Supplementary Materials

Supplementary tables

Table S1 The datasets used in this protocol — sample preparation

Data type	**Non-labeling**	**^{14}N/^{15}N-labeling**	**SILAC-labeling**	**TMT-labeling**
RAW files	ID_data_HeLa_QE_HF_120min_rep1.raw ID_data_HeLa_QE_HF_120min_rep2.raw	Quant_15Ndata_Celegans_QE_105min.raw	Quant_SILACdata_HeLa_QE_110min.raw	Quant_TMTdata_RMS_QE_105min.raw
Source sample	Unlabeled total proteins from a human cell lysate	^{14}N/^{15}N-labeled *C. elegans* lysates	Phosphopeptides enriched from SILAC-labeled human cell lysates	TMT-labeled peptides from human cell lysates
Light/Heavy mixing ratio	Not applied	1:1 ^{14}N/^{15}N	1:1 SILAC K0R0/K8R10	1:1:1:1:1:1 of MS2 reporter ions 126/127/128/129/130/131
Reduction	TCEP			
Alkylation	IAA			
Enzyme digestion	Trypsin			

Table S2 The datasets used in this protocol — LC-MS/MS

Data type	**Non-labeling**	**$^{14}N/^{15}N$-labeling**	**SILAC-labeling**	**TMT-labeling**
RAW files	ID_data_HeLa_QE_ HF_120min_rep1.raw ID_data_HeLa_QE_ HF_120min_rep2.raw	Quant_15Ndata_ Celegans_QE_ 105min.raw	Quant_SILACdata_ HeLa_QE_ 110min.raw	Quant_TMTdata_ RMS_QE_ 105min. raw
Effective LC gradient				
Running time	2–100 min	2–85 min	0–80 min	2–58 min
Gradient (% of acetonitrile)	7–30	3–27	3–27	5–28
MS data acquisition				
Instrument type	Q Exactive HF	Q-Exactive	Q-Exactive	Q-Exactive
MS AGC target	3×10^6	3×10^6	1×10^6	3×10^6
MS Resolution	60,000	70,000	70,000	70,000
MS Maximum IT	60 ms	60 ms	60 ms	60 ms
MS1 Scan range (m/z)	200–2000	300–2000	400–2000	300–2000
MS2 AGC target	1×10^5	1×10^5	5×10^4	1×10^5
MS2 Resolution	15,000	17,500	17,500	17,500
MS2 Maximum IT	60 ms	60 ms	250 ms	60 ms
Isolation Window	2 m/z	2 m/z	2 m/z	2 m/z
NCE (HCD)	27	27	27	32
Intensity threshold	3×10^4	5×10^4	4×10^3	5×10^4
Dynamic exclusion	45 s	30 s	60 s	30 s

Table S3 Annotation of pFind parameters in the *MS Data* panel

Annotation of pFind parameters in the MS Data panel	
MS data format	pFind 3 accommodates four data formats (MGF, RAW, WIFF, mzML)
MS instrument	pFind 3 accommodates four MS instrument data types HCD-FTMS, HCD-ITMS, CID-FTMS, CID-ITMS
Data file list	List of a user's MS data files to be searched
Data extraction	
Place of decimal	
M/*Z*: 5	*M*/*Z* values will be recorded up to the fifth decimal point
Intensity:1	Intensity values will be recorded up to the first decimal point
Precursor Score	
Model	Normal is the right setting in most cases. 15N is only for data with ^{15}N-labeling
Threshold (–0.5)	Cutoff score for pParse to export precursor ions, the smaller the value, the more exported precursors. The default value is –0.5
Mixture Spectra	Leave it checked to allow a mixture spectrum (2 or more precursors fragmented together) to be searched under different precursors masses. It increases the number of identified peptides by about 10% for cell lysate samples
Output files	MS1/MS2, MGF will be exported as options. These files may be useful for manual checking or file sharing between database search engines

