



OPEN

# Global chemical modifications comparison of human plasma proteome from two different age groups

Yongtao Liu<sup>1,3</sup>, Xuanzhen Pan<sup>1,3</sup>, Mindi Zhao<sup>2</sup> & Youhe Gao<sup>1</sup>✉

In this study, two groups of human plasma proteome at different age groups (old and young) were used to perform a comparison of global chemical modifications, as determined by tandem mass spectrometry (MS/MS) combined with non-limiting modification identification algorithms. The sulfhydryl in the cysteine A total of 4 molecular modifications were found to have significant differences passing random grouping tests: the succinylation and phosphorylation modification of cysteine (Cys, C) and the modification of lysine (Lys, K) with threonine (Thr, T) were significantly higher in the old group than in the young group, while the carbamylation of lysine was lower in the young group. We speculate that there is an increase in certain modified proteins in the blood of the old people which, in turn, changes the function of those proteins. This change may be one of the reasons why old people are more likely than young people to be at risk for age-related diseases, such as metabolic diseases, cerebral and cardiovascular diseases, and cancer.

The biological processes of living things/organisms can perform and operate in mutual coordination and high-efficiency cooperation simultaneously, relying on the synthesis, catalysis, and regulatory reactions in which proteins are involved. Proteins are biological macromolecule with complex structures, and the difference in the high-level structures of proteins determines their different biochemical activities, which can perform specific functions through a higher-level network formed by the combination of proteins<sup>1</sup>. Chemical modification of proteins refers to the covalent group reaction of amino acid residues or chain ends. In general, a few changes in chemical structure do not affect the biological activity of proteins, which are called modification of nonessential parts of the protein. However, in most cases, changes in molecular structure will significantly change the physical and chemical properties of the protein, change the conformation of the protein, and make the activity of the protein vary, and then the function that it performs is changed<sup>2</sup>. Therefore, even if there is no change in the protein content level, but some small changes in the level of chemical modification, the function of the proteins will change significantly, which means that the chemical modification of the proteins enriches the different functions of the protein in another dimension. The effects of chemical modification of proteins on the functions of proteins are mainly shown in the following three aspects: (1) Even if a modification occurs, the functions of proteins will be affected. (2) The same type or different type of modifications of the different amino acids have different effects on the same protein's function. (3) The same protein may undergo many types of chemical modification, which makes the biological process in which it participates a more variable and complex process<sup>3</sup>. The main types of chemical modifications related to proteins are as follows: (1) Post-translational modifications (PTMs) refer to the chemical modification of a protein after translation<sup>4</sup>. The precursor of post-translationally modified protein often has no biological activity. Post-translational modification of a specific modified enzyme is usually required for it to become a functional mature protein and perform its specific biological functions<sup>5,6</sup>. (2) Chemical-derived modification of proteins refers to a kind of modification that introduces new groups or removes original/intrinsic groups in the protein side-chain, usually including spontaneous non-enzymatic modifications, as well as the modifications introduced by cross-linking agents and artificial agents. (3) Amino

<sup>1</sup>Department of Biochemistry and Molecular Biology, Beijing Key Laboratory of Gene Engineering Drug and Biotechnology, Beijing Normal University, Beijing, China. <sup>2</sup>Department of Laboratory Medicine, National Geriatrics Center, Beijing Hospital, Beijing, China. <sup>3</sup>These authors contributed equally: Yongtao Liu and Xuanzhen Pan. ✉email: gaoyouhe@bnu.edu.cn

acid substitution refers to the change in protein properties and functions caused by the substitution of amino acids in the protein side-chain by other kinds of amino acids. These changes are the types of modifications that significantly affect protein function.

Mass spectrometry can not only realize the acquisition of large-scale data and in-depth mining but also achieve the accurate determination of specific protein targeted modifications. With the continuous development of scientific instruments, ultrahigh-resolution and tandem mass spectrometry (MS/MS) provide more abundant information or data for proteomics and chemical modification research, which also facilitates the accurate identification of chemical modification sites in the protein chain. In the process of proteomic data analysis, it is usually necessary to search and compare the proteomic database of species, and high-resolution tandem mass spectrometry is the primary method for obtaining a large amount of protein modification information. When using the search engine to search the database, known types of protein chemical modification were usually set. This type of search method is called a restricted search, but it is difficult to identify a new type of modification with a type in the product that is unknown<sup>7</sup>. Therefore, comprehensive and non-limiting modification identification plays an important role in understanding all the chemical modification information contained in the sample proteome. Open-pFind is an open sequence library search algorithm that integrates the UniMod database, analyses, and processes the collected mass spectrum data through an open search to obtain the global chemical modification information of samples<sup>8–10</sup>.

