Q12805	EGF-containing fibulin-like extracellular matrix protein 1	148
Q14957	Glutamate receptor ionotropic, NMDA 2C	194
Q15113	Procollagen C-endopeptidase enhancer 1	162
Q16270	Insulin-like growth factor-binding protein 7	221
Q5TFQ8	Signal-regulatory protein beta-1 isoform 3	177
Q6NVU6	Inactive Ufm1-specific protease 1	12
Q7Z3B1	Neuronal growth regulator 1	127
Q8NDZ4	Divergent protein kinase domain 2A	316
Q8WZ42	Titin	18362
Q92876	Kallikrein-6	111
Q92876	Kallikrein-6	91
Q92876	Kallikrein-6	232
Q92878	DNA repair protein RAD50	441

Q96KN2	Beta-Ala-His dipeptidase	87
Q9BPX3	Condensin complex subunit 3	246
Q9BQ39	ATP-dependent RNA helicase DDX50	182
Q9BXJ9	N-alpha-acetyltransferase 15, NatA auxiliary subunit	684
Q9P202	Whirlin	497
Q9P202	Whirlin	498
Q9P2T0	Testicular haploid expressed gene protein	120
Q9UBP4	Dickkopf-related protein 3	224
Q9UHG2	ProSAAS	104
Q9UL18	Protein argonaute-1	275
Q9Y283	Inversin	60

Table S2. List of enriched Gene Ontology terms

Category	Term	Count	*p* value[a]
Cellular Component	blood microparticle	26	2.60E-33
	extracellular space	39	4.30E-24
	extracellular region	40	1.20E-22
	extracellular exosome	46	6.00E-20
	platelet alpha granule lumen	8	3.50E-08
	fibrinogen complex	3	1.20E-02
	extracellular matrix	7	1.60E-02
	chylomicron	3	2.20E-02
	endocytic vesicle lumen	3	2.60E-02
	immunoglobulin complex, circulating	3	3.30E-02
	very-low-density lipoprotein particle	3	3.30E-02
	high-density lipoprotein particle	3	3.70E-02
Molecular Function	protease binding	7	3.40E-04
	serine-type endopeptidase activity	9	3.90E-04
	endopeptidase inhibitor activity	5	8.10E-04
	serine-type endopeptidase inhibitor activity	6	1.30E-03
	immunoglobulin receptor binding	4	4.10E-03
	phosphatidylcholine-sterol O-acyltransferase activator activity	3	5.70E-03
	small molecule binding	3	6.80E-03
	heparin binding	6	6.80E-03
	cholesterol binding	4	8.90E-03
	antigen binding	5	9.20E-03
	peptidase activator activity	3	9.20E-03
	protein binding	46	1.80E-02
	cholesterol transporter activity	3	1.80E-02
	lipid transporter activity	3	3.00E-02

Biological Process	platelet degranulation	11	4.40E-09
	complement activation	9	4.80E-07
	complement activation, classical pathway	9	9.10E-07
	negative regulation of endopeptidase activity	8	6.70E-05
	proteolysis	12	4.10E-04
	phagocytosis, engulfment	5	1.00E-03
	retina homeostasis	5	1.50E-03
	innate immune response	10	2.90E-03
	reverse cholesterol transport	4	2.90E-03
	retinoid metabolic process	5	5.20E-03
	cellular protein metabolic process	6	5.20E-03
	positive regulation of B cell activation	4	6.80E-03
	phagocytosis, recognition	4	7.80E-03
	regulation of complement activation	4	9.00E-03
	high-density lipoprotein particle assembly	3	1.50E-02
	neuron projection regeneration	3	1.50E-02
	acute-phase response	4	1.60E-02
	positive regulation of cholesterol esterification	3	1.60E-02
	lipoprotein biosynthetic process	3	1.60E-02
	response to dietary excess	3	1.90E-02
	cellular iron ion homeostasis	4	1.90E-02
	receptor-mediated endocytosis	6	2.10E-02
	complement activation, alternative pathway	3	2.80E-02
	positive regulation of peptidase activity	3	2.80E-02
	B cell receptor signaling pathway	4	2.90E-02
	phospholipid efflux	3	3.00E-02
	high-density lipoprotein particle remodeling	3	3.30E-02
	response to wounding	4	4.00E-02
	cholesterol homeostasis	4	4.00E-02
	blood coagulation, intrinsic pathway	3	4.30E-02

[a]*p* values are adjusted by Benjamini–Hochberg correction

Table S3. List of citrullination sites only identified in AD or healthy samples

Presence	Accession	Description	Position
Only in AD	P06396	Gelsolin	43
	P0C0L5	Complement C4-B	337
	P10909	Clusterin	286
	P02787	Serotransferrin	346
	P0C0L5	Complement C4-B	138
	P51693	Amyloid-like protein 1	124
	P55010	Eukaryotic translation initiation factor 5	73
	Q12805	EGF-containing fibulin-like extracellular matrix protein 1	148
Only in healthy	O15240	Neurosecretory protein VGF	371
	P00450	Ceruloplasmin	62
	P00734	Prothrombin	263
	P01023	Alpha-2-macroglobulin	1014
	P01023	Alpha-2-macroglobulin	598
	P01024	Complement C3	1512
	P01834	Immunoglobulin kappa constant	35
	P01860	Immunoglobulin heavy constant gamma 3	113
	P01876	Immunoglobulin heavy constant alpha 1	273
	P02649	Apolipoprotein E	258
	P02649	Apolipoprotein E	108
	P02649	Apolipoprotein E	292
	P02787	Serotransferrin	69
	P02790	Hemopexin	151
	P02790	Hemopexin	208
	P05060	Secretogranin-1	202
	P10645	Chromogranin-A	455
	P10909	Clusterin	167
	P23142	Fibulin-1	497
	P23142	Fibulin-1	607
	Q5TFQ8	Signal-regulatory protein beta-1 isoform 3	177
	Q14957	Glutamate receptor ionotropic, NMDA 2C	194
	Q15113	Procollagen C-endopeptidase enhancer 1	162
	Q7Z3B1	Neuronal growth regulator 1	127
	Q8NDZ4	Divergent protein kinase domain 2A	316
	Q92876	Kallikrein-6	111
	Q92876	Kallikrein-6	91
	Q96KN2	Beta-Ala-His dipeptidase	87
	Q9UHG2	ProSAAS	104
	O00533	Neural cell adhesion molecule L1-like protein	713
	O15240	Neurosecretory protein VGF	176

O43583	Density-regulated protein	18
P00734	Prothrombin	486
P01024	Complement C3	573
P02647	Apolipoprotein A-I	51
P02794	Ferritin heavy chain	64
P05060	Secretogranin-1	101
P05060	Secretogranin-1	162
P10645	Chromogranin-A	150
P31323	cAMP-dependent protein kinase type II-beta regulatory subunit	383
P41222	Prostaglandin-H2 D-isomerase	85
P98171	Rho GTPase-activating protein 4	371
Q9BXJ9	N-alpha-acetyltransferase 15, NatA auxiliary subunit	684
Q9P202	Whirlin	497
Q9P202	Whirlin	498
Q9P2T0	Testicular haploid expressed gene protein	120

Table S4. List of citrullination sites identified in both AD and healthy samples

Accession	Description	Site	*p* value	log_2(AD/Ctrl)
P00450	Ceruloplasmin	258	0.01333	-2.96
P00751	Complement factor B	593	0.26149	-0.88
P01009	Alpha-1-antitrypsin	247	0.42428	-1.06
P01023	Alpha-2-macroglobulin	853	0.22275	-1.37
P01023	Alpha-2-macroglobulin	586	0.51708	-0.74
P01024	Complement C3	439	0.00750	-0.75
P01024	Complement C3	1582	0.00857	-2.22
P01024	Complement C3	425	0.13990	-0.81
P01024	Complement C3	1042	0.23744	-1.18
P01024	Complement C3	1376	0.89065	0.11
P01701	Immunoglobulin lambda variable 1-51	74	0.01352	-1.40
P01859	Immunoglobulin heavy constant gamma 2	16	0.23820	-1.86
P02647	Apolipoprotein A-I	212	0.00857	-0.75
P02647	Apolipoprotein A-I	107	0.02338	-0.82
P02647	Apolipoprotein A-I	147	0.14215	-1.54
P02649	Apolipoprotein E	152	0.01075	-1.41
P02649	Apolipoprotein E	269	0.01200	-1.11
P02649	Apolipoprotein E	278	0.01392	-2.30
P02649	Apolipoprotein E	224	0.03462	-1.98
P02649	Apolipoprotein E	132	0.35331	-1.20
P02649	Apolipoprotein E	207	0.90017	0.12
P02671	Fibrinogen alpha chain	573	0.87272	0.17
P02686	Myelin basic protein	199	0.11589	-0.97
P02751	Fibronectin	953	0.01091	-1.75
P02751	Fibronectin	2447	0.19506	-1.18
P02753	Retinol-binding protein 4	173	0.01450	-1.01
P02763	Alpha-1-acid glycoprotein 1	167	0.00000	-2.18
P02766	Transthyretin	123	0.12714	-1.92
P02774	Vitamin D-binding protein	218	0.04813	-1.66
P02787	Serotransferrin	132	0.00003	-1.02
P02787	Serotransferrin	587	0.00056	-3.65
P02787	Serotransferrin	696	0.01000	-1.95
P02787	Serotransferrin	239	0.04788	-1.81
P02787	Serotransferrin	600	0.12985	1.36
P02787	Serotransferrin	651	0.28346	-1.07
P02787	Serotransferrin	273	0.63041	-0.16
P02787	Serotransferrin	251	0.75145	-0.43
P02790	Hemopexin	371	0.01582	-3.17
P04217	Alpha-1B-glycoprotein	90	0.10753	-1.13
P04217	Alpha-1B-glycoprotein	402	0.90412	0.10

P05060	Secretogranin-1	642	0.02459	-2.12
P05090	Apolipoprotein D	60	0.00029	-0.94
P05090	Apolipoprotein D	82	0.14000	-2.04
P05155	Plasma protease C1 inhibitor	366	0.00490	-3.92
P06396	Gelsolin	188	0.02371	-0.66
P06396	Gelsolin	48	0.16169	-1.54
P06727	Apolipoprotein A-IV	180	0.00800	-2.62
P06727	Apolipoprotein A-IV	304	0.01111	-1.81
P06727	Apolipoprotein A-IV	154	0.31479	-1.09
P07998	Ribonuclease pancreatic	113	0.01200	-1.16
P08603	Complement factor H	175	0.12379	-0.52
P08603	Complement factor H	166	0.16478	-1.13
P0C0L5	Complement C4-B	1349	0.04968	1.28
P0C0L5	Complement C4-B	185	0.48669	-0.48
P0C0L5	Complement C4-B	729	0.51760	0.52
P0C0L5	Complement C4-B	155	0.78109	-0.34
P0C0L5	Complement C4-B	954	0.93099	0.06
P10909	Clusterin	336	0.01000	-1.11
P10909	Clusterin	182	0.01012	-2.67
P10909	Clusterin	194	0.05133	-1.78
P10909	Clusterin	306	0.13674	-3.15
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	259	0.36236	0.35
P22680	Cytochrome P450 7A1	112	0.15882	-3.60
P51693	Amyloid-like protein 1	388	0.01513	-1.94
Q16270	Insulin-like growth factor-binding protein 7	221	0.01140	-0.97
Q92876	Kallikrein-6	232	0.00923	-2.82
Q92878	DNA repair protein RAD50	441	0.51856	-1.07
Q9BPX3	Condensin complex subunit 3	246	0.59568	-0.39
Q9UBP4	Dickkopf-related protein 3	224	0.11566	-1.01
Q9UL18	Protein argonaute-1	275	0.71225	-0.45
Q9Y283	Inversin	60	0.32133	-0.63

Chapter 7

Mass spectrometry imaging of N-glycans from formalin-fixed paraffin-embedded tissue sections using a novel sub-atmospheric pressure ionization source

Adapted from: Shi, Y.; **Li, Z.**; Felder, M. A.; Yu, Q.; Shi, X.; Peng, Y.; Cao, Q.; Wang, B.; Puglielli, L.; Patankar, M. S.; Li, L., Mass Spectrometry Imaging of N-Glycans from Formalin-Fixed Paraffin-Embedded Tissue Sections Using a Novel Sub-atmospheric Pressure Ionization Source. *Analytical Chemistry* **2019,** *91* (20), 12942-12947. Li, Z. contributed to experimental design, data collection and helped prepare the manuscript. DOI https://doi.org/10.1021/acs.analchem.9b02995

Abstract

N-linked glycosylation, featuring various glycoforms, is one of the most common and complex protein post-translational modifications (PTMs) controlling protein structures and biological functions. It has been revealed that abnormal changes of protein N-glycosylation patterns are associated with many diseases. Hence, unraveling the disease-related alteration of glycosylation, especially the glycoforms, is crucial and beneficial to improve our understanding about the pathogenic mechanisms of various diseases. In past decades, given the capability of in-situ mapping of biomolecules and their region-specific localizations, matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) has been widely applied to the discovery of potential biomarkers for many diseases. In this study, we coupled a novel sub-atmospheric pressure (SubAP)/MALDI source with a Q Exactive HF hybrid quadrupole-orbitrap mass spectrometer for in-situ imaging of N-linked glycans from formalin-fixed paraffin-embedded (FFPE) tissue sections. The utility of this new platform for N-glycan imaging analysis was demonstrated with a variety of FFPE tissue sections. A total of 55 N-glycans were successfully characterized and visualized from a FFPE mouse brain section. Furthermore, 29 N-glycans with different spatial distribution patterns could be identified from a FFPE mouse ovarian cancer tissue section. High-mannose N-glycans exhibited elevated expression levels in the tumor region, indicating the potential association of this type of N-glycans with tumor progression.

Introduction

N-linked glycosylation is a type of common post-translational modification (PTM), which plays an important role in various physiological processes. Prior studies revealed that aberrant glycosylation was associated with many diseases including Alzheimer's disease,[1] cancers[2,3] and cardiovascular diseases.[4] Although electrospray ionization mass spectrometry (ESI-MS)-based glycomic[5,6] and glycoproteomic analyses[7-10] have been applied to investigate N-glycan expression changes between healthy and disease specimens, the homogenization of the tissue samples conducted in the vast majority of previous studies inevitably result in the loss of spatial information regarding N-glycan localization. The region-specific N-glycan distribution patterns, especially in the disease area, could be critical in discovering N-glycan biomarkers and revealing disease mechanisms.