Table S4 Annotation of pFind parameters in the *Identification* panel

Annotation of pFind parameters in Identification panel	
Database search	
Database	A protein sequence database, to be selected by user from the pFind Database collection
Enzyme	Select digestion enzyme used in sample preparation for *in silico* digestion of protein sequences
Digestion specificity	Full-Specific: enzyme specificity is to be observed strictly on both ends of an *in silico* digested peptide. Semi-specific: enzyme specificity is to be observed on either or both ends of a peptide, that is, at least one end results from a specific cut. Non-specific: no requirement
Missed cleavages	The maximum number of missed enzyme digestion site allowed in identified peptides
Precursor Tolerance	Maximum mass deviation of an observed precursor ion from the theoretical mass of a peptide.
Fragment Tolerance	Maximum mass deviation of an observed fragment ion from the theoretical mass
Open search	If checked, an open-pFind search will be performed; otherwise, a restricted search that considers only the user specified modifications will be performed
Add modification	
Fixed	Fixed modification on peptides
Variable	Variable modification on peptides
Display All	Leave it unchecked unless you wish to see everything in the modification list. The current number of entries is 2594
Result Filter	
Show Spectra or Peptides	
FDR $\leqslant$ 1% (peptide/spectra)	FDR at the peptide or spectrum level. The peptide level is set by default, which is more stringent than the alternative
Peptide mass	Min and max peptide mass used to filter the search result
Peptide Length	Min and max peptide length used to filter the search result
Show proteins	
Number of peptides $\geqslant$ 1	The min number of peptides required for a protein identification. This is usually set to 1, Changing this value to 2 will greatly reduce false protein identifications (Chi *et al.* 2018)
FDR $\leqslant$ 1%	FDR at the protein level

Table S5 Annotation of the *Summary* panel in pBuild

Annotation of the *Summary* panel in pBuild	
Peptide Level	
Spectra number	Number of identified spectra
Scans number	number of MS2 scans that have yielded peptide identification results
Peptide number	Number of identified peptides
Sequence number	Number of identified peptide sequences
Protein number	Number of identified proteins
Protein group number	Number of identified protein groups
Decoy spectra number	Number of MS2 spectra that happen to be matched with sequences from the decoy database
Decoy peptide number	Number of identified peptides from the decoy database
Decoy protein number	Number of identified proteins from the decoy database
Decoy protein group number	Number of identified protein groups from the decoy database
Cleavage	
Specific	Among all the identified peptides, the percentage of those that could result from specific cutting by the enzyme of choice at both ends
C-term specific	The percentage of peptides whose C-termini, and only the C-termini, could result from specific cutting by the enzyme of choice
N-term specific	The percentage of peptides whose N-termini, and only the N-termini, could result from specific cutting by the enzyme of choice
Non-specific	The percentage of peptides that do not fit the specificity of the enzyme at both ends
Quantitation	
NaN number (no contaminants)	Not-a-Number, an invalid quantitation ratio of $^{15}N/^{14}N$ or $^{13}C/^{12}C$
Mean	The mean value of all quantitation ratios; the calculation of each ratio is triggered by an identified MS2 spectrum
Median	The median value of all quantitation ratios
Standard Deviation	The standard deviation of all quantitation ratios
Modification	The ten most abundant modifications identified are displayed
Missed Cleavage	The percentages of identified peptides that contain from zero, one, two, or more missed cleavage sites are displayed
Mixed Spectra	The percentages of MS2 spectra that may result from isolation and fragmentation of one, two, three, or more precursor ions
Charge	The charge state distribution of identified peptides
MassError	
Precursor mass error (mean)	For all PSMs identified, the mean of the mass differences between an observed precursor ion and the theoretical mass of the peptide identified from the precursor ion's MS2 spectrum, in ppm
Precursor mass error (std)	The standard deviation of the precursor mass error among all identified peptides