Plasma is an important part of the internal environment and homeostasis, and it plays a role in transporting the substances needed for maintaining the life activities and the wastes generated off the body. Proteins are rich in variety and content in the plasma, and the components are easily affected by metabolism, physiology, and pathology. All metabolites or wastes in cells and tissues are transported and exchanged through the blood; the study of plasma proteomics may reflect the physiological state of the body at a specific stage. Chemical modification levels are another important research area of plasma proteomics. A comprehensive comparison of the changes in plasma proteome chemical modification levels will provide multiple dimensions of information for the study of physiological changes in the body<sup>11</sup>. At present, there are more than 1,500 kinds of chemical modifications in the UniMod<sup>12</sup>, PSI-MOD<sup>13</sup>, and RESID databases. There are many kinds of chemical modifications in the human plasma proteome, such as N-terminal acetylation, phosphorylation in the side chain, methylation, glycosylation, ubiquitination, and disulfide bonds between two chains. The plasma proteome can reflect the nuances defining the differences between age and ageing<sup>14</sup>. A study of the proteome changes in another body fluid, urine, demonstrated that this fluid can be analyzed to elucidate body ageing<sup>15</sup>. Even the common physiological process of hunger can be analyzed to characterize the urine proteome<sup>16</sup>. However, the comparison of the global proteome chemical modifications in two kinds of body fluids (plasma and urine) showed the differences in modifications between different types of samples<sup>17</sup>. Based on the importance of chemical modifications in the human plasma proteome, this study attempted to compare the differences in the global chemical modification levels of the plasma proteome at two different age groups, as determined by high-resolution tandem mass spectrometry combined with non-limiting modification identification (Open-pFind).

## Results

**Identification of total protein by using the bottom-up proteomics technology.** In the label-free proteomic analysis, 20 (10/10) samples were analyzed by LC-MS/MS. After retrieving data (.raw) based on pFind studio, the analysis results can be browsed and exported in pBuild studio. The 120-min liquid chromatogram gradient was analyzed, and 628–781 kinds of proteins (average 724) and 7,526–11,464 kinds of peptides (average 10,238) were identified in the plasma samples without high-abundance protein removal. About the raw data of pFind, we had submitted to iProX Datasets (<https://www.iprox.org/page/HMV006.html>) under the Project ID: IPX0002313000. And the results of pFind were showed in Supplementary Table 1.

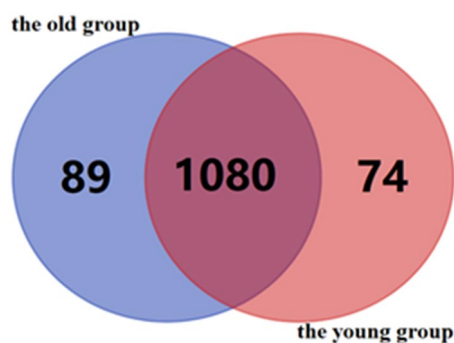
**Comparison of post-translational modifications of the plasma proteome between the young and old groups.** A total of 1,169 modifications (including low abundance modifications, that is, each modification is identified at least once) were identified in the plasma samples of the old group, and 1,154 modifications (including low abundance modifications) were identified in the plasma samples of the young group. Eighty-eight percent of the modifications are of the same type, and the remaining 12% are related (Fig. 1).

Among the 163 unique protein chemical modifications in the two groups, 158 of them had less than 50% repeatability in each group, and only 5 of them met the conditions of more than 50% reproducibility. They are four kinds of modifications in the old group: Arg → Phe [R], Arg → Trp [R], Xlink\_DMP [K], DTT\_ST [S] and bisANS-sulphonate [T] modifications in the young group. In general, the unique chemical modification types are the molecular modifications with low identification quantity and poor reproducibility in the group.

Non-low-abundance chemical modifications (i.e., each modification is identified 10 times or more in the sample) are counted, and the identification coverage in each group is required to be greater than 50%. There are 120 kinds of chemical modifications in 1,080 kinds that meet this condition. Unsupervised cluster analysis can roughly distinguish young group and old group samples, but 40% of young group and old group samples are clustered into one group, which is not classified as being the same as other samples in the group. Figure 2 shows the unsupervised clustering results of specific samples. The screening conditions for different chemical modifications are as follows: the p-value between two groups is less than 0.05, and the mean value of the number of chemical modifications of each sample in the group is calculated by the normalization of the total number of chemical modification spectra identified. There are four modifications with a multiple of change greater than 1.5 between groups: 2-succinyl [C], Lys → Thr [k], phosphor [C], and carbamyl [k]. Table 1 shows the details.