Complementary to ESI-MS technique, matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) technique has been introduced for in-situ biomolecular analysis due to its capability of retaining and directly visualizing biomolecules in tissue sections.[11-13] Depending on the source pressure, MALDI-MS can be divided into three types: vacuum MALDI-MS, atmospheric MALDI-MS (AP/MALDI-MS) and sub-atmospheric pressure MALDI-MS (SubAP/MALDI-MS). Vacuum MALDI-MS is the one that commonly used, and generally has higher sensitivity than that of AP/MALDI-MS, enabled by the low pressure (less than 1.33 Pascal (Pa)) in the source.[14] However, to gain the high vacuum in the source, it typically requires a

dedicated MALDI MS system to perform these in-situ profiling and imaging experiments. Different from the traditional vacuum MALDI source, the AP/MALDI source and SubAP/MALDI source are independent of mass spectrometer and can be switched back and forth with an ESI source on an MS instrument platform, providing greater instrumental versatility and flexibility to allow both ESI-based analysis and MALDI-based analysis on a single mass spectrometer. Furthermore, the SubAP/MALDI source is operated at a subatmospheric pressure (133 Pa-1330 Pa) and offers better sensitivity than AP/MALDI source, making it a more attractive alternative to the traditional vacuum MALDI source.[15]

In past decades, MALDI-MS has been successfully applied to map the spatial distributions of various biomolecules on tissue sections, such as lipids,[16,17] small metabolites[18,19] and neuropeptides.[20] Recently, the scope of the vacuum MALDI-MS, has been successfully extended to the in-situ N-glycan imaging analysis,[21-23] while the application of SubAP/MALDI-MS to N-glycan imaging has rarely been reported, even though its sensitivity has been proven to be sufficient for many biological applications.[15] Considering the advantages of SubAP/MALDI-MS, it could be beneficial to systematically evaluate the potential of SubAP/MALDI-MS, as a complementary tool of vacuum MALDI-MS, in the imaging of N-glycans from tissue sections.

In this study, a new platform was developed by coupling the novel SubAP/MALDI source to a Q Exactive HF (QE HF) hybrid quadrupole-Orbitrap mass spectrometer. After optimization with N-glycan mixtures, this platform was successfully employed to characterize and visualize N-

glycans released from formalin-fixed paraffin-embedded (FFPE) tissue sections including mouse brain and murine ovarian cancer tissue sections. Our results, for the first time, demonstrated the applicability of the SubAP/MALDI-MS platform for MS imaging (MSI) of N-glycans with high spatial resolution and mass accuracy.

Methods

Chemicals and reagents. Methanol (MeOH), ethanol (EtOH), acetonitrile (ACN), formic acid (FA), sodium chloride (NaCl), trifluoroacetic acid (TFA), citric acid and ammonium bicarbonate were purchased from Fisher Scientific (Pittsburgh, PA). Xylene and microcon-30kDa centrifugal filter unit were purchased from EMD Millipore Sigma (Burlington, MA). Dithiothreitol and PNGase F were purchased from Promega (Madison, WI). Distilled water mentioned in this work was Milli-Q water from a Millipore filtration system (Bedford, MA). α-cyano-4-hydroxycinnamic acid (CHCA), Tris (2-carboxyethyl) phosphine (TCEP), triethylammonium bicarbonate buffer (TEAB) and bovine thyroglobulin (BTG) were purchased from Sigma-Aldrich (St. Louis, MO). 2,5-dihydroxybenzoic acid (DHB) was purchased from Acros Organics (Morris Plains, NJ). All reagents were used without additional purification. Microscope glass slides were purchased from VWR international, LLC (Radnor, PA). Indium tin oxide (ITO)-coated glass slides were purchased from Delta technologies (Loveland, CO).

Preparation of FFPE tissue sections. Mouse studies were approved by the Animal Care and Use Committee of the University of Wisconsin-Madison. For the FFPE mouse brain section, mouse

was anesthetized with carbon dioxide. The brain was collected immediately after transcardial PBS perfusion, fixed overnight in 10% neutral buffered formalin, and paraffin embedded using standard techniques. 6 µm coronal tissue sections were prepared using a microtome (HistoCore MULTICUT, Leica Biosystems).

For the development of ovarian tumor in a mouse model, syngeneic murine ovarian cancer cells (Mouse Ovarian Surface Epithelial Cells, known as MOSEC[24]) were suspended in phosphate buffered saline and injected into the peritoneal cavity of 8 to 10-week-old female C57BL/6 mice, $10x10^6$ cells/animal. The MOSEC were a gift from Dr. Kathy Roby, Kansas University Medical Center, Kansas City, KS. The mice were housed under controlled conditions in the vivarium and monitored over a period of 2-3 months for the development of ascites as determined by behavioral changes/indications of discomfort, weight gain and presence of a distended belly, and body condition evaluation. Accumulation of ascites typically occurred 10-12 weeks post implantation of the cancer cells. Animals showing significant accumulation of ascites and with worsening body condition were sacrificed and tumors from the peritoneal cavity were obtained during necropsy. The tumors were immediately stored in formalin and embedded in paraffin the following day. 6 µm tissue section was cut from the paraffin block, adhered to the ITO-coated slide, and used for mass spectrometry-based N-glycan imaging.

Preparation of N-glycans released from bovine thyroglobulin. N-glycans were released from glycoprotein standard-bovine thyroglobulin (BTG) with slightly modified filter-aided N-Glycan

separation (FANGS) strategy.[25] Glycoproteins were dissolved in water to have a concentration of 2 μg μL^{-1} and mixed with 5 μL of 0.5 M TCEP. Heat-denaturation was performed by switching sample tubes between 100°C water bath and room temperature for four cycles of 15 seconds each. The mixture was then loaded onto a 30 kDa molecular weight cutoff (MWCO) filter and buffer exchanged with 100 μL 0.5 M triethylammonium bicarbonate buffer (TEAB) by centrifugation at 14000g for three cycles (15 minutes for each) to remove contaminants with low molecular weight. 4 μL of PNGase F in 96 μL of 0.5 M TEAB buffer was added to each filter and incubated at 37°C overnight. N-glycans were separated with deglycosylated proteins and eluted by centrifugation at 14000g for 10 minutes. The filter was then washed with 100 μL 0.5 M TEAB buffer to ensure complete elution of N-glycans. Two fractions were combined and dried *in vacuo*. The released N-glycans were analyzed using the SubAP/MALDI QE HF mass spectrometer.

Pretreatment of FFPE tissue section for SubAP/MALDI N-glycan imaging. SubAP/MALDI imaging of N-glycans from FFPE tissue section was performed using workflow illustrated in **Figure S1**. 6μm FFPE tissue sections were placed on a hot plate and heated at 65°C for 20 min. After cooling down, FFPE tissue sections were consecutively washed twice with xylene, ethanol, 95% ethanol and 70% ethanol to remove paraffin. Following that, the antigen retrieval process was performed by boiling tissue sections in 20 mM freshly prepared citric acid buffer for 1 hour. Enzyme and matrix application were performed by a robotic TM sprayer system (HTX Technologies, Carrobo, NC). For enzyme deposition, 20 μl PNGase F dissolved in 330 μl 50 mM

ammonium bicarbonate solution was sprayed at a flowrate of 0.02 ml min^{-1}, and in total 16 passes were performed. The nozzle temperature was set to 35 °C with a moving velocity of 800 mm min^{-1}. Then, FFPE tissue sections were incubated in a humidity chamber at 37°C for 12 hours. CHCA dissolved in ACN: H_2O: TFA (v: v: v, 50: 50: 0.1) solution at a concentration of 7 mg ml^{-1} was used as the matrix for N-glycan imaging. 24 passes of matrix spraying were performed at a flow rate of 0.05 ml min^{-1}, and 30 seconds drying time was set between each pass. The nozzle temperature was set to 80°C with a moving velocity of 800 mm min^{-1}. After matrix spraying, tissue slides were dried in a vacuum chamber for 30 min and stored at -20°C until use.

Mass spectrometry. The QE HF mass spectrometer is coupled with a novel SubAP/MALDI (ng) ion source equipped with a 355 nm Nd:YAG laser (**Figure 1**). The laser spot size is 10 μm and the maximum output frequency reaches 10 kHz. Spiral plate motion was used for profiling experiments, while constant speed raster mode was employed for the imaging of tissue sections. A microscopic slide in the sample holder is approximately 2 mm away from the MS inlet capillary. Enabled by two novel design features, the SubAP/MALDI source has better sensitivity than the regular AP/MALDI source. Firstly, it operates at pressures from 1 Torr to 10 Torr instead of regular atmospheric pressure. Secondly, an ion funnel interface is designed to improve the ion collection and transfer into the mass spectrometer. The source is controlled by the Target software (MassTech Inc., Columbia, MD), and the ImageQuest software (Thermo Fisher Scientific, Waltham, MA) is applied to construct the images of target analytes using the XY coordinates of laser spots along

with the acquired MS data file. For N-glycan imaging experiments, laser energy of 20 % and repetition rate of 200 Hz were used. Ion funnel parameters were set as below: Voltage (V) values for V1 through V7 were set as 0, 0, 5, 10, 170, 200 and 300, respectively; RF amplitude was set as 250 V; QE-HF mass spectrometer was operated at positive full MS scan mode with the mass range of m/z 1100-3000. The capillary temperature and S-lens RF level were set to 200 °C and 50, respectively. Automatic gain control (AGC) of 5e6 was used to ensure that the maximum ion injection time of 300 ms could be reached at each pixel. Resolution was sample-dependent and will be discussed in Results and discussion section.

Histology staining. H&E staining was performed as described previously.[26] Tissue sections were deparaffinized through a xylene and graded ethanol series, then rinsed in hematoxylin, in 1% acid alcohol, and then in eosin. After rinsing in running water to remove excessive staining, slides were dehydrated in graded ethanol and xylene.

Data processing. MS spectra were processed by Xcalibur (Thermo Scientific, Bremen, Germany), and observed N-glycans were annotated by using the GlycoWorkbench[27] software with less than 10 ppm mass tolerance. Signal intensities of N-glycans were normalized to total ion chromatogram (TIC) and ImageQuest (Thermo Scientific, Bremen, Germany) was used to construct N-glycan images with the mass tolerance window of 5 ppm. N-glycan compositions were tentatively identified by searching against UniCarbKB database.

Results and discussion

Optimization of operating pressure. SubAP/MALDI source was operated at sub-atmospheric pressure normally ranging from 133 Pa to1330 Pa. Given that the ion source pressure is critical to ion transport efficiency, it is necessary to examine the impact of source pressure on the N-glycan signal intensities and select an optimal pressure for N-glycan imaging analysis. When coupled to QE HF mass spectrometer, a minimum source pressure of 386 Pa was required to maintain the stability of the entire platform. Hence, the pressure optimization was initiated from 400 Pa using N-glycans released from bovine thyroglobulin (A representative MS spectrum is shown in **Figure S2**). N-glycan mixture was mixed with CHCA, and further spotted onto sample plate for analysis. Ion abundance of 4 dominant N-glycans were normalized to TIC chromatogram, and used for pressure optimization. As shown in **Figure 2a-c**, TIC-normalized ion abundance and ion abundance of N-glycans significantly dropped following the increase of operating pressure, while no apparent change of TIC intensities was observed. This result suggested that relatively lower pressure could improve the sensitivity of SubAP/MALDI measurements probably due to enhanced transportation of N-glycans. Therefore, to balance the stability and sensitivity of the platform, operating pressure of 400 Pa was selected for N-glycan analysis.

Matrix optimization. DHB and CHCA are two common matrices used for MALDI imaging of small biomolecules. In general, high laser energy is required to ionize analytes co-crystalized with DHB matrix, while lower laser energy is needed for the ionization of CHCA-coated analytes. The performance of these two matrices in N-glycan imaging on the SubAP/MALDI-MS platform was

studied. When CHCA was used, increasing laser energy improved TIC intensities (**Figure 2f**), though the optimal N-glycan ion abundance, either absolute or normalized, was obtained at the laser energy of 6.8 micro Joules per square centimeter ($\mu J/cm^2$) **(Figure 2d-e**). For DHB, higher TIC-normalized ion abundance of N-glycans were observed due to the much lower TIC intensities in comparison to CHCA (**Figure 2g**). Ion abundance and TIC augmented following the increase of laser energy, while higher energy was required to get similar results as that of CHCA (**Figure 2h-i**). Thus, to get longer laser lifespan and higher glycan signal intensities, matrix of CHCA and laser energy of $6.8\mu J/cm^2$ were used for SubAP/MALDI-based N-glycan imaging.

High-resolution imaging of N-glycans from FFPE mouse brain sections. Recent N-glycoproteomic studies revealed that N-glycoproteins with heterogeneous N-glycans were highly expressed in mouse brain.[28-31] Therefore, to test the applicability of the novel SubAP/MALDI-MS platform for high resolution MS imaging acquisition, a 6 µm-thick FFPE mouse brain coronal section was prepared and treated by PNGase F to release N-glycans for imaging analysis with a pixel size of 25µm. A total of 55 N-glycans including high-mannose N-glycans, fucosylated N-glycans and sialylated glycans were detected and annotated by the accurate mass matching (**Figure S3**, **Table S1**). The new SubAP/MALDI-MS platform detected more N-glycans in comparison to previous N-glycan imaging studies of either frozen or FFPE mouse brain section using vacuum MALDI-MS platforms (**Figure 3a**, **Table S2**).[32,33] Also, as shown in **Figure 3b**, the new SubAP/MALDI source is capable of detecting most N-glycans identified by vacuum MALDI-MS.

Furthermore, by matching with the optical image of the mouse brain section (**Figure 3c**), N-glycans with different distribution patterns on the brain tissue section could be revealed and representative N-glycans were shown in **Figures 3d-3h**. These results manifested that the new instrument platform had adequate sensitivity for the detection and visualization of N-glycans from FFPE tissue sections, and therefore could be employed as an alternative to vacuum MALDI-MS platform for *in-situ* N-glycan imaging analysis.

Mapping of N-glycan distribution on FFPE mouse ovarian cancer tissue section. N-glycosylation is a protein PTM that plays an important role in many devastating diseases. Accumulating evidence have indicated that aberrant alteration of N-glycosylation can occur during the onset and progression of ovarian cancer.[22,34-36] In particular, high-mannose N-glycans, which are synthesized in the endoplasmic reticulum (ER) and are crucial for proper protein folding, have been shown to be dominantly expressed in the cancer area.[22] Herein, a mouse model with ovarian cancer was established and a 6 μm FFPE ovarian tissue section containing tumor region was prepared. Following that, the new SubAP/MALDI-MS platform was employed to map the distribution patterns of N-glycans in the tissue section. Considering that the mouse ovarian tissue section was bigger, but less heterogeneous than the mouse brain section, a pixel size of 50 μm was selected without significant compromise on the image quality. In total, 29 N-glycans, including complex N-glycans and high-mannose N-glycans, were successfully detected and annotated by the accurate mass matching (**Figure 4**, **Table S3**). More homogenous distributions of complex N-

glycans on tissue section were observed and shown in **Figures 5a-h**. In contrast, high-mannose N-glycans, from $Hex_4HexNAc_2$ to $Hex_{10}HexNAc_2$, were highly expressed in the cancer region though no apparent accumulation was observed for $Hex_3HexNAc_2$ as the simplest N-glycan (**Figures 5i-p**). This observation is consistent with the previous finding that increasing N-glycan branches could lead to structural and functional changes of N-glycoproteins, contributing to the progression of cancers.[3,37] Overall, our MSI results revealed that alteration of high-mannose N-glycans was involved in mouse ovarian cancer development, which was consistent with the findings of previous N-glycomic profiling studies,[22,38,39] and further demonstrating the capability of the new platform for *in-situ* N-glycan imaging analysis.