ID Rate	The percentage of identified MS2 scans, for each individual data file searched or the whole	
Parameter		
Thread number	2	The number of CPU threads used for data analysis
ms tolerance	20 ppm	Maximal allowed mass deviation of a precursor ion from the theoretical mass of a peptide
msms tolerance	20 ppm	Maximal allowed mass difference between an experimental fragment ion and a theoretical fragment ion generated *in silico*
Open search	True	Type of search performed by pFind 3. 'True' means that an Open-pFind search is performed, which is the default setting
Input format	raw	The format of the input data is .raw
Fix modification	NULL	User-specified fixed modification(s) on peptides. None in this case
Variable modification	NULL	User-specified variable modification(s) on peptides. None in this case
Enzyme	Trypsin KR_C	Which protease is used to digest samples? Trypsin in this case
Max missing cleavage number	3	The maximum number of missed enzyme digestion sites allowed in identified peptides
Coelute	True	True: export all potentially co-eluting precursors in a mixed spectrum; False: export from each MS2 the one precursor that has the best score in pParse
II info label	The type(s) of labeling specified in the experiment. The value of each labeling type can be changed in the configuration file named quant.ini	
Chrom tolerance	15 ppm	This is the maximal allowed mass deviation specified for pQuant when it looks for the signal of a precursor ion to build a chromatogram
Label efficiency	99.0%	The labeling efficiency of a stable isotope labeling reagent; 99.0% is the default value
File		
aa path 1	Location of amino acids labeled for 'none'	
aa path 2	Location of amino acids labeled for 'heavy'	
Madification path	Location of the database of modifications	
Fasta path	Location of the protein sequence database	
Contaminant path	Location of the database of contaminant proteins	
Task path	Location of the parameter and the result files	
Raws		
Raw path 1	Location of raw data	

Table S6 Annotation of pFind parameters in the *Quantitation* panel

Annotation of pFind parameters in the *Quantitation* panel	
MS1 quantitation	
Type	pFind 3 accommodates five data types, Labeling-None, Labeling-15N, Labeling-SILAC *etc*., LabelFree, Report-Intensity
Multiplicity	The number of labeling reagents used in sample labeling, including the unlabeled one
Light Label	Usually this is 'None' for unlabeled, that is, labeled with naturally occurring light isotopes such as ^{1}H, ^{12}C, and ^{14}N
Heavy Label	The heavy stable isotope label, *e.g*. deuterium or ^{15}N
Medium label	For triple-SILAC labeled data: users choose 3 in 'multiplicity', the 'Light label', 'Medium label', 'Heavy label' will appear in the panell. One such triple SILAC labeling consists of Arg0Lys0 for the light label, Arg6($^{13}C_6$)Lys4(D_4) for the medium label, and Arg10($^{13}C_6{}^{15}N_4$)Lys8($^{13}C_6{}^{15}N_2$) for the heavy label
Advance panell	
Number of Scans Per Half CMTG: 200	The number of MS1 scans along half of the chromatogram peak. Generally, it is 100–200
Number of Holes in CMTG: 2	If the peptide (or reporter ion) intensity in two consecutive MS1 (or MS2) spectra drops below 10% of the maximal intensity of the chromatographic peak under construction, the chromatographic peak stops extending in that direction, the start/end point of a chromatographic peak is thus determined. In general, this parameter is set to 2
Calibration in ppm: 0	The systemic mass deviation in ppm of the peaks in MS1 scans. Enter zero if your MS instrument is well calibrated
Half Window Accuracy Peak Tolerance in ppm: 15	Similar to precursor tolerance in the identification panell. pQuant will find the experimental *m*/*z* of precursor to reconstruct the chromatographic peak. A value of 15 means pQuant will allow ± 15 ppm mass accuracy when it constructs a chromatographic peak
TYPE_SAME_START_END_BETWEEN_EVIDENCE (For 1:1 Mixed Samples /For10:1 or 1:10 Mixed Samples/ Independent)	For 1:1 mixed sample, the length of chromatograph of light and heavy isotope are the same. For the 10:1 or 1:10 mixed sample, the low abundance isotope has the shorter chromatograph
ELEMENT_ENRICHMENT_CALIBRIATION (none/15N/13C)	If 15N or 13C is entered here, pQuant will estimate the atomic enrichment ratio of ^{15}N or ^{13}C
MS2	Since quantitation here is based on MS1, make sure the option 'MS2 Quantitation' is unchecked. It is off by default
MS2 quantitation	