Order	No.*	Name	Type	p-value	Old avg	Young avg	Old/young
1	957	2-Succinyl[C]	Chemical derivative	0.001	0.00091	0.00046	1.98↑
2	594	Lys → Thr[K]	AA substitution	0.003	0.00107	0.00065	1.65↑
3	21	Phospho[C]	PTM	0.013	0.00087	0.000548	1.59↑
4	1,175	Pro → Trp[P]	AA substitution	0.042	0.00091	0.00061	1.49↑
5	21	Phospho[S]	PTM	0.027	0.00105	0.00077	1.36↑
6	64	Succinyl[AnyN-term]	PTM	0.032	0.00414	0.00317	1.31↑
7	23	Dehydrated[S]	PTM	0.022	0.0041	0.00446	1.09↓
8	562	Glu- > Asp[E]	AA substitution	0.29	0.0015	0.00184	1.22↓
9	5	Carbamyl[K]	Multiple	0.049	0.01858	0.03081	1.67↓

**Table 1.** Differential molecular modification information of plasma proteome between the old and young group. The number is the database entry number, from <https://www.unimod.org/>; AA means amino acid.



**Figure 1.** Statistics of the global plasma proteome chemical modifications in the old and young groups. The old and young groups have 1,080 of the same type of chemical modification, of which 89 are unique to the old group, and 74 are unique to the young group. The above entire modifications include all low-abundance modifications.

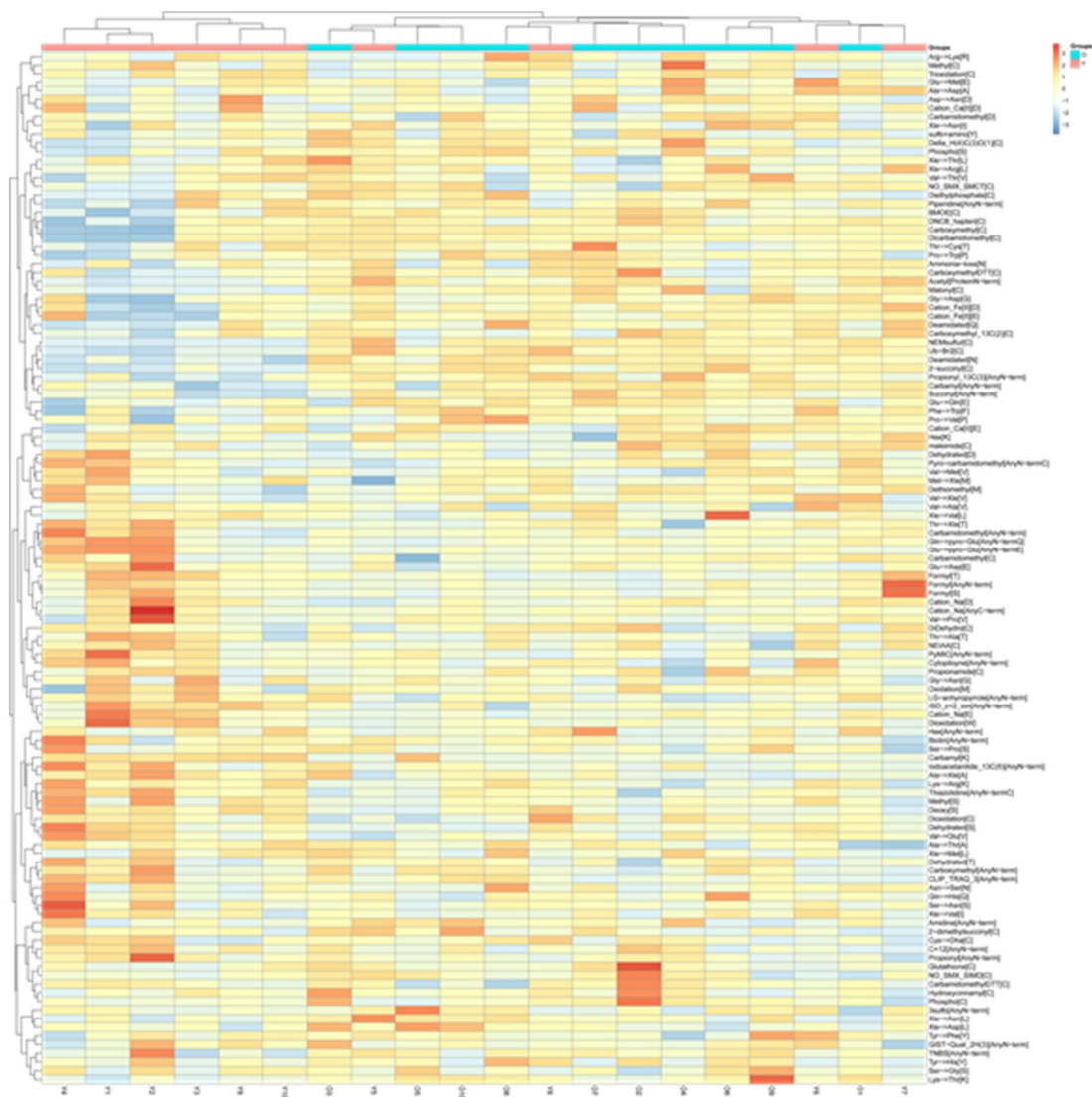
**Random grouping test.** Four different modifications satisfying different screening conditions (p-value less than 0.05, fold change greater than 1.5) were performed by random grouping tests to verify the false-positive rate of each modification. Ten young group datasets and 10 old group datasets were randomly divided into two groups with 10 samples in each group and 92,378 different combinations ( $\frac{1}{2} C_{20}^{10}$ ). Each combination was statistically analyzed by the same difference screening conditions. After extensive calculation and statistical analysis, four different modified random combinations were obtained. See Table 2 for details.