Conclusions

A novel MS-based imaging platform was developed by integrating a new SubAP/MALDI source with a Q Exactive HF orbitrap mass spectrometer for the in-situ N-glycan imaging analysis with high resolution and high mass accuracy for the first time. Several parameters, such as the operation pressure of the SubAP/MALDI source, the type of matrix and laser energy, were optimized with a mixture of N-glycans released from a glycoprotein. After optimization, the new platform was successfully applied to map distribution patterns of N-glycans from FFPE mouse brain and ovarian cancer tissue sections. More N-glycans can be identified on brain section in comparison to that of commonly used vacuum MALDI-MS platforms, and high mannose N-glycans show accumulation in the tumor region. The new platform was demonstrated to have improved sensitivity for N-glycan

profiling analysis from biological tissue sections. For future directions, by simply switching the PNGase F to trypsin, it is entirely possible to extend the application of the new platform from N-glycan imaging to peptide imaging, and even in-situ glycopeptide identification enabled by the powerful Q Exactive HF orbitrap mass spectrometer or similar high performance MS instrument platforms.

Acknowledgments

The authors would like to thank Dr. Eugene Moskovets and Dr. Vladimir Doroshenko from MassTech, Inc. for access to the SubAP MALDI source and technical support. The research was supported in part by NIH R01 DK071801, R56MH110215, RF1AG052324, and U01CA231081. The MALDI Orbitrap instrument was purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531). LL acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

References

(1) Liang, H. C.; Russell, C.; Mitra, V.; Chung, R.; Hye, A.; Bazenet, C.; Lovestone, S.; Pike, I.; Ward, M. *J. Proteome. Res.* **2015**, *14*, 5063-5076.

(2) Lau, K. S.; Dennis, J. W. *Glycobiology* **2008**, *18*, 750-760.

(3) de Leoz, M. L.; Young, L. J.; An, H. J.; Kronewitter, S. R.; Kim, J.; Miyamoto, S.; Borowsky, A. D.; Chew, H. K.; Lebrilla, C. B. *Mol. Cell. Proteomics* **2011**, *10*, M110 002717.

(4) Yang, S.; Mishra, S.; Chen, L.; Zhou, J. Y.; Chan, D. W.; Chatterjee, S.; Zhang, H. *Anal. Chem.* **2015**, *87*, 9671-9678.

(5) Park, D.; Brune, K. A.; Mitra, A.; Marusina, A. I.; Maverakis, E.; Lebrilla, C. B. *Mol. Cell. Proteomics* **2015**, *14*, 2910-2921.

(6) Yang, S.; Hoti, N.; Yang, W.; Liu, Y.; Chen, L.; Li, S.; Zhang, H. *Clin Proteom* **2017**, *14*, 3.

(7) Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L. W.; Kent, K. C.; Li, L. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1751-1764.

(8) Sun, S.; Shah, P.; Eshghi, S. T.; Yang, W.; Trikannad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Hoti, N.; Zhang, Z.; Chan, D. W.; Zhang, H. *Nat. Biotechnol.* **2016**, *34*, 84-88.

(9) Chen, Z.; Yu, Q.; Hao, L.; Liu, F.; Johnson, J.; Tian, Z.; Kao, W. J.; Xu, W.; Li, L. *Analyst* **2018**, *143*, 2508-2519.

(10) Hong, Q.; Ruhaak, L. R.; Stroble, C.; Parker, E.; Huang, J.; Maverakis, E.; Lebrilla, C. B. *J. Proteome. Res.* **2015**, *14*, 5179-5192.

(11) Jones, E. E.; Dworski, S.; Canals, D.; Casas, J.; Fabrias, G.; Schoenling, D.; Levade, T.; Denlinger, C.; Hannun, Y. A.; Medin, J. A.; Drake, R. R. *Anal. Chem.* **2014**, *86*, 8303-8311.

(12) Lukowski, J. K.; Weaver, E. M.; Hummon, A. B. *Anal. Chem.* **2017**, *89*, 8453-8458.

(13) Chen, B.; OuYang, C.; Tian, Z.; Xu, M.; Li, L. *Anal. Chim. Acta* **2018**, *1007*, 16-25.

(14) Keller, C.; Maeda, J.; Jayaraman, D.; Chakraborty, S.; Sussman, M. R.; Harris, J. M.; Ane, J. M.; Li, L. *Front Plant Sci*. **2018**, *9*, 1238.

(15) Li, G.; Cao, Q.; Liu, Y.; DeLaney, K.; Tian, Z.; Moskovets, E.; Li, L. *Rapid Commun Mass Spectrom.* **2019**, *33*, 327-335.

(16) Shi, Y.; Johnson, J.; Wang, B.; Chen, B.; Fisher, G. L.; Urabe, G.; Shi, X.; Kent, K. C.; Guo, L. W.; Li, L. *J. Proteome. Res.* **2019**, *18*, 1669-1678.

(17) Cerruti, C. D.; Benabdellah, F.; Laprevote, O.; Touboul, D.; Brunelle, A. *Anal. Chem.* **2012**, *84*, 2164-2171.

(18) Zhang, J.; Lamotte, L.; Dodds, E. D.; Lebrilla, C. B. *Anal. Chem.* **2005**, *77*, 4429-4438.

(19) Cornett, D. S.; Frappier, S. L.; Caprioli, R. M. *Anal. Chem.* **2008**, *80*, 5648-5653.

(20) Ye, H.; Hui, L.; Kellersberger, K.; Li, L. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 134-147.

(21) Scott, D. A.; Casadonte, R.; Cardinali, B.; Spruill, L.; Mehta, A. S.; Carli, F.; Simone, N.; Kriegsmann, M.; Del Mastro, L.; Kriegsmann, J.; Drake, R. R. *Proteomics Clin Appl* **2019**, *13*, e1800014.

(22) Everest-Dass, A. V.; Briggs, M. T.; Kaur, G.; Oehler, M. K.; Hoffmann, P.; Packer, N. H. *Mol. Cell. Proteomics* **2016**, *15*, 3003-3016.

(23) Holst, S.; Heijs, B.; de Haan, N.; van Zeijl, R. J.; Briaire-de Bruijn, I. H.; van Pelt, G. W.; Mehta, A. S.; Angel, P. M.; Mesker, W. E.; Tollenaar, R. A.; Drake, R. R.; Bovee, J. V.; McDonnell, L. A.; Wuhrer, M. *Anal. Chem.* **2016**, *88*, 5904-5913.

(24) Roby, K. F.; Taylor, C. C.; Sweetwood, J. P.; Cheng, Y.; Pace, J. L.; Tawfik, O.; Persons, D. L.; Smith, P. G.; Terranova, P. F. *Carcinogenesis* **2000**, *21*, 585-591.

(25) Abdul Rahman, S.; Bergstrom, E.; Watson, C. J.; Wilson, K. M.; Ashford, D. A.; Thomas, J. R.; Ungar, D.; Thomas-Oates, J. E. *J. Proteome. Res.* **2014**, *13*, 1167-1176.

(26) Titford, M. *Biotechnic & histochemistry : official publication of the Biological Stain Commission* **2005**, *80*, 73-78.

(27) Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M. *J. Proteome. Res.* **2008**, *7*, 1650-1659.

(28) Fang, P.; Wang, X. J.; Xue, Y.; Liu, M. Q.; Zeng, W. F.; Zhang, Y.; Zhang, L.; Gao, X.; Yan, G. Q.; Yao, J.; Shen, H. L.; Yang, P. Y. *Oncotarget.* **2016**, *7*, 38796-38809.

(29) Tian, Y.; Kelly-Spratt, K. S.; Kemp, C. J.; Zhang, H. *J. Proteome. Res.* **2010**, *9*, 5837-5847.

(30) Liu, J.; Wang, F.; Mao, J.; Zhang, Z.; Liu, Z.; Huang, G.; Cheng, K.; Zou, H. *Anal. Chem.* **2015**, *87*, 2054-2057.

(31) Qiao, X.; Tao, D.; Qu, Y.; Sun, L.; Gao, L.; Zhang, X.; Liang, Z.; Zhang, L.; Zhang, Y. *Proteomics* **2011**, *11*, 4274-4278.

(32) Powers, T. W.; Jones, E. E.; Betesh, L. R.; Romano, P. R.; Gao, P.; Copland, J. A.; Mehta, A. S.; Drake, R. R. *Anal. Chem.* **2013**, *85*, 9799-9806.

(33) Toghi Eshghi, S.; Yang, S.; Wang, X.; Shah, P.; Li, X.; Zhang, H. *ACS Chem. Biol.* **2014**, *9*, 2149-2156.

(34) Kui Wong, N.; Easton, R. L.; Panico, M.; Sutton-Smith, M.; Morrison, J. C.; Lattanzio, F. A.; Morris, H. R.; Clark, G. F.; Dell, A.; Patankar, M. S. *J. Biol. Chem.* **2003**, *278*, 28619-28634.

(35) Abbott, K. L.; Nairn, A. V.; Hall, E. M.; Horton, M. B.; McDonald, J. F.; Moremen, K. W.; Dinulescu, D. M.; Pierce, M. *Proteomics* **2008**, *8*, 3210-3220.

(36) Saldova, R.; Wormald, M. R.; Dwek, R. A.; Rudd, P. M. *Dis. Markers* **2008**, *25*, 219-232.

(37) Kaprio, T.; Satomaa, T.; Heiskanen, A.; Hokke, C. H.; Deelder, A. M.; Mustonen, H.; Hagstrom, J.; Carpen, O.; Saarinen, J.; Haglund, C. *Mol. Cell. Proteomics* **2015**, *14*, 277-288.

(38) Chen, H.; Deng, Z.; Huang, C.; Wu, H.; Zhao, X.; Li, Y. *Tumor biol.* **2017**, *39*, 1010428317716249.

(39) Anugraham, M.; Jacob, F.; Nixdorf, S.; Everest-Dass, A. V.; Heinzelmann-Schwarz, V.; Packer, N. H. *Mol. Cell. Proteomics* **2014**, *13*, 2213-2232.

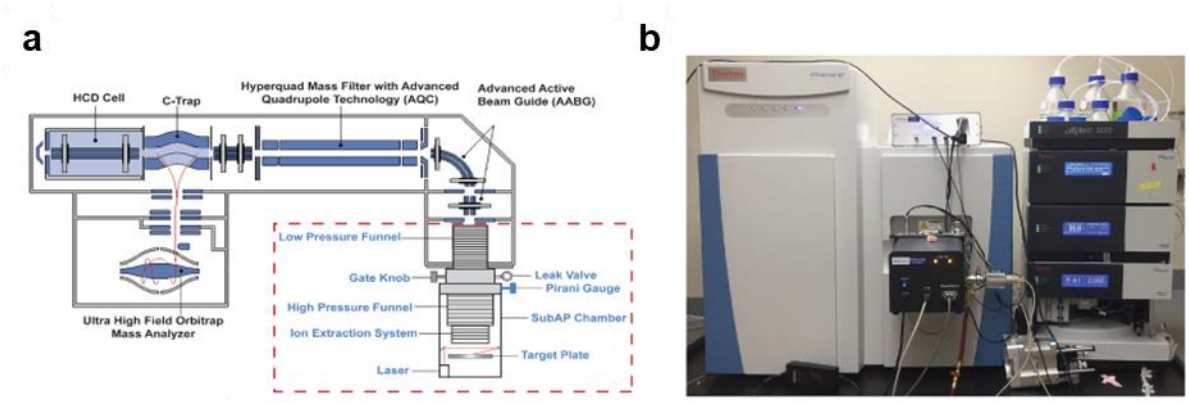

Figure 1. The new SubAP/MALDI-QEHF platform for N-glycan imaging. a, Schematic illustration of the SubAP/MALDI-QEHF platform. **b,** A snapshot showing the QEHF mass spectrometer integrated with the SubAP/MALDI source instead of an ESI source.

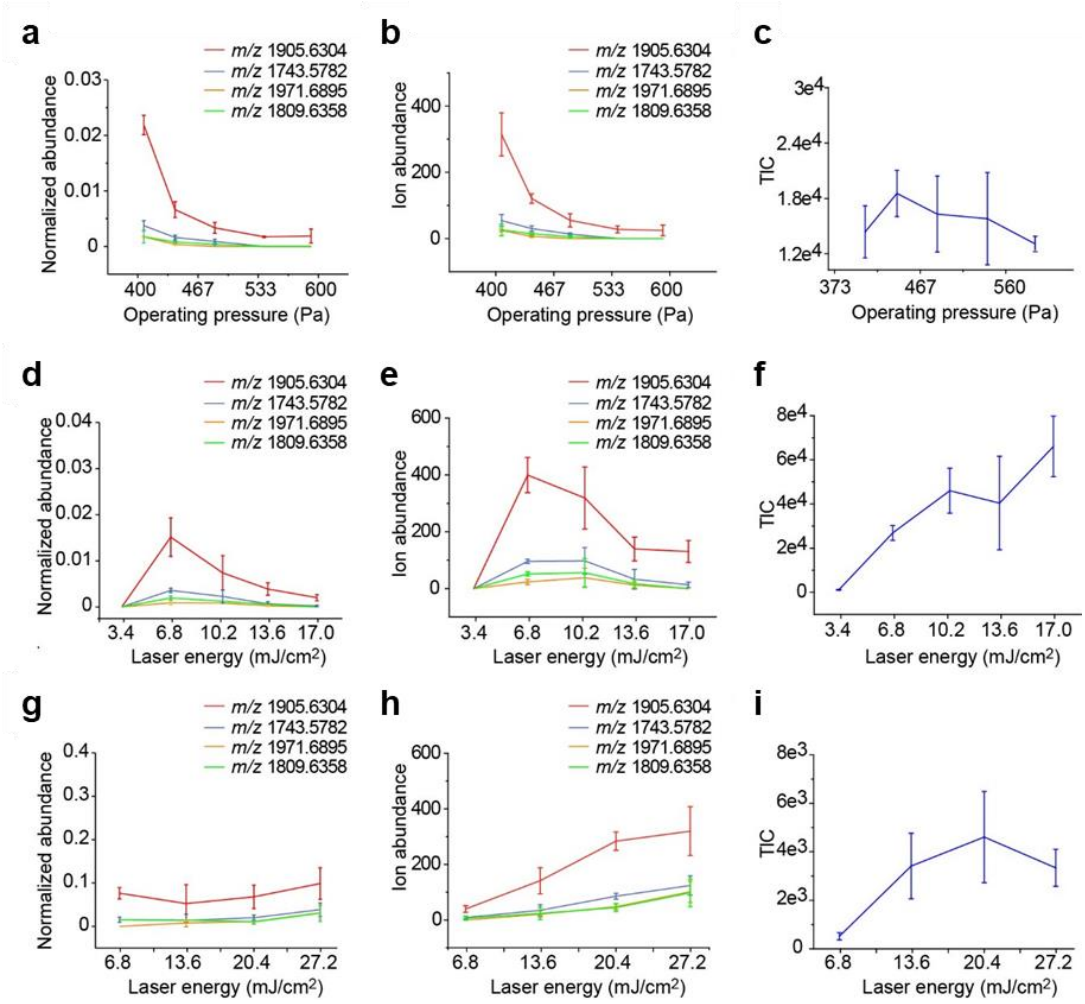

Figure 2. Parameter optimization for N-glycan analysis on SubAP/MALDI MS platform (n=3). TIC-normalized abundance (**a**), ion abundance (**b**) and TIC (**c**) of four representative N-glycans detected at different operating pressure. TIC-normalized abundance (**d**), ion abundance (**e**) and TIC (**f**) of four representative N-glycans detected at different laser energy using CHCA as matrix. TIC-normalized abundance (**g**), ion abundance (**h**) and TIC (**i**) of four representative N-glycans detected at different laser energy using DHA as matrix.

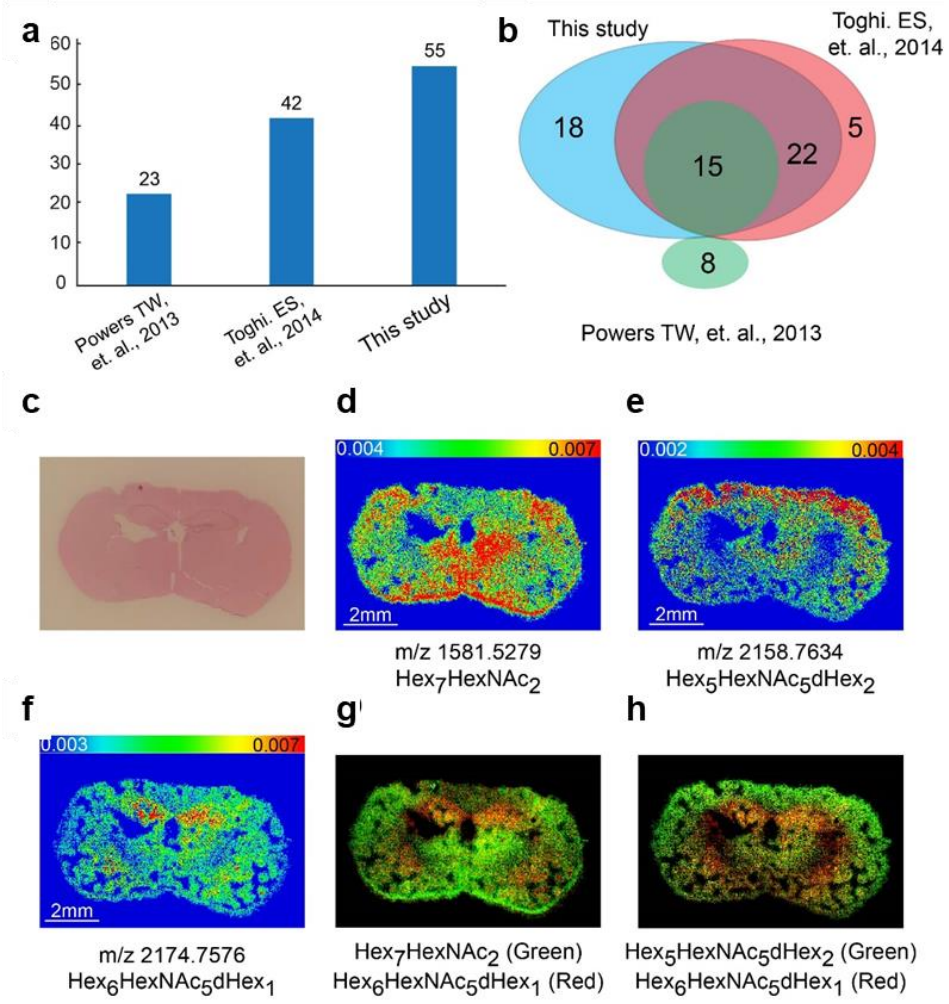

Figure 3. High-resolution imaging of N-glycans from FFPE mouse brain sections. a, More N-glycans were detected and imaged from FFPE mouse brain tissue section by using the novel SubAP/MALDI-MS platform. **b,** Venn diagram showing the overlap of N-glycans detected in this study with N-glycans reported in prior studies (Refs. 32 & 33) using vacuum MALDI-MS platform. **c,** H&E stained FFPE mouse brain section post N-glycan imaging. **d-f,** MS images of representative N-glycans detected from FFPE mouse tissue section. **g-h,** Overlap of different N-glycan images clearly revealed different spatial distribution patterns of N-glycans on mouse brain section.

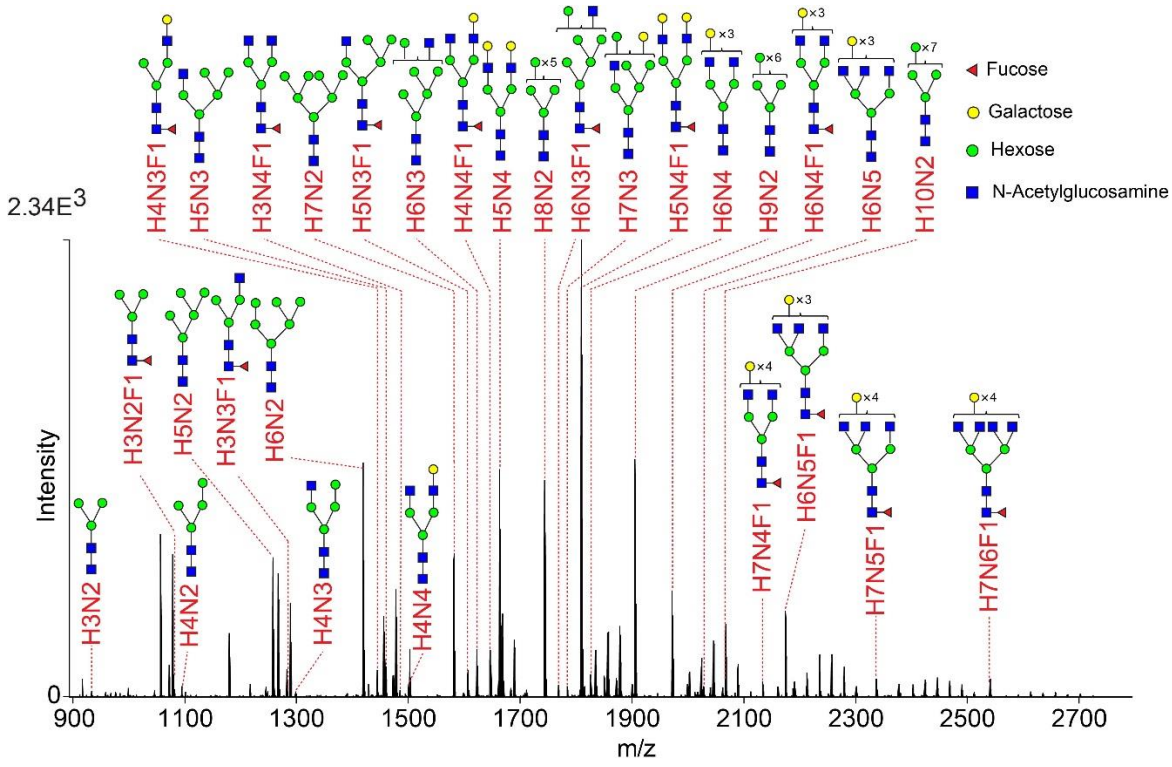

Figure 4. N-glycans detected from FFPE mouse tissue section with ovarian cancer. The annotated glycan compositions were tentatively identified by searching against UniCarbKB database. H: Hexose; N; N-Acetyl glucosamine; F: Fucose.

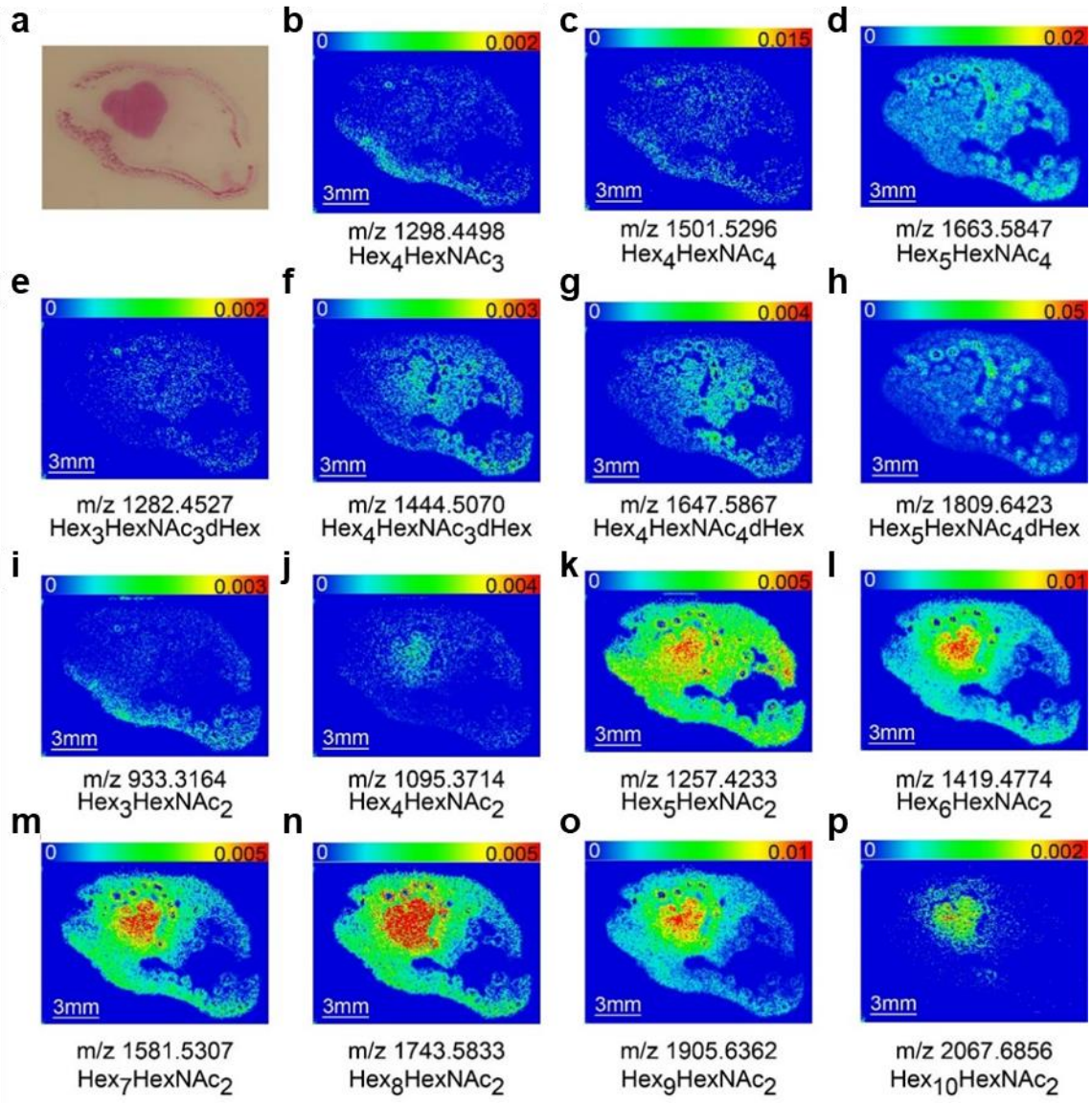

Figure 5. Images of N-glycans showing different spatial distribution patterns on FFPE mouse tissue section with ovarian cancer. a, H&E stained FFPE mouse tissue section with ovarian cancer. **b-h,** Complex N-glycans showed similar distribution in cancer area in comparison to peripheral area; **i-p,** High mannose N-glycans accumulated in cancer area except $Hex_3HexNAc_2$.

Supplemental Information

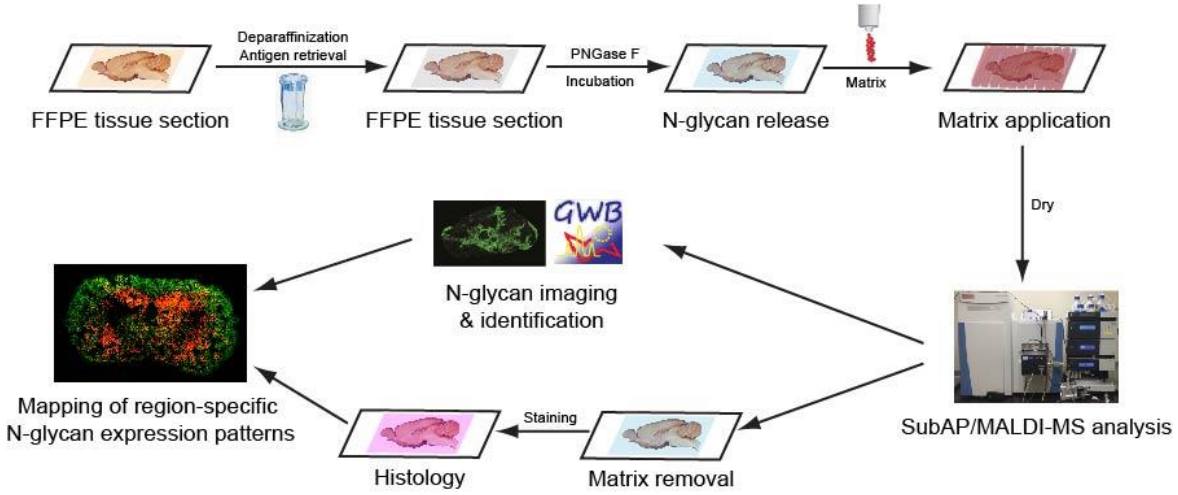

Figure S1. Typical workflow for imaging of N-glycans from FFPE tissue sections using the SubAP/MALDI-MS platform.