Through the random grouping test, it is found that the randomness of the four kinds of different modifications is approximately 5%, and the reliability is more than 90%. It is shown that the difference between the four kinds of modifications in different age groups is less likely to be generated at random. About the random grouping test details, the Supplementary 1–2 will show.

## Discussion

The entire life cycle of proteins, extending from translation assembly to final degradation, involves many toxic environments. These environments modify protein molecules to change the structure and function of the protein. This process is also called molecular ageing<sup>18</sup>. These reactions are mainly caused by the non-enzymatic binding of active small molecules on proteins with functional groups. The modification of proteins can be divided into reversible and irreversible modifications. Through a non-limiting modification search, several significantly different protein chemical modifications were found in the plasma of the two different age groups. These modifications include succinamide modification of cysteine, the substitution of lysine residues by threonine residues, phosphorylation of cysteine, the substitution of proline residues by tryptophan residues, phosphorylation of serine, N-terminal modification of succinamide, dehydration of serine, the substitution of glutamic acid residues by aspartic acid residues, and carbamylation of lysine. Among these changes, succinylation, phosphorylation of cysteine, and L-lysine replacement with threonine were significantly higher in the older group than in the younger group, and carbamylation of lysine was higher in the younger group than in the older group.

Cysteine is one of the amino acids with a lower abundance in the protein. The statistics of protein residues show that the average frequency of cysteine in eukaryotes is approximately 2%, and 70% of the reduced sulfhydryl source protein is present in the in vivo environment<sup>19</sup>. Studies have shown that the abundance of cysteine in protein is affected by the function of the protein. The high activity of the sulfhydryl group makes facilitates its participation in many chemical modifications. In contrast, this group also determines the distribution and topological properties of cysteine residues in protein structure. The sulfhydryl group is also an important group to form a disulfide bond and maintain protein structure<sup>20</sup>.



**Figure 2.** Unsupervised cluster analysis of overall plasma proteome molecular modifications in the young and old groups. The red mark (Y) is the sample of the young group, and the blue mark (O) is the sample of the old group. The abscissa is the unsupervised clustering and sample-specific information, and the ordinate is the specific molecular modification name.

No	Name	Total times in random groups	Frequency of occurrence	Confidence level%
1	2-succinyl[C]	4,857	0.0525	94.74
2	Lys → Thr[K]	1,135	0.0123	98.77
3	Phospho[C]	2,126	0.0230	97.70
4	Carbamyl[K]	4,664	0.0505	94.95

**Table 2.** Random grouping calculation results of 4 different chemical modifications in the plasma proteome.

The succinylation and phosphorylation of cysteine in this study involve the participation of the sulfhydryl group, which may affect the formation of the disulfide bond. Fumaric acid was added to the dissociative sulfhydryl sites of some Cys residues in proteins by a Michael addition reaction to form S-(2-succinic acid) cysteine<sup>21</sup>. This modification was initially detected in plasma proteins (including albumin) and formed by irreversible reactions. This modification has been reported in diabetes, obesity, fumaric acid hydratase-related diseases, and the model of RIE's syndrome. At the same time, it was also found that the content of succinic acid protein increased in mouse 3T3-L1 adipocytes cultured in high glucose medium (30 mm, while the physiological level was 5 mm), as well as in rats treated with streptozotocin<sup>21–24</sup>. It has been reported that an excess of nutrients (sugars) will lead to an increase in ATP: ADP, NADH: NAD<sup>+</sup> and mitochondrial membrane potential, while

an increase in NADH: NAD<sup>+</sup> will inhibit oxidative phosphorylation, resulting in the continuous accumulation of mitochondrial intermediates (including fumaric acid), leading to an increase in protein succinylation<sup>21</sup>. The accumulation of succinate protein is also caused by a decrease in fumarate hydratase activity<sup>25,26</sup>. Fumaric acid hydratase catalyzes the reversal of fumaric acid to malic acid in the tricarboxylic acid cycle. It is important to note that loss of function and mutations in fumaric hydratase is known to predispose affected individuals to multiple skin diseases and uterine leiomyomas, as well as hereditary leiomyomas and renal cell carcinoma (HLRCC)<sup>19</sup>.

Although the exact role of succinic acid-modified cysteine and other related proteins has not been fully elucidated, it is related to the cancer response related to fumarate hydratase<sup>27–29</sup>. The increase in protein succinic acid was also described in the brain stem of *ndufs4* knockout mice (a model of Leigh syndrome)<sup>30</sup>, indicating that this type of protein chemical modification has a potential role in the pathogenesis of this mitochondrial disease. Park et al. found that in the detected protein succinylation sites, 16 succinylation sites appeared in the cofactor binding area or enzyme catalytic area, and 74 succinylation sites existed around the enzyme active site<sup>31</sup>. Baynes et al. found that cysteine succinyl modification in human skin collagen increased with age<sup>22,32,33</sup>. Cysteine phosphorylation was recently found in prokaryotic and eukaryotic systems and is believed to play a key role in signal transduction and regulation of cellular responses. Due to the low chemical stability of thiophosphates in peptides, the *in vivo* phosphorylation of cysteine side chains is rarely studied<sup>33,34</sup>. Phosphorylation of cysteine is an important function of cysteine-dependent protein phosphatase (CDP), which belongs to a subfamily of protein tyrosine phosphatases (PTPs) and catalyzes the hydrolysis of phosphate ester bonds through the formation of phosphate cysteine intermediates. Studies have shown that this reversible post-translational regulation (PTM) is crucial in regulating the expression of virulence determinants and bacterial resistance to antibiotics. Besides, it has also been shown that phosphorylation of cysteine, previously considered a rare modification, may be more common in nature and may play an important role in the biological regulation of various organisms<sup>35–37</sup>.

The replacement of lysine with threonine changed the acid–base properties of the protein, which affected the activity and function of the protein. It has long been observed that the substitution of lysine residues with threonine residues on hemoglobin will reduce its oxygen affinity<sup>38</sup>. Sun WY et al. found that the mutation of nucleotide 20,040 in exon 14, with Thr replacing Lys at amino acid 556, would reduce the pro-coagulative activity of prothrombin by 50%<sup>39</sup>. The decrease in oxygen affinity, including some other recessions belonging to coagulation activity and other physiological conditions, also reflects the slow metabolism of the body, which may be the earlier manifestation of ageing. Also, it was found that human and mouse embryonic stem cells need specific amino acids to proliferate. MES cells need threonine (Thr) metabolism to complete epigenetic histone modification. Thr is converted to glycine and acetyl coenzyme A, and glycine metabolism specifically regulates the trimethylation of lysine (Lys) residues in histone H3 (H3K4me3)<sup>40</sup>. Besides, we also found that the modification of L-lysine carbamylation in the old group was lower than that in the young group. Carbamylation is an irreversible non-enzyme modification process. The process is the side chain reaction between the decomposition product of urea and the N-terminus of protein or lysine residue, which was previously reported to be related to the ageing of proteins<sup>41</sup>. L-lysine carbamylation can promote the coordination interaction of metal ions to specific enzyme activities. Some studies have pointed out that the amount of carbamylation in the plasma of patients with increased urea levels (such as nephrotic patients) is significantly increased<sup>42</sup>.

Ageing is an inevitable and spontaneous process undergone by the organism over time. Ageing is a complex natural phenomenon that is manifested by the degeneration of structure, the decline of function, and the recession of adaptability and resistance. Ageing is one of the largest known risk factors for most human diseases: approximately two-thirds of the world's 150,000 people die every day from ageing-related causes. At present, the cause of ageing has not been determined. The current mainstream theory explaining ageing is damage theory, and DNA damage is considered as the common foundation of cancer and ageing. Some people think that the internal cause of DNA damage is the most important driving force of ageing. According to the waste accumulation theory, the accumulation of waste in cells may interfere with metabolism. For example, a waste called lipofuscin is formed by a complex reaction of fat and protein combining in cells. These wastes accumulate in cells in the form of small particles, and their size will increase with age. Plasma protein can reflect changes in the body during the process of ageing. At the same time, the accumulation of biological macromolecules whose structure is destroyed or even inactivated may lead to the gradual failure of the biological body and system, which is also considered to be the concept of ageing.

Through the research described above, we found several types of chemical modifications and replacement of proteins in different age groups. These chemical modifications change the structure and properties of proteins and then affect the function of proteins. It is speculated that the gradual accumulation of some kinds of harmful and irreversible protein modifications in the plasma of the elderly may reflect the ageing process of the body. The accumulation of harmful protein modifications may be one of the reasons why the old are more likely than the young to suffer from ageing-related diseases, such as metabolic diseases, cardiovascular and cerebrovascular diseases, and tumor risks.