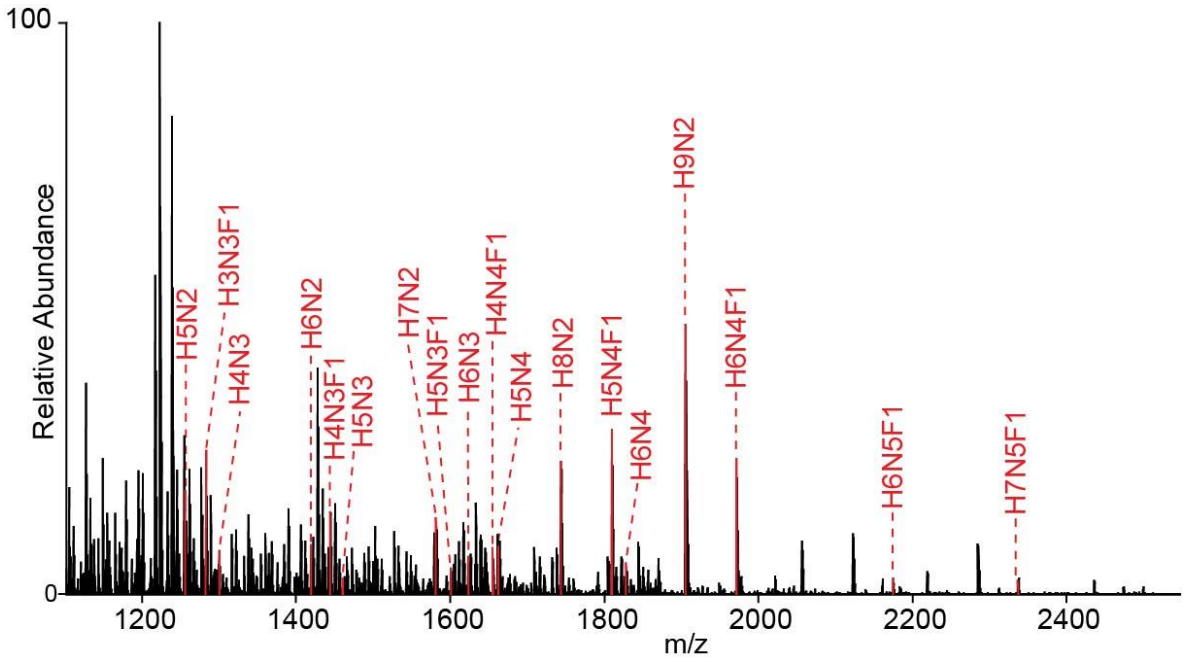

Figure S2. Representative MS spectrum of N-glycans released from bovine thyroglobulin collected with SubAP/MALDI-QEHF MS. H: Hexose; N: N-Acetylglucosamine; F: Fucose.

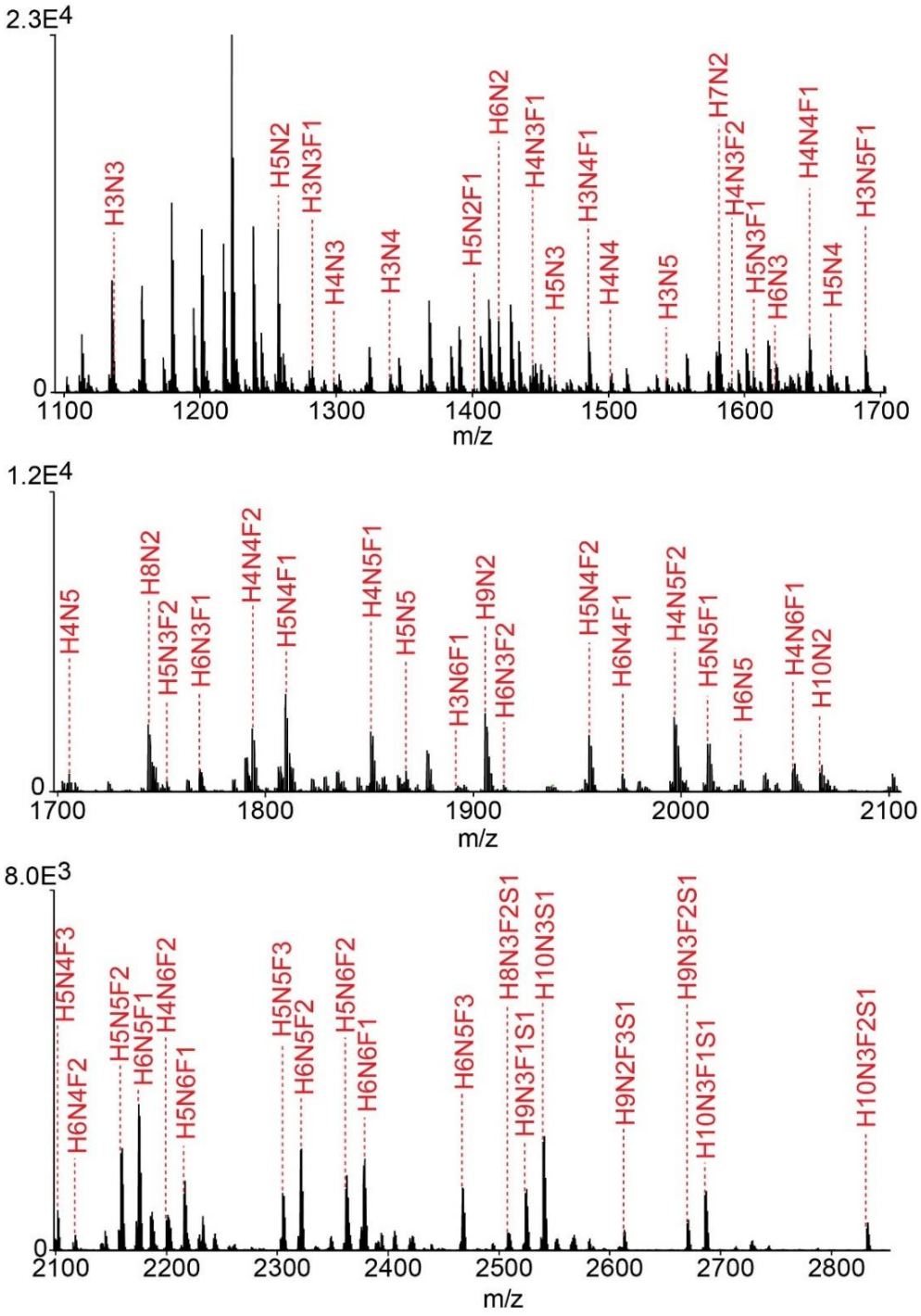

Figure S3. N-glycans detected from FFPE mouse brain section using SubAP/MALDI-MS platform. H: Hexose; N: N-Acetylglucosamine; F: Fucose; S: Sialic acid.

Table S1. N-glycans detected from FFPE mouse brain section using SubAP/MALDI-MS platform.

m/z (+Na)	Identified N-glycans	Mass deviation (ppm)	m/z (+Na)	Identified N-glycans	Mass deviation (ppm)
1136.3952	$Hex_3HexNAc_3$	1	1905.6299	$Hex_9HexNAc_2$	2
1257.4232	$Hex_5HexNAc_2$	0.5	1914.6621	$Hex_6HexNAc_3dHex_2$	4
1282.4527	$Hex_3HexNAc_3dHex_1$	1	1955.6928	$Hex_5HexNAc_4dHex_2$	2
1298.4488	$Hex_4HexNAc_3$	0.3	1971.6812	$Hex_6HexNAc_4dHex_1$	5.5
1339.4741	$Hex_3HexNAc_4$	1	1996.7157	$Hex_4HexNAc_5dHex_2$	4
1403.4740	$Hex_5HexNAc_2dHex_1$	4	2012.7104	$Hex_5HexNAc_5dHex_1$	4
1419.4784	$Hex_6HexNAc_2$	2	2028.6978	$Hex_6HexNAc_5$	7
1444.5090	$Hex_4HexNAc_3dHex_1$	1	2053.7325	$Hex_4HexNAc_6dHex_1$	6
1460.5029	$Hex_5HexNAc_3$	0.6	2067.6873	$Hex_{10}HexNAc_2$	0
1485.5353	$Hex_3HexNAc_4dHex_1$	1	2101.7387	$Hex_5HexNAc_4dHex_3$	8
1501.5285	$Hex_4HexNAc_4$	0	2117.7344	$Hex_6HexNAc_4dHex_2$	7
1542.5542	$Hex_3HexNAc_5$	0.6	2158.7634	$Hex_5HexNAc_5dHex_2$	6
1581.5279	$Hex_7HexNAc_2$	0.2	2174.7576	$Hex_6HexNAc_5dHex_1$	5
1590.5617	$Hex_4HexNAc_3dHex_2$	2	2199.7842	$Hex_4HexNAc_6dHex_2$	8
1606.5598	$Hex_5HexNAc_3dHex_1$	0	2215.7813	$Hex_5HexNAc_6dHex_1$	7
1622.5511	$Hex_6HexNAc_3$	2	2304.8148	$Hex_5HexNAc_5dHex_3$	8
1647.5874	$Hex_4HexNAc_4dHex_1$	0.6	2320.8122	$Hex_6HexNAc_5dHex_2$	7
1663.5806	$Hex_5HexNAc_4$	0.5	2361.8348	$Hex_5HexNAc_6dHex_2$	9
1688.6129	$Hex_3HexNAc_5dHex_1$	0	2377.8285	$Hex_6HexNAc_6dHex_1$	9
1704.6021	$Hex_4HexNAc_5$	3	2466.8614	$Hex_6HexNAc_5dHex_3$	10
1743.5799	$Hex_8HexNAc_2$	0.7	2507.8749 (+H)	$Hex_8HexNAc_3NeuAc_1dHex_2$	6
1752.6130	$Hex_5HexNAc_3dHex_2$	3	2523.8793 (+H)	$Hex_9HexNAc_3NeuAc_1dHex_1$	2
1768.6097	$Hex_6HexNAc_3dHex_1$	0	2539.8764 (+H)	$Hex_{10}HexNAc_3NeuAc_1$	1
1793.6421	$Hex_4HexNAc_4dHex_2$	1	2612.9016 (+H)	$Hex_9HexNAc_2NeuAc_1dHex_3$	7
1809.6370	$Hex_5HexNAc_4dHex_1$	1	2669.9299 (+H)	$Hex_9HexNAc_3NeuAc_1dHex_2$	5
1850.6631	$Hex_4HexNAc_5dHex_1$	1.5	2685.9269 (+H)	$Hex_{10}HexNAc_3NeuAc_1dHex_1$	4
1866.6469	$Hex_5HexNAc_5$	7.5	2831.9749 (+H)	$Hex_{10}HexNAc_3NeuAc_1dHex_2$	7
1891.6793	$Hex_3HexNAc_6dHex_1$	7			

Table S2. N-glycans reported in previous studies (Refs. 32 & 33) and detected by SubAP/MALDI-MS platform from FFPE mouse brain section. H: Hexose; N: N-Acetylglucosamine; F: Fucose; S: Sialic acid. √: Detected; ×: Undetected

N-glycan	Powers TW, et. al (Ref 32)	Toghi. ES, et. al (Ref 33)	This study	N-glycan	Powers TW, et. al (Ref 32)	Toghi. ES, et. al (Ref 33)	This study
H3N3	×	×	√	H6N5	×	×	√
H5N2	√	√	√	H4N6F1	×	√	√
H3N3F1	×	√	√	H10N2	×	×	√
H4N3	×	√	√	H5N4F3	√	√	√
H3N4	×	√	√	H6N4F2	√	√	√
H5N2F1	×	√	√	H5N5F2	√	√	√
H6N2	√	√	√	H6N5F1	×	√	√
H4N3F1	×	√	√	H4N6F2	×	√	√
H5N3	×	√	√	H5N6F1	√	√	√
H3N4F1	×	√	√	H6N6	×	√	×
H4N4	√	√	√	H5N5F3	×	√	√
H3N5	√	√	√	H6N5F2	×	√	√
H7N2	×	√	√	H5N6F2	×	√	√
H4N3F2	×	×	√	H6N6F1	×	×	√
H5N3F1	×	√	√	H6N5F3	×	√	√
H6N3	×	×	√	$H_5N_6F_1S_1$	√	×	×
H4N4F1	√	√	√	$H_5N_4F_1S_2$	√	×	×
H5N4	×	√	√	$H_5N_6F_2S_1$	√	×	×
H3N5F1	√	√	√	H5N6F3	×	√	×
H4N5	×	×	√	H6N6F2	×	√	×
H8N2	√	√	√	$H_7N_6S_1$	√	×	×
H5N3F2	×	×	√	$H_6N_5F_3S_1$	√	×	×
H6N3F1	×	√	√	$H_6N_5F_2S_2$	√	×	×
H4N4F2	√	√	√	$H_7N_6F_3S_1$	√	×	×
H5N4F1	√	√	√	$H_7N_6F_4S_1$	√	×	×
H4N5F1	×	√	√	H6N5F4	×	√	×
H5N5	×	×	√	H6N6F3	×	√	×
H3N6F1	×	×	√	H8N3F2S1	×	×	√
H9N2	×	√	√	H9N3F1S1	×	×	√
H6N3F2	×	×	√	H10N3S1	×	×	√
H5N4F2	√	√	√	H9N2F3S1	×	×	√

H6N4F1	×	√	√	H9N3F2S1	×	×	√
H4N5F2	√	√	√	H10N3F1S1	×	×	√
H5N5F1	×	√	√	H10N3F2S1	×	×	√

Table S3. List of N-glycans detected from FFPE mouse ovarian cancer tissue section.

m/z (+Na)	Identified N-glycans	Mass deviation (ppm)	m/z (+Na)	Identified N-glycans	Mass deviation(ppm)
933.3164	$Hex_3HexNAc_2$	0.6	1663.5847	$Hex_5HexNAc_4$	2
1079.3758	$Hex_3HexNAc_2dHex_1$	1	1743.5833	$Hex_8HexNAc_2$	1
1095.3714	$Hex_4HexNAc_2$	1.5	1768.6121	$Hex_6HexNAc_3dHex_1$	0
1257.4233	$Hex_5HexNAc_2$	0.5	1784.6084	$Hex_7HexNAc_3$	0.4
1282.4527	$Hex_3HexNAc_3dHex_1$	1	1809.6423	$Hex_5HexNAc_4dHex_1$	2
1298.4498	$Hex_4HexNAc_3$	0.5	1825.6335	$Hex_6HexNAc_4$	0
1419.4774	$Hex_6HexNAc_2$	1	1905.6362	$Hex_9HexNAc_2$	1
1444.5070	$Hex_4HexNAc_3dHex_1$	0	1971.6936	$Hex_6HexNAc_4dHex_1$	1
1460.5031	$Hex_5HexNAc_3$	0.7	2028.7106	$Hex_6HexNAc_5$	1.5
1485.5338	$Hex_3HexNAc_4dHex_1$	0	2067.6856	$Hex_{10}HexNAc_2$	0.5
1501.5296	$Hex_4HexNAc_4$	1	2133.7442	$Hex_7HexNAc_4dHex_1$	0
1581.5307	$Hex_7HexNAc_2$	1	2174.7692	$Hex_6HexNAc_5dHex_1$	1
1606.5614	$Hex_5HexNAc_3dHex_1$	1	2336.8212	$Hex_7HexNAc_5dHex_1$	1.3
1622.5553	$Hex_6HexNAc_3$	0	2539.8946	$Hex_7HexNAc_6dHex_1$	3
1647.5867	$Hex_4HexNAc_4dHex_1$	0			

Chapter 8

Conclusions and future directions

Conclusions

In this dissertation, multi-dimensional mass spectrometry (MS)-based approaches are applied to quantitatively characterizing proteins and their post-translational modifications (PTMs), mapping their spatial distributions on tissue, and understanding their involvement in physiological processes and pathological development. Novel analytical tools are also developed to facilitate the detection and confident identification of PTMs, specifically protein citrullination. The dissertation presents extensive method development and valuable datasets of various biological systems, both of which contribute to improved understanding of the diverse and critical roles of proteins and their PTMs *in vivo*, and will greatly inspire and benefit future investigations in related fields.