## Materials and methods

**Sample collection.** The plasma samples of 20 candidates who had medical examinations with passing medical tests were collected from the clinical laboratory of Beijing Hospital, and all of the samples had been discarded from the clinical laboratory. All candidates fully understood and signed the informed consent. The samples were divided into two groups according to age: a young group and an old group. The samples were randomly selected from the clinical laboratory samples, and there were no restrictions or requirements on the diet, drugs, and other factors of the blood sampling donors. The study was approved by the Beijing Hospital and the ethics committee of Beijing Normal University and all experiments were performed under relevant guidelines and regulations. This experiment provided volunteers with detailed information about the study, including

Sample type	Plasma in the young group (n = 10)	Urine in the old group (n = 10)
<b>Gender</b>		
Male	2	3
Female	8	7
<b>Age</b>		
Median	31.8 ± 4.19	59.3 ± 3.46
Average	26–37	56–66

**Table 3.** Statistical data of plasma samples.

its purpose, method, and process, and kept the personal information of volunteers strictly confidential. Only age and gender information are mentioned for discarded samples from the laboratory. Table 3 shows the sample information. Raw data and results of pFind were submitted to iProX Datasets (<https://www.iprox.org/page/HMV006.html>) under the Project ID: IPX0002313000.

**Protein sample preparation and trypsin enzymolysis.** The plasma was centrifuged after whole blood anticoagulant treatment. The plasma sample (n = 20) was diluted 40 times with Milli-Q water, and then 100 µL was taken for subsequent experiments. A 20 mmol/L dithiothreitol (DTT) was used to react with the sample at 37 °C for 1 h to denature the disulfide bond in the protein structure, and then 55 mmol/L iodoacetamide (IAM) was added and reacted in the dark for 30 min to alkylate the disulfide bond sites. The supernatant was precipitated with three volumes of precooled acetone at – 20 °C for 2 h and then centrifuged at 4 °C and 12,000×g for 30 min to obtain protein precipitation. The precipitate was then resuspended in an appropriate amount of protein lysis buffer solution (8 mol/L urea, 2 mol/L thiourea, 25 mmol/L DTT, and 50 mmol/L Tris). The concentration of protein extract was measured by Bradford analysis. Using filter-assisted sample preparation (FASP), trypsin gold (mass spec grade, Promega, Fitchburg, WI, USA) was used to hydrolyze 100 µg protein at a ratio of 50:1 for each sample. The dried peptides were sealed at –80 °C after drying by a vacuum centrifugal concentrator.

**Liquid chromatography-tandem mass spectrometry (LC–MS/MS) analysis.** Before analysis, the dried polypeptide samples were dissolved in 0.1% formic acid solution, and the final concentration was controlled at 0.1 µg/µL. Each sample was analyzed according to 1 µg polypeptide quality: a Thermo Easy-nlc1200 chromatographic system was loaded on the precolumn and the analytical column. Proteomic data were collected by the Orbitrap Fusion Lumos mass spectrometry system (Thermo Fisher Scientific Bremen, Germany). Liquid chromatography: pre-column: 75 µm × 2 cm, nanoviper C18, 2 µm, 100 Å; analytical column: 50 µm × 15 cm, nanoviper C18, 2 µm, 100 Å; injection volume: 10 µL, flow rate: 250 nL/min, mobile phase: phase A: 100% mass spectrometry grade water (Fisher Scientific, span)/1% formic acid (Fisher Scientific), phase B: 80% acetonitrile (USA)/20% water/1% formic acid, 120 min gradient washing off: 0 min, 3% B phase; 0–3 min, 8% B phase; 3–93 min, 22% B phase; 93–113 min, 35% B phase; 113–120 min, B phase; mass spectrometry, ion source: spray voltage: 2.0kV, capillary temperature: 320 °C, resolution setting: first-order (Orbitrap) 120,000 @m/z 200, second-order (Orbitrap) 30,000 (Orbitrap) @m/z 200, parent ion scanning range: m/z 350–1,350; sub ion scanning range: start from m/z 110, MS1 AGC: 4E5, charge range: 2–7, ion implantation time: 50 ms, MS 2 AGC: 1E5, ion implantation time: 50 ms, ion screening window: m/z 2.0, fragmentation mode: HCD, energy NCE 32, data dependent MS/MS: Top 20, dynamic exclusion time: 15 s, internal calibration mass number: 44 5.12003.