The dissertation starts with a comprehensive study of the development and maturation of human pancreas (**Chapter 2**). We delineate proteome-wide and extracellular matrix (ECM)-specific alterations in four age groups of healthy human pancreas: fetal (18-20 weeks gestation), juvenile (5-16 years old), young adults (21-29 years old) and older adults (50-61 years old) using 12-plex *N,N*-dimethyl leucine (DiLeu) isobaric tags. We also visualize and verify the alterations of specific ECM proteins using immunofluorescent staining and investigate their localization changes within islet or acinar compartments. There are many interesting observations in this study and some key take-home messages include: (1) The proteome of human pancreas is dynamically changing and remodeled in different age groups and becomes mature and stable in the adulthood. (2) Many previously reported pancreatic tumor biomarkers display significant changes in protein expression across multiple developmental stages, which suggests that biomarker identification and clinical implementation studies could be further strengthened by taking into consideration the age of the patient. (3) ECM proteins change not only in total abundance in pancreas but also in localization within islet or acinar cells which may provide insights for studying islet development,

function, and disease. More importantly, we believe our baseline data will provide a valuable resource to the scientific community so we try to make our dataset as accessible as we can, by providing all detailed information and creating a searchable online database. We anticipate this work will greatly benefit future studies especially in various pancreatic pathologies.

Chapters 3-6 depict our journey to improve the detection and understanding of an important PTM, namely citrullination, by developing various qualitative and quantitative approaches and applying them to different biological systems. In **Chapter 3**, we design and develop a novel biotin thiol tag which can react with citrulline residues with high specificity and efficiency. Detection of protein citrullination has long been challenging mainly because it is a very low-abundance PTM with a small mass shift of less than 1 Da, which makes it difficult to generate high-quality mass spectra that are easily distinguishable from the background. Our method not only enlarges the mass addition of this PTM to allow for more confident identification and localization, but also enables effective enrichment of target molecules. After careful proof-of-principle and optimization experiments, we demonstrate the utility of this pipeline by applying it to globally mapping protein citrullination in multiple mouse tissues. We present the first tissue-specific citrullination atlas in mice and generate the largest dataset to date. We unveil the widespread distribution of this PTM than previously anticipated and the complex biological functions that may be regulated through protein citrullination, such as neuronal signal transmission and metabolic processes. In addition, this method enables simultaneous detection of another chemically similar PTM called homocitrullination that occurs on lysine residues. To extend the applicability of the current technique in comparative studies, we integrate the workflow with isotopic dimethyl labeling to achieve a multiplexed quantitative analysis of protein citrullination.

Continuing the work described in the previous chapter, we find that dimethyl labeling suffers from limited multiplexing capability and lower identification rate due to complicated MS1 spectra. Therefore, in **Chapter 4**, we combine the qualitative analytical method with 12-plex DiLeu isobaric tags to achieve a high-throughput quantitative analysis of protein citrullination for the first time. We demonstrate that this strategy provides accurate and precise quantification results using both standard peptides and complex biological samples. We then take advantage of this optimized approach to investigate how citrullination is involved in the DNA damage response in human cell lines. We observe increased expression of citrullination after DNA damage treatment and the results indicate that citrullination may be closely associated with the regulation of translational activities and DNA repair processes.

Chapter 5 describes an enhanced citrullination profiling strategy by using a combination of various digestion and fragmentation methods. We first demonstrate the possibility of using LysargiNase, which is complementary to trypsin, during enzymatic digestion step for citrullination analysis, and explore the unique fragmentation patterns of LysargiNase digested peptides. Thanks to the orthogonality of different digestion and fragmentation techniques, we can achieve an in-depth characterization of citrullination in mouse brain and generate the most comprehensive analysis of a single tissue type. This method offers an attractive alternative towards a complete mapping of citrullinome from complex biological samples.

Chapter 6 represents an application of our novel methods to understanding biological questions that were previously impeded by the bottleneck of analytical tools. In this study, we investigate citrullination changes during the progression of Alzheimer's disease (AD) using human cerebrospinal fluid (CSF) samples. This is the first report of the presence of citrullinated proteins in CSF and their alterations with diseases. Overall, we find the expression level of protein

citrullination is decreased in AD patients compared to healthy individuals, which might be associated with the abnormal formation of plaques and tangles during disease progression. Specific regulation pathways are exemplified by the identification of citrullination sites on apolipoprotein E, clusterin and complement C3 for the first time, all of these proteins have been implicated in AD pathogenesis and development. This study uncovers the possible involvement of citrullination in AD pathogenesis and reveal potential biomarkers that may benefit diagnosis and development of therapeutic approaches of the disease.

Chapter 7 focuses on mass spectrometry imaging (MSI) analysis which retains information on spatial distribution of the analytes. In this work, we explore the application of a novel sub-atmospheric pressure (SubAP)/matrix assisted laser desorption/ionization (MALDI) source coupled with a high-resolution accurate mass spectrometer for *in situ* imaging of N-linked glycans from formalin-fixed paraffin-embedded (FFPE) tissue sections. We first optimize the acquisition parameters of the SubAP/MALDI instrument and streamline the workflow for the imaging of glycans. We further apply the novel platform to mapping the spatial distribution of N-glycans from mouse ovarian cancer tissue sections. Different spatial distribution patterns of N-glycans, especially the abnormal accumulation of high-mannose N-glycans in cancer regions, are observed, demonstrating the capability of the new platform for high-resolution imaging analysis of biological samples.

Future directions

In **Chapter 2**, the whole-tissue mass spectrometry methodology we employ cannot be used exclusively to make conclusions about ECM and proteins relevant specifically to islet development. While our main goal is to characterize the broad changes of the tissue, we also think it is very important and meaningful to sort out ECM proteins enriched in islets. For this reason, we use

immunofluorescent staining to visualize the localization of selected ECM proteins in islets compared to acinar regions throughout development. However, a large-scale screening of region-specific proteins in pancreas is precluded using this method and alternative strategies are needed. Investigating isolated islets is a common strategy for such purposes. However, isolated islets are not always pure and therefore often contain a substantial fraction of acinar tissue as well. In addition, effective ways to isolate islets from the human pancreas inherently destroy the ECM and several ECM proteins are not regenerated during islet culture.[1-2] Therefore, future studies may focus on using laser-capture microdissection mass spectrometry (LCM-MS)[3-4] or MSI[5] techniques to characterize the proteome of human islets *in situ* and to investigate the localization of specific proteins in the pancreas, without the need for islet isolation. This will avoid the inherent damage of islets during the process of isolation and culture, and give much more meaningful data regarding islet development in healthy human tissues. Possible pitfalls in LCM-MS experiments include instrumentation difficulties and the low amount of starting materials that can be collected in a reasonable timeframe. MSI methodology may also suffer from low ionization efficiency of certain analytes and ambiguous identifications without fragmentation validation.

Chapters 3-6 demonstrate a promising method for large-scale citrullination analysis both qualitatively and quantitatively, which enables many fascinating ongoing projects and future directions. First of all, we are planning to investigate the association of protein citrullination in other pathological states such as in various cancers with this technological advancement, since previous studies were far from comprehensive due to limitations of methodology. In **Chapter 6**, we explore the roles of citrullination during the progression of AD. However, the CSF sample is obtained by invasive lumbar punctures which precludes it to be easily used in diagnostic approaches. In contrast, the blood sample is more easily accessible and share some common

components with CSF. Therefore, it would be worth exploring the possibility of translating the CSF biomarkers to diagnostic markers in blood. Second, we hope to optimize aspects of the workflow to further improve the depth and reproducibility of current analysis. For example, we have been using data-dependent acquisition (DDA) for MS analysis. However, the stochastic precursor selection of DDA leads to inconsistent detection of peptides and high variation across replicates. Alternatively, data-independent acquisition (DIA) is a recent development in quantitative proteomics and has been shown to be superior to DDA by producing a higher number of quantified proteins in shorter analysis time, fewer missing values, and lower coefficients of variation across replicates.[6-7] A preliminary experiment shows that DIA acquisition greatly improves the identification rate by more than 50% (**Figure 1**), suggesting its potential for enhanced citrullination analysis. Last but not least, we are also aware of a few intrinsic drawbacks of our method. The biotin tag itself generates two fragment ions during HCD events, which likely complicates the spectra and suppresses the backbone fragmentation of peptides for sequence annotation. The reaction, enrichment and release steps involving biotin-avidin interactions can be a little tedious and may also result in non-specific binding and thus background noise in MS acquisition. To address this limitation, we seek to develop a next-generation method to boost citrullination detection step further. For instance, a solid support could be used in lieu of biotin head group for enrichment purposes and a cleavable linker (e.g., trypsin-cleavable) could be incorporated to allow for release of captured peptides (**Figure 2**). However, the reaction efficiency on the solid support needs to be carefully evaluated. The length and sequence of the linker should also be optimized with caution to enable effective cleavage while keeping a moderate mass addition to the analytes.

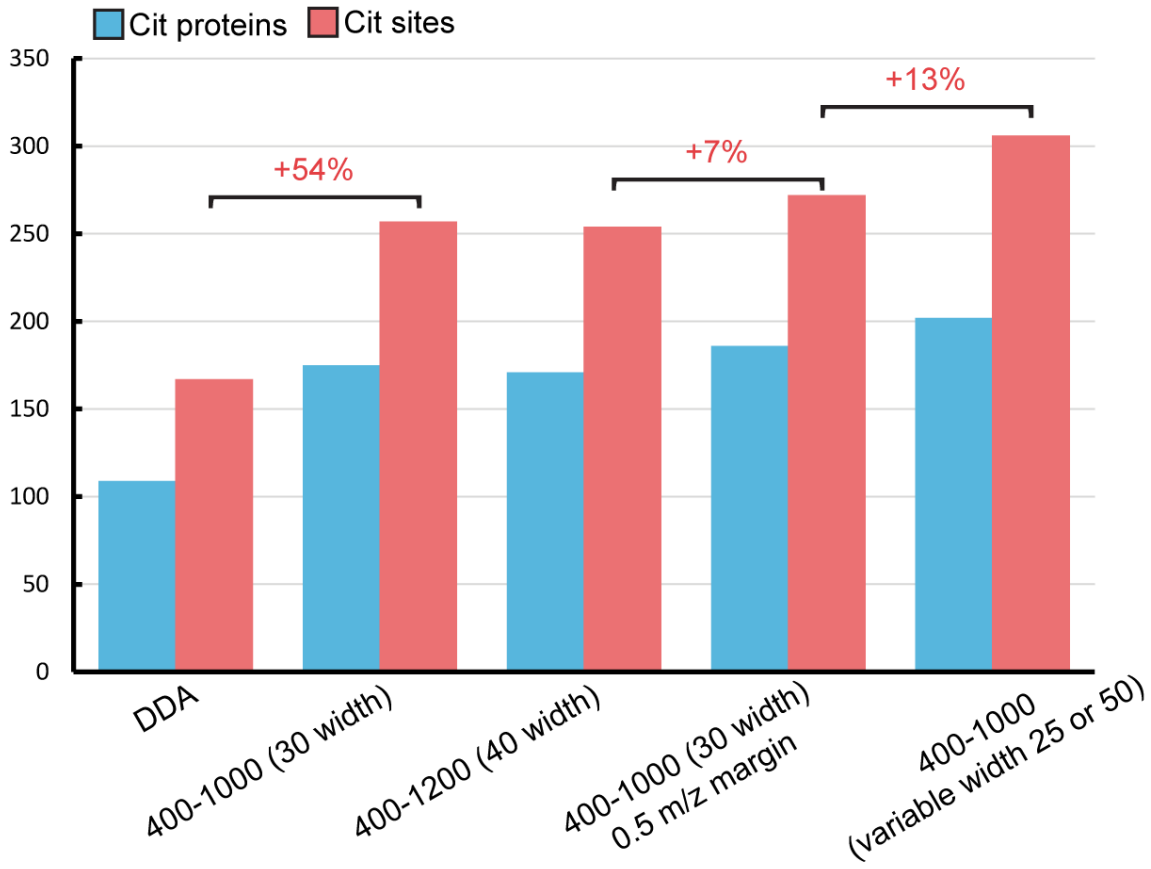

Figure 1. Comparison of citrullination identification numbers using DDA and four DIA-based MS acquisition methods.

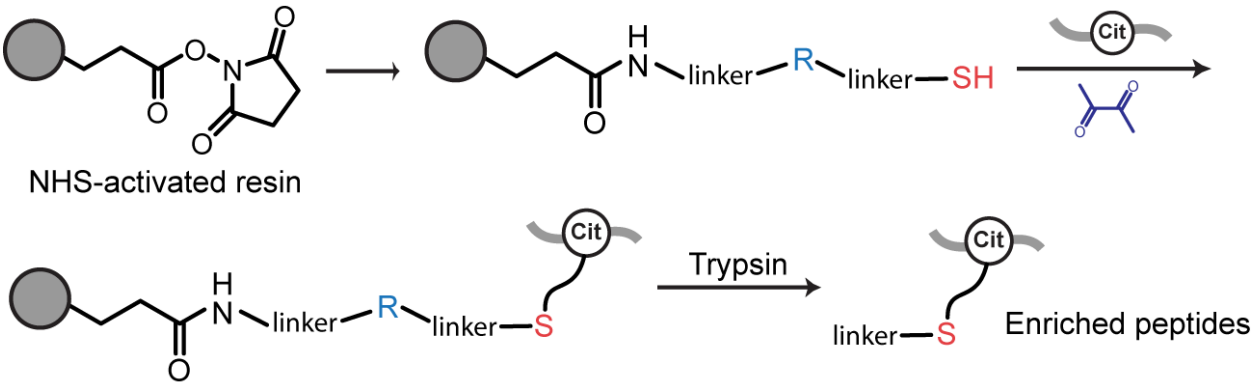

Figure 2. Potential scheme of next-generation method for citrullination analysis.

References

(1) Cross, S. E.; Vaughan, R. H.; Willcox, A. J.; McBride, A. J.; Abraham, A. A.; Han, B.; Johnson, J. D.; Maillard, E.; Bateman, P. A.; Ramracheya, R. D.; Rorsman, P.; Kadler, K. E.; Dunne, M. J.; Hughes, S. J.; Johnson, P. R. *Am. J. Transplant.* **2017,** *17*, 451-461.

(2) Meier, R. P. H.; Meyer, J.; Muller, Y. D.; Szot, G. L.; Bedat, B.; Andres, A.; Masse, N.; Lablanche, S.; Puppa, G.; Bosco, D.; Berney, T. *Transpl. Int.* **2020,** *33*, 1516-1528.

(3) Zhang, L.; Lanzoni, G.; Battarra, M.; Inverardi, L.; Zhang, Q. *J Proteomics* **2017,** *150*, 149-159.

(4) Nyalwidhe, J. O.; Grzesik, W. J.; Burch, T. C.; Semeraro, M. L.; Waseem, T.; Gerling, I. C.; Mirmira, R. G.; Morris, M. A.; Nadler, J. L. *PloS one* **2017,** *12*, e0183908.

(5) Dilillo, M.; Pellegrini, D.; Ait-Belkacem, R.; de Graaf, E. L.; Caleo, M.; McDonnell, L. A. *Journal of proteome research* **2017,** *16*, 2993-3001.

(6) Koopmans, F.; Ho, J. T. C.; Smit, A. B.; Li, K. W. *Proteomics* **2018,** *18*.

(7) DeLaney, K.; Li, L. *Anal. Chem.* **2019,** *91*, 5150-5158.

Appendix I

Methods and protocols

1.1. Overview of experimental workflow for bottom-up quantitative proteomics and post-translational modification (citrullination) analyses

Typical workflows for bottom-up quantitative proteomics and citrullination analysis are shown in **Figure 1**. To prepare peptides from complex biological samples, proteins will be first extracted out of cells or tissues. Various protein extraction methods have been explored to either improve the depth of proteome coverage or facilitate the detection of certain types of proteins. This appendix will detail the procedures of three universal methods for protein extraction, including in-solution digestion, Filter Aided Sample Preparation (FASP) method, and Surfactant and Chaotropic Agent assisted sequential extraction/on-pellet Digestion (SCAD) method. After extraction, protein amount is determined by bicinchoninic acid (BCA)-based assays with commercial kits. This step is usually necessary for samples without previous knowledge of protein concentration, and is aimed to provide reference for the amount of enzymes added during following digestion steps. Some studies also use protein BCA assay as the only normalization step to ensure accurate quantification results. Protein extracts will then be reduced to break the disulfide bonds, which helps improve digestion efficiency. Alkylation is usually recommended to block free thiol groups and prevent formation of disulfide bonds again. FASP and SCAD methods involve the utilization of surfactant such as sodium dodecyl sulfate (SDS) to improve the extraction efficiency and solubilization capability of proteins in biological samples. However, common surfactant is not compatible with enzymatic digestion and is also detrimental to MS acquisition even at a very low concentration. To efficiently remove surfactant, FASP uses filter-aided buffer exchange while SCAD method takes advantage of protein precipitation. Subsequent enzymatic digestion step usually with trypsin cleaves large proteins into small peptides and solid phase extraction (SPE) clean-up is followed to get rid of salts and contaminants. Peptide assay which

measures peptide concentration through colorimetry (similar to BCA assay) is experimentally more accurate to decide the MS injection volume and normalize the peptide amount for further chemical labeling processes.

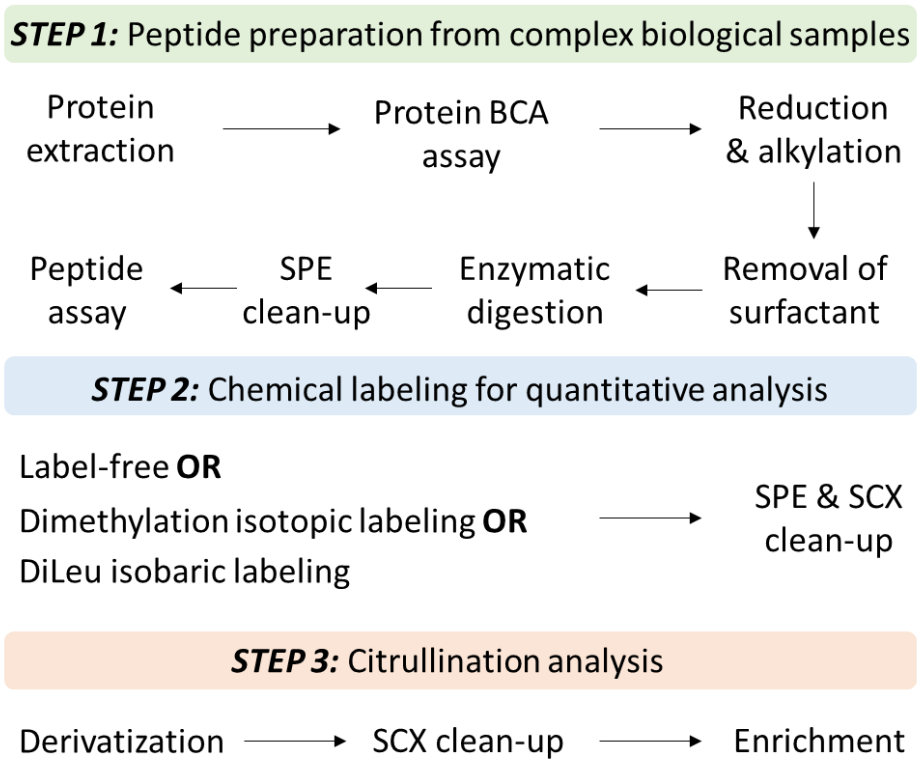

Figure 1. Workflows for quantitative proteomics and citrullination analysis

Quantitative analysis can be achieved mainly through three methods: label-free approach, isotopic labeling (e.g., reductive dimethylation) and isobaric labeling (DiLeu). In label-free approach, multiple samples are prepared in parallel and analyzed separately in MS. Quantification is achieved using peak areas or spectral counting of certain molecules. Although easy to perform, label-free approach often requires longer instrument time and can suffer from run-to-run variations. Isotopic labeling method includes a chemical labeling step of different samples, which leads to different mass addition to these labeled samples. Samples can then be mixed for analysis and distinguished at MS1 level. Quantitation can be performed by comparing the peak intensities of these differentially labeled samples. Isotopic labeling provides certain degree of multiplexing

capability (up to five plex for dimethylation but two or three plex is more often utilized) but often suffers from lower identification rate due to complicated MS1 spectra. Isobaric labeling also takes advantage of chemical labeling using different stable isotope-incorporated reagents. However, in this case different channels of the chemical tags will have same nominal mass yet different isotope combination. Consequently, differentially labeled samples will exhibit single peaks at MS1 level but can generate distinct reporter ions at low-mass range upon MS2 fragmentation which can be used for quantitative comparison. Isobaric labeling enables high-throughput quantitative analysis for up to 21 samples (DiLeu). Sample clean-ups are usually necessary after these labeling steps to remove excess tags or reagents. SPE desalting is sufficient for dimethylation but strong cation exchange is needed after DiLeu labeling due to its similar hydrophobic properties to peptide samples.

Citrullination analysis starts with the derivatization of target citrullinated peptides (ureido group reaction) using chemical probes (e.g., biotin thiol tag). Excess tags need to be removed using SCX before enrichment such as using streptavidin beads. To improve the identification rate in proteomics or many other PTM analyses, high pH fractionation is sometimes utilized prior to LC-MS/MS. However, this approach turns out to be less helpful for citrullination characterization probably due to the very low abundance of this PTM.

1.2. Peptide preparation methods from complex biological samples

1.2.1. In-solution digestion

(1) Prepare 50 mM tris buffer:

181.71 mg tris base; 30 mL H_2O; add 60 uL HCl to adjust pH~8.

Prepare 8 M urea buffer:

Urea 4.8048 g (8 M); $CaCl_2$ 7.3 mg (5 mM); NaCl 17.5 mg (30 mM); 1 tablet of protease inhibitor; 1 tablet of phosphate inhibitor; add tris buffer to 10 mL; pH~8.

*Note: $CaCl_2$ and NaCl are optionally added to improve extraction efficiency for cells. Phosphate inhibitor is not needed if not working on phosphorylation.

(2) Add appropriate amount of urea buffer to each sample. Use probe sonicator to extract proteins from tissues. Alternatively, vortex, sonication and heating are enough for most cell samples.

(3) Perform BCA protein assay according to the manufacturer's protocols.

(4) Prepare 100 mM DTT in tris buffer. Add DTT to each sample to a final concentration of 10 mM, incubate at room temperature for 1 h.

(5) Prepare 200 mM IAA in tris buffer. Add IAA to each sample to a final concentration of 50 mM, incubate in dark at room temperature for 0.5 h.

(6) Add same amount of DTT in (4), incubate for 5 min to react with excess IAA.

(7) Digestion (trypsin/LysC for example). Add enzyme to each sample (1:50 to 1:200 enzyme/protein w/w ratio), incubate at 37 ℃ for 4 h. Add tris buffer to dilute urea to <1 M, incubate at 37 ℃ overnight.

(8) Add 10% TFA to each sample to a final concentration of 0.25% to quench the digestion, votex, centrifuge at 14,000 g for 15 min.

(9) Desalt the supernatant using Sap-Pak (<2 mg) or Omics tips (<100 ug) according to the manufacturer's protocols. Dry down the sample.

(10) Resuspend samples with H_2O, perform peptide assay according to the manufacturer's protocols. Aliquot appropriate amount of samples and dry down. Samples are ready for LFQ analysis or subjected to labeling steps or PTM enrichment.

1.2.2. FASP method

(1) Prepare 4% SDS, add 50 mM tris buffer to 10 mL (4 mL 10% SDS, 2.5 times of dilution), check pH~8. Prepare 8 M urea in tris buffer (no need to add protease inhibitor due to the strong denaturing ability of SDS). Prepare 100 mM ammonium bicarbonate (ABC) solution.

(2) Lyse cells or tissues in buffer using probe sonication (below 200 µL buffer works best for following centrifugation of filters), centrifugate at 14,000 g for 5 min. Transfer the supernatant to a new tube.

(3) Perform BCA protein assay.

(4) Transfer appropriate amount of samples to 30 kDa filters.

(5) Prepare 100 mM DTT in tris buffer. Add DTT to a final concentration of 10 mM. Incubate at room temperature for 30-60 min.

(6) Prepare 200 mM IAA in tris buffer. Add IAA to a final concentration of 50 mM. Incubate in dark at room temperature for 30 min.

(7) Centrifuge at 14,000 g for 15 min at 20 ℃.

(8) Add 200 µL of urea to the sample to rinse off small molecule contaminants, centrifuge at 14, 000 g for 15 min at 20 ℃, repeat 2x times.

(9) Add 100 µL of ABC, centrifuge at 14,000 g for 15 min at 20 ℃, repeat 2x times.

(10) Resuspend appropriate amount of enzymes in 50 uL ABC, add to the filters. Incubate overnight at 37 ℃.

(11) Transfer the filter unit to new collection tube and centrifuge at 14,000 g for 10 min.

(12) Add 50 µL 0.5 M NaCl and centrifuge the filter units at 14,000 g for 10 min. Repeat for 1x time.

(13) Add 350 µL DI water and 12.5 µL 10% TFA to the vial (bigger volume to decrease sample loss, finally 0.25% TFA, pH 2~3, finally 500 µL).

(14) Use Sep-Pak to desalt (or Omics tips if less than 100 µg).

(15) Perform peptide assay.

1.2.3. SCAD method

(1) Prepare 50 mM tris buffer, prepare 4% SDS in tris buffer, prepare 8 M urea in tris buffer. Prepare 80% acetone and store in -20 ℃.

(2) Lyse cells or tissues in buffer using probe sonication, centrifugate at 14,000 g for 5 min. Transfer the supernatant to a new tube.

(3) Perform BCA protein assay.

(4) Prepare 100 mM DTT in tris buffer. Add DTT to each sample to a final concentration of 10 mM, incubate at room temperature for 30 min.

(5) Prepare 200 mM IAA in tris buffer. Add IAA to each sample to a final concentration of 50 mM, incubate in dark at room temperature for 30 min.

(6) Add same amount of DTT in (4), incubate for 5 min to react with excess IAA.

(7) Add 4 times volume of cold 80% acetone to the samples (should observe precipitation of proteins). Sit in -20 ℃ freezer overnight.

(8) Centrifuge at 14,000 g for 15 min, discard supernatant containing SDS.

(9) Add same amount of cold 80% acetone, sit in -20 ℃ freezer for 1 h. Centrifuge at 14,000 g for 10 min, discard supernatant. Repeat 1x time.

(10) Open tube lid, let air dry for 12 min.

(11) Add 8 M urea to dissolve the protein pellet (150 uL is enough for 1 mg protein, slightly increase the amount if not dissolved).

(12) Digestion (trypsin/LysC for example). Add enzyme to each sample (1:50 to 1:200 enzyme/protein w/w ratio), incubate at 37 ℃ for 4 h. Add tris buffer to dilute urea to <1 M, incubate at 37 ℃ overnight.

(13) Add 10% TFA to each sample to a final concentration of 0.25% to quench the digestion, votex, centrifuge at 14,000 g for 15 min.

(14) Use Sep-Pak to desalt (or Omics tips if less than 100 µg).

(15) Perform peptide assay.

1.3. Chemical labeling for quantitative analysis

1.3.1. Reductive dimethylation

The two-plex uses formaldehyde (OCH_3) and deuterium formaldehyde (OCD_3) as dimethylation agents, and the three-plex adds carbon-13 deuterium formaldehyde ($O^{13}CD_3$) as an additional dimethylation agent. The two-plex protocol does not add sodium ions to the sample due to the use of a milder H- donor, borane pyridine as opposed to $NaBH_3CN$. However, for the heaviest label, D- must be donated instead of H- to produce a total mass shift of four, because simply using carbon-13 deuterium formaldehyde will only produce a mass difference of two, which could have interference from naturally occurring isotopes of the compound labeled with deuterium formaldehyde. For this reason, it uses $NaBH_3CN$ for the light and intermediate label, and $NaBD_3CN$ for the heavy label. A light label incorporates 2(CH_2) per primary amine; an intermediate label incorporates 2(CD_2) per primary amine; and a heavy label incorporates 2(CD_2-H+D) per primary amine. All peptide N-termini and lysine side chains will be labeled. Take duplex dimethylation experiment for 400 ug peptides as an example.

(1) Dissolve sample in ~40 µL of water. Sonicate to ensure full dissolution.

(2) Prepare 1% formaldehyde (dilute from 37%), 1% deuterium formaldehyde (dilute from 20%), and 30 mM borane pyridine (dilute from 8 M) solutions.

(3) Add 20 μL of formaldehyde solution to samples for light labeling or 20 μL of deuterium formaldehyde solutions to samples for intermediate labeling.

(4) Add 20 μL of borane pyridine solution to each sample.

(5) Incubate at 37 °C for 15 min.

(6) Add 20 μL100 mM ammonium bicarbonate solution to quench the reaction.

(7) Add TFA/FA to adjust the pH <3, desalt as needed.

1.3.2. DiLeu labeling

1.3.2.1. DiLeu labeling of peptides

Protocols used to label 50-100 ug peptides (1:10-1:20 tag ratio).

(1) Dry down DiLeu tags to fully dehydrate them.

(2) Prepare activation solution (100 uL needed for each channel):

1500 uL anhydrous DMF; 7.076 uL NMM, 21.114 mg DMTMM

(3) Add 100 uL of activation solution to each vial of DiLeu, sonication, shake for 45 min.