**Database search.** The pFind Studio software (version 3.1.3, Institute of Computing Technology, Chinese Academy of Sciences) was used to analyze the LC–MS/MS data with label-free quantification. The target retrieval database is from the Homo Sapiens database downloaded from UniProt (updated to October 2018). Raw files generated by the Orbitrap Fusion were searched directly using a ±20-ppm precursor mass tolerance and a ±20-ppm fragment mass tolerance. At the time of retrieval, the instrument type is HCD-FTMS, the full specificity of an enzyme is trypsin, and there are at most two missing sites, peptides with at least six amino acids were retained. Open-search is selected. Screening conditions: the FDRs were estimated by the program from the number and quality of spectral matches to the decoy database; for all data sets, the FDRs at spectrum, peptide, and protein level were < 1%, and the Q-value at the protein level is less than 1%. Data are analyzed using both forward and reverse database retrieval strategies.

Received: 22 April 2020; Accepted: 17 August 2020  
Published online: 14 September 2020

## References

1. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
2. Sakamoto, S. & Hamachi, I. Recent progress in chemical modification of proteins. *Anal. Sci. Int. J. Jpn. Soc. Anal. Chem.* **35**, 5–27 (2019).

3. Wang, C. *et al.* Dynamic modifications of biomacromolecules: mechanism and chemical interventions. *Sci. China Life Sci.* **62**, 1459–1471 (2019).
4. Minguez, P. *et al.* Deciphering a global network of functionally associated post-translational modifications. *Mol. Syst. Biol.* **8**, 599 (2012).
5. Deribe, Y. L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* **17**, 666–672 (2010).
6. Pieroni, L. *et al.* Enrichments of post-translational modifications in proteomic studies. *J. Sep. Sci.* **43**, 313–336 (2020).
7. An, Z. *et al.* PTMiner: localization and quality control of protein modifications detected in an open search and its application to comprehensive post-translational modification characterization in human proteome. *Mol. Cell. Proteomics MCP* **18**, 391–405 (2019).
8. Chi, H. *et al.* pFind-Alioth: a novel unrestricted database search algorithm to improve the interpretation of high-resolution MS/MS data. *J. Proteomics* **125**, 89–97 (2015).
9. Li, D. *et al.* pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. *Bioinformatics (Oxford, England)* **21**, 3049–3050 (2005).
10. Wang, L.-H. *et al.* pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **21**, 2985–2991 (2007).
11. Goto, T., Kudo, Y., Lee, S. H. & Oe, T. Efficient strategy for screening chemical modifications on human serum albumin: use of LC/MS/MS and differential analysis software. *Bunseki Kagaku* **64**, 653–659 (2015).
12. Creasy, D. M. & Cottrell, J. S. Unimod: protein modifications for mass spectrometry. *Proteomics* **4**, 1534–1536 (2004).
13. Montecchi-Palazzi, L. *et al.* The PSI-MOD community standard for representation of protein modification data. *Nat. Biotechnol.* **26**, 864–866 (2008).
14. Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
15. Li, X. & Gao, Y. Potential urinary aging markers of 20-month-old rats. *PeerJ* **4**, e2058 (2016).
16. Yuan, Y. & Gao, Y. Urinary protein changes in a rat starvation model. *MOJPB* **6**, 323–331 (2017).
17. Zhao, M., Pan, X., Gao, Y. & Liu, Y. Why are there proteins in the urine of healthy people?. *Sci. Sin. Vitae* **50**, 338–348 (2020).
18. Gillery, P. & Jaisson, S. Usefulness of non-enzymatic post-translational modification derived products (PTMDPs) as biomarkers of chronic diseases. *J. Proteomics* **92**, 228–238 (2013).
19. Péér, I. *et al.* Proteomic signatures: amino acid and oligopeptide compositions differentiate among phyla. *Proteins* **54**, 20–40 (2004).
20. Gunnoo, S. B. & Madder, A. Chemical protein modification through cysteine. *Chembiochem Eur. J. Chem. Biol.* **17**, 529–553 (2016).
21. Alderson, N. L. *et al.* S-(2-Succinyl)cysteine: a novel chemical modification of tissue proteins by a Krebs cycle intermediate. *Arch. Biochem. Biophys.* **450**, 1–8 (2006).
22. Blatnik, M., Thorpe, S. R. & Baynes, J. W. Succination of proteins by fumarate: mechanism of inactivation of glyceraldehyde-3-phosphate dehydrogenase in diabetes. *Ann. N. Y. Acad. Sci.* **1126**, 272–275 (2008).
23. Thomas, S. A., Storey, K. B., Baynes, J. W. & Frizzell, N. Tissue distribution of S-(2-succino)cysteine (2SC), a biomarker of mitochondrial stress in obesity and diabetes. *Obesity (Silver Spring Md.)* **20**, 263–269 (2012).
24. Frizzell, N. *et al.* Succination of thiol groups in adipose tissue proteins in diabetes: succination inhibits polymerization and secretion of adiponectin. *J. Biol. Chem.* **284**, 25772–25781 (2009).
25. Frizzell, N., Thomas, S. A., Carson, J. A. & Baynes, J. W. Mitochondrial stress causes increased succination of proteins in adipocytes in response to glucotoxicity. *Biochem. J.* **445**, 247–254 (2012).
26. Isaacs, J. S. *et al.* HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability. *Cancer Cell* **8**, 143–153 (2005).
27. Frezza, C., Pollard, P. J. & Gottlieb, E. Inborn and acquired metabolic defects in cancer. *J. Mol. Med. (Berlin, Germany)* **89**, 213–220 (2011).
28. Yang, M., Soga, T., Pollard, P. J. & Adam, J. The emerging role of fumarate as an oncometabolite. *Front. Oncol.* **2**, 85 (2012).
29. Zheng, L. *et al.* Fumarate induces redox-dependent senescence by modifying glutathione metabolism. *Nat. Commun.* **6**, 6001 (2015).
30. Piroli, G. G. *et al.* Succination is increased on select proteins in the brainstem of the NADH dehydrogenase (ubiquinone) Fe-S protein 4 (Ndufs4) Knockout Mouse, a model of leigh syndrome. *Mol. Cellular Proteomics MCP* **15**, 445–461 (2016).
31. Piroli, G. G. *et al.* Identification of protein succination as a novel modification of tubulin. *Biochem. J.* **462**, 231–245 (2014).
32. Frizzell, N., Lima, M. & Baynes, J. W. Succination of proteins in diabetes. *Free Radic. Res.* **45**, 101–109 (2011).
33. Zu, X.-L., Besant, P. G., Imhof, A. & Attwood, P. V. Mass spectrometric analysis of protein histidine phosphorylation. *Amino Acids* **32**, 347–357 (2007).
34. Attwood, P. V., Piggott, M. J., Zu, X. L. & Besant, P. G. Focus on phosphohistidine. *Amino Acids* **32**, 145–156 (2007).
35. Pengo, N., Agrotis, A., Prak, K., Jones, J. & Ketteler, R. A reversible phospho-switch mediated by ULK1 regulates the activity of autophagy protease ATG4B. *Nat. Commun.* **8**, 294 (2017).
36. Sun, F. *et al.* Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. *Proc. Natl. Acad. Sci. USA* **109**, 15461–15466 (2012).
37. Kalantari, A., Derouiche, A., Shi, L. & Mijakovic, I. Serine/threonine/tyrosine phosphorylation regulates DNA binding of bacterial transcriptional regulators. *Microbiology (Reading, England)* **161**, 1720–1729 (2015).
38. Shih, D. T. *et al.* Hemoglobin Chico beta 66(E10)Lys--Thr: a new variant with decreased oxygen affinity. *Hemoglobin* **11**, 453–464 (1987).
39. Sun, W. Y., Smirnow, D., Jenkins, M. L. & Degen, S. J. Prothrombin scranton: substitution of an amino acid residue involved in the binding of Na<sup>+</sup> (LYS-556 to THR) leads to dysprothrombinemia. *Thromb. Haemost.* **85**, 651–654 (2001).
40. van Winkle, L. J. & Ryznar, R. One-carbon metabolism regulates embryonic stem cell fate through epigenetic DNA and histone modifications: implications for transgenerational metabolic disorders in adults. *Front. Cell Dev. Biol.* **7**, 300 (2019).
41. Badar, A., Arif, Z. & Alam, K. Role of carbamylated biomolecules in human diseases. *IUBMB Life* **70**, 267–275 (2018).
42. Dalle-Donne, I. *et al.* Protein carbonylation, cellular dysfunction, and disease progression. *J. Cell Mol. Med.* **10**, 389–406 (2006).

## Acknowledgements

We thank Dr. Hao Chi (Institute of Computing Technology, CAS) for help in using pFind, and thank Dr. Shisheng Wang (West China Hospital, Sichuan University) and Dr. Chengpin Shen (OmicSolution Co., Ltd) for giving some advice about data analysis and ‘Wu Kong’ platform (<https://www.omicsolution.org/wkomics/main/>) for relative cluster analysis.

## Author contributions

Y.L. designed and conducted the experiments, analyzed the data, prepared the figures, and wrote the paper. X.P. retrieved references, helped with writing the paper and did later period work. These authors contributed equally to this work, Y.L. and X.P. should be co-first authors. M.Z. provided samples and helped in conduct the experiments. Y.G. as corresponding author, provided the main ideas, designed experiments, and helped with writing the paper.

## Funding

The National Key Research and Development Program of China (2018YFC0910202, 2016YFC1306300), the Fundamental Research Funds for the Central Universities (2020KJZX002), Beijing Natural Science Foundation (7172076), Beijing Cooperative Construction Project (110651103), Beijing Normal University (11100704), Peking Union Medical College Hospital (2016-2.27).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-72196-z>.

**Correspondence** and requests for materials should be addressed to Y.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020