(4) Resuspend each sample in 20 uL 0.5 M TEAB.

(5) Centrifuge activated DiLeu solution, pipette out 100 uL supernatant to each vial.

(6) Incubate for 2 h with shaking.

(7) Add 6.3 uL of 5% hydroxylamine to quench (final concentration of 0.25%), shake 10 min. Dry down.

1.3.2.2. SCX clean-up

Excess DiLeu tags need to be removed by SCX as shown below.

(1) Prepare buffer:

Loading and washing buffer: 25% ACN, 0.1% formic acid, pH <3

Elution buffer: 25% ACN, 0.4 M ammonium formate, pH ~7

(2) Condition the column. Sit the tip onto a 2 mL tube (with an adaptor). Add 100 uL loading buffer to the tip, centrifuge (1,000 or 1,200 rpm, 2 min) and discard flowthrough, repeat this step for a total of 3 times.

(3) Reconstitute the sample in 200 uL loading buffer, add to the tip and centrifuge (1,000 or 1,200 rpm, 4 min). Reload the flowthrough sample once to ensure complete binding.

(4) Add 100 uL loading buffer to the tip, centrifuge (1,000 or 1,200 rpm, 2 min) and discard flowthrough, repeat this step for a total of 5-10 times.

(5) Transfer the tip to a clean tube, add 100 uL elution buffer, centrifuge (1,000 or 1,200 rpm, 2 min) and collect flowthrough, repeat for a total of 2 times (or add 80 uL * 3 times) and dry down.

(6) Desalt as needed.

1.4. Citrullination analysis

1.4.1. Derivatization of citrullinated peptides

(1) Check the mass accuracy of biotin tag on MALDI instrument. Should obverse m/z 304 and its sodium adduct (m/z 326). Add TCEP to 2% if peaks at m/z 605 and 627 are high (dimer).

(2) Add 30 uL (0.3 mg) of biotin tag to each sample, dry down.

(3) Resuspend the samples using 40 uL 12.5% TFA.

(4) Prepare 2,3-butanadione solution by adding 1 uL of the pure reagent to 114 uL 12.5% TFA.

(5) Add 10 uL 2,3-butanadione solution to each sample. Incubate at 37 °C in dark with shaking for 6 h. Dry down the samples.

(6) Perform SCX to remove excess tags.

1.4.2. Enrichment of derivatized peptides

All steps are performed in 1.5 ml tubes.

(1) For 400 ug starting peptides, withdraw 120 uL streptavidin beads (contains ~60 uL beads plus stock solution).

(2) Add 1000 uL PBS (1 ×), manually vortex up and down, centrifuge (3,000 g, 2 min), discard supernatant, repeat for 5 times in total.

(3) Resuspend sample in 1000 uL PBS, add to beads, rotate for 2 h.

(4) Discard supernatant, add 1000 uL PBS, manually vortex up and down, centrifuge (3,000 g, 2 min), discard supernatant, repeat for 3 times in total.

(5) Add 1000 uL PBS with 5% ACN, manually vortex up and down, centrifuge (3,000 g, 2 min), discard supernatant, repeat for 3 times in total.

(6) Add 1000 uL H_2O, manually vortex up and down, centrifuge (3,000 g, 2 min), discard supernatant, repeat for 10 times in total.

(7) Elute with 80% ACN, 0.2% TFA, 0.1% FA. Add 300 uL, manually vortex up and down, centrifuge (3,000 g, 2 min), collect supernatant.

(8) Add 300 uL elution solution, manually vortex up and down, heat to 90 ℃ for 5 min, centrifuge (3,000 g, 2 min), collect supernatant. Repeat for 3 times. Combine all four elution.

(9) Filter the samples (1.2 mL elute in total) with 0.22 um filters (safe for MS), dry down.

1.5. High pH fractionation

(1) Sign the form to reserve instrument

(2) Prepare mobile phase A: H_2O, 10 mM ammonium formate, pH=10 (126 mg is required to make 200 mL, 920 µL ammonium hydroxide to adjust pH).

Mobile phase B: 90% ACN, 10 mM ammonium formate, pH=10 (126 mg is required to make 200 mL, 4.4 mL ammonium hydroxide to adjust pH).

(3) C18 column is stored at 4-degree refrigerator. Install the column and set the column temperature to be 30 degrees.

(4) Ignite UV detector for 20 min (press “shift”+”1” for two times).

(5) Check all solvents on the instrument.

(6) Following the operation manual below.

a) Seal Wash

10% methanol, no acetonitrile. This wash cleans behind the pump-head seals to ensure proper lubrication. Minimum once per week.

1. On instrument interface, navigate Menu/Status > Diag > Prime SealWsh > Start
2. Press Stop after 5 minutes

b) Prime Injector

10% methanol for maintenance; high organic solvent for dirty runs (e.g. 95% acetonitrile). Done before injecting any real samples to ensure no bubbles are present in the injector. Minimum once per week.

1. On instrument interface, navigate Menu/Status > Diag > Prime NdlWsh > Start
2. After completion, press Close

c) Purge Injector (exit----status)

Solvent is dependent on run. Run this protocol at beginning of experiments each day. Minimum once per week for maintenance.

1. On instrument interface, navigate Menu/Status > Status screen
2. Navigate Direct Function > 4 Purge Injector > OK
3. Set Sample loop volumes 6.0, leave Compression Check unchecked > OK

d) Prime Solvent Pumps

Solvent is dependent on run. If solvents are changed, run this protocol. Minimum once per week for maintenance.

1. On instrument interface, navigate Menu/Status > Status screen
2. Using arrow keys, choose composition A, type 100%, Enter x4
3. Navigate Direct Function > 3 Wet Prime > OK
4. Set Flow Rate 7.000 mL/min, Time 1.00 min > OK
5. Repeat for all changed/active solvent pumps

e) Condition Column

Dependent on user. Use starting conditions for run.

1. On instrument interface, navigate Menu/Status > Status screen
2. Using arrow keys, type starting solvent compositions for run (100% A, 0.2 mL/min)
3. Navigate Direct Function > 6 Condition Column > OK
4. Set Time as desired

(7) Set the pump flow rate to be 0.2 mL/min, normal column pressure is 1500-2000 psi.

(8) Method: 2015_Xiaofan_HpH_C18_94min. The maximum injection volume is 100 µL. Dissolve sample in 108 µL mobile phase A for injection.

(9) Run a blank sample before sample analysis.

(10) Set the fractionation collector. Discard the first 8 min flow-through, and collect the flow-through every 4 to 6 min depending on applications.

(11) Run another blank sample to wash the column.

(12) Change the mobile phase to water. Wash the column for another 10 min.

(13) Close the UV light and put the column back to the 4-degree refrigerator.

(14) Concatenate the non-adjacent fractions into desired number of fractions. Dry down samples.

Appendix II

List of selected publications and presentations

Publications

1. **Li, Z.**[#]; Tremmel, D.[#]; Ma, F.; Yu, Q.; Ma, M.; Delafield, D.; Shi, Y.; Wang, B.; Mitchell, S.; Feeney, A.; Jain, V.; Sackett, S.; Odorico, J.; Li, L., Proteome-wide and matrisome-specific alterations during human pancreas development and maturation. *Nature Communications* **2021.** 12 (1), 1020.

2. Shi, Y.[#]; **Li, Z.**[#]; Wang, B.; Shi, X.; Ye, H.; Delafield, D.; Lv, L.; Ye, Z.; Chen, Z.; Ma, F; Li, L., Enabling global analysis of protein citrullination and homocitrullination via biotin thiol tag-assisted mass spectrometry. **2021.** Under review.

3. **Li, Z.**; Wang, B.; Yu, Q.; Shi, Y.; Li, L., 12-plex DiLeu isobaric labeling enabled high-throughput investigation of citrullination alterations in DNA damage response. **2021.** Under review.

4. **Li, Z.**; Ma, M.; Wang, B.; Yu, Q.; Shi, Y.; Li, L., LysargiNase complements trypsin for enhanced profiling of protein citrullination. **2021.** To be submitted.

5. **Li, Z.**; Wang, B.; Shi, Y.; Yu, Q.; Li, L., Mass spectrometric profiling of citrullination alterations in cerebrospinal fluid of patients with Alzheimer's disease. In preparation.

6. Yu, Q.[#], Ma, M.[#], Delafield, D., Cui, Y., **Li, Z.**, Wu, W., Shi, X., Westmark, P., Westmark, C., Li, L., On-tissue spatial proteomics integrating MALDI-MS imaging with shotgun proteomics reveals soy consumption-induced biomarkers in a fragile X syndrome mouse model. **2021.** Under review.

7. Yu, Q., Feng, Y., **Li, Z.**, Ma, M., Liu, F., Xu, W., Li, L. SUGAR tag enabled multiplexed quantitative N-glycan analysis including sialic acid linkage isomers. **2021.** To be submitted.

8. Gant, K.; Jambor, A.; **Li, Z.**; Rentchler, E.; Weisman, P.; Li, L.; Patankar, M.; Campagnola, P., Imaging Collagen Alterations in Early Precursor Lesions of High Grade Serous Ovarian Cancer by Second Harmonic Generation Microscopy. *Cancers* **2021.** 13 (11), 2794.

9. Zhang, H.; Xu, M.; Shi, X.; Liu, Y.; **Li, Z.**; Jagodinsky, J.; Ma, M.; Welham, N.; Morris, Z.; Li, L., Characterization, Quantification and Molecular Imaging of Fatty Acid Isomers from Biological Samples Using Peracetic Acid Induced Epoxidation Coupled with Mass Spectrometry. *Chemical Science* **2021.** 12, 8115-8122.

10. Tabang, D.; Cui, Y.; Tremmel, D.; Ford, M.; **Li, Z.**; Sackett, S.; Odorico, J.; Li, L., Analysis of Pancreatic Extracellular Matrix Protein Post-Translational Modifications via Electrostatic Repulsion-Hydrophilic Interaction Chromatography Coupled with Mass Spectrometry. *Molecular Omics* **2021.**

11. Atkinson, L.; Liu, Y.; McKay, F.; Vandewyer E.; Viau, C.; Irvine, A.; Rosa, B.; **Li, Z.**; Marks, N.; Maule, A.; Mitreva, M.; Beets, I.; Li, L., Mousley, A., *Ascaris suum* Informs Extrasynaptic Volume Transmission in Nematodes. *ACS Chemical Neuroscience* **2021.** 12 (17) 3176-3188.

12. Zhang, H.; Shi, X.; Vu, N.; Li, G.; **Li, Z.**; Shi, Y.; Li, M.; Wang, B.; Welham, N.; Patankar, M.; Li, L., On-tissue derivatization with Girard's reagent P enhances N-glycan signals for formalin-fixed paraffin-embedded tissue sections in MALDI mass spectrometry imaging. *Analytical Chemistry* **2020.** 92 (19), 13361-13368.

13. Liu, R.; Wei, P.; Keller, C.; Orefice, N.; Shi, Y.; **Li, Z.**; Huang, J.; Cui, Y.; Frost, D.; Han, S.; Cross, T.; Rey, F.; Li, L., Integrated Label-Free and 10-plex DiLeu Isobaric Tag Quantitative Methods for Profiling Changes in the Mouse Hypothalamic Neuropeptidome and Proteome: Assessment of the Impact of the Gut Microbiome. *Analytical Chemistry* **2020.** 92 (20), 14021-14030.

14. Shi, Y.; **Li, Z.**; Felder, M.; Yu, Q.; Shi, X.; Peng, Y.; Cao Q.; Wang, B.; Puglielli, L.; Patankar, M; Li, L., High-resolution mass spectrometry imaging of N-glycans from formalin-fixed paraffin-embedded tissue sections using a novel sub-atmospheric pressure ionization source. *Analytical Chemistry* **2019.** 91 (20), 12942-12947.

15. Ma, F.; Tremmel, D.; **Li, Z.**; Lietz, C.; Sackett, S.; Odorico, J.; Li, L., In Depth Quantification of Extracellular Matrix Proteins from Human Pancreas. *Journal of Proteome Research* **2019.** 18 (8), 3156-3165.

16. Liu, Y.; Buchberger, A.; DeLaney, K.; **Li, Z.**; Li, L., Multifaceted Mass Spectrometric Investigation of Neuropeptide Changes in Atlantic Blue Crab, *Callinectes sapidus*, in Response to Low pH Stress. *Journal of Proteome Research* **2019.** 18 (7), 2759-2770.

Presentations

1. **Li, Z.**; Tremmel, D.; Ma, F.; Yu, Q.; Ma, M.; Delafield, D.; Shi, Y.; Wang, B.; Mitchell, S.; Feeney, A.; Jain, V.; Sackett, S.; Odorico, J.; Li, L., Proteome-wide and matrisome-specific alterations during human pancreas development and maturation revealed by 12-plex isobaric DiLeu tags. 1st CASMS Conference. 2021. **Lightening talk and poster presentation (virtual)**.
2. **Li, Z.**; Tremmel, D.; Ma, F.; Yu, Q.; Ma, M.; Delafield, D.; Shi, Y.; Wang, B.; Mitchell, S.; Feeney, A.; Jain, V.; Sackett, S.; Odorico, J.; Li, L., Proteome-wide and matrisome-specific alterations during human pancreas development and maturation revealed by 12-plex isobaric DiLeu tags. US HUPO Annual Conference. 2021. **Lightening talk and poster presentation (virtual)**.
3. **Li, Z.**; Shi, Y., Wang, B., Li, L., Discovery of novel citrullination biomarkers using cerebrospinal fluid in patients with Alzheimer's disease. 68th Annual ASMS Conference on Mass Spectrometry and Allied Topics. Houston, TX. 2020. **Poster presentation (virtual)**.
4. **Li, Z.**; Tremmel, D.; Ma, F.; Sackett, S.; Odorico, J.; Li, L., 12-plex DiLeu isobaric tags (DiLeu) enable quantitative proteomic analysis of human pancreatic tissues at various developmental stages. Pittcon conference & expo. Chicago, IL. 2020. **Oral presentation**.
5. **Li, Z.**; Tremmel, D.; Ma, F.; Sackett, S.; Odorico, J.; Li, L., 12-plex DiLeu isobaric tags (DiLeu) enable quantitative proteomic analysis of human pancreatic tissues at various developmental stages. Rising Stars in Analytical Chemistry and Materials Science Symposium. Rahway (Merck & Co.), NJ. 2019. **Oral presentation**.
6. **Li, Z.**; Shi, Y., Ma, F., Gant, K., Patankar, M., Li, L., MS-based strategies reveal extracellular matrix alterations and N-glycan spatial distribution changes with the progression of ovarian cancer. 67th Annual ASMS Conference on Mass Spectrometry and Allied Topics. Atlanta, GA. 2019. **Poster presentation**.
7. **Li, Z.**; Liu, Y.; Li, L., Mass spectrometric investigation of neuropeptide changes in blue crab Callinectes sapidus in exposure to silver nanoparticles. 66th Annual ASMS Conference on Mass Spectrometry and Allied Topics. San Diego, CA. 2018. **Poster presentation**.