Method	pFind 3 accommodates five methods, iTRAQ-8plex, iTRAQ-6plex, TMT-6plex, TMT-10plex, pIDL
TMT/iTRAQ Quantitation	
Reporter ions MZ	The accurate m/z values of reporter ions formed by the labeling reagents in MS2 scans. MS2 quantification is based on their relative intensities
Advanced	
Fragment tolerance	The range of mass deviation of fragment ions
Peak range	*m/z* range of reporter ions for MS2 quantitation
PIF	PIF stands for precursor intensity fraction, the fraction of the intensity of the intended precursor ion among all the ions present in the isolation window. The higher the PIF, the more accurate the quantification
PSM FDR	PSM stands for peptide-spectrum match. PSM FDR is the FDR cutoff at the spectrum level. The smaller the FDR, the more accurate the identification result, and more false negatives, too
Protein FDR	FDR cutoff at the protein level. The smaller the FDR, the more accurate the identification result, and more false negatives
Correction matrix	The atomic enrichment ratio of 13C or 15N used in the tags. See reagents and kits guide for details
Run VSN	Use Variance Stabilizing Normalization (VSN) method to normalize the protein intensities
pIDL Quantitation	
Nterm-modification	The name given to the modification at the peptide N terminus
Mass (Nterm-modification)	The mass change brought about by the Nterm-modification
Cterm-modification	The name given to the modification at the peptide C terminus
Mass (Cterm-modification)	The mass change brought about by the Cterm-modification

Table S7 Annotation of the *Protein* panel in pBuild

Annotation of the *Protein* panel in pBuild	
#	Serial number
AC	Accession number or protein entry identifier
DE	Description of a protein
SQ Length	Protein sequence length
PSM Count	Total number of MS2 spectra matched to a peptide
Coverage	Sequence coverage of a protein
Score	The expected false discovery rate at the protein level
Group	The identifier of a protein group of which the protein in question is a member; it is the accession number of the leading member of this protein group
Flag	NULL: leading protein; Sameset/Subset: the protein in question is not the leading member of its protein group and relative to the leading protein, it is a sameset/subset protein. In other words, the peptides that could be assigned to this protein could all be assigned to the leading protein; the reverse can be said for sameset but not for subset
Ratio	Quantitation ratio

Table S8 Annotation of the header in the file ‘pFind.protein’

Annotation of the header in the file ‘pFind.protein’	
First row	
ID	Serial number
AC	Accession number or protein entry identifier
Score	Score of a protein from a SVM model
Q-value	The expected false discovery rate at the protein level
Coverage	Sequence coverage of the protein
No.Peptide	Total number of peptides identified in this protein
No.Sameset	The number of sameset proteins in this protein group. A protein is a "sameset protein" of another one if the exact same peptides identify them both. See annotation of ‘Flag’ in Supplementary Table S7
No.Subset	The number of subset proteins in this protein group. A protein is a "subset protein" of another one if the peptides identifying the former are a subset of those identifying the latter. See also annotation of ‘Flag’ in Supplementary Table S7
Have_Distinct_Pep	A value of ‘1’ means that the leading member of a protein group possesses at least ONE unique peptide not shared with any other proteins in the database searched.
Description	Description of a protein
Second row	
ID	Serial number of sameset and subset proteins
Sequence	Peptide sequence
Calc.MH+	Calculated theoretical mass of a peptide with one proton attached to it
Mass_Shift(Exp.-Calc.)	Mass deviation of a peptide, in ppm
Raw_Score	The preliminary score of a peptide-spectrum match
Final_Score	The final score of a peptide-spectrum match generated by a percolator-like scoring model
Modification	Modifications at different positions are separated by semicolon. For each modification, the number indicates the position of the modification in a peptide, with ‘0’ referring to the N terminus of a peptide and ‘1’ the first amino acid residue (usually the side chain of the first amino acid residue) and so on
Specificity	Enzyme specificity associated with the genesis of a peptide by protease digestion; 0: non-specific; 1: specific at the C terminus non-specific at the N terminus; 2: specific at the N terminus, non-specific at the C terminus; 3: specific at both termini
Proteins	Proteins from which the peptide may originate
Positions	The position of the first amino acid of a peptide in a protein sequence
Label	If there are k modifications in a peptide, then the value is a string with $k + 1$ numbers separated by k vertical bars ('|'). The first number denotes the labeling

	type of the peptide sequence, and the other k numbers indicate the labeling types of the modifications found on the peptide. The range of all of the $k + 1$ numbers is [0, q], where q is the number of different labeling types set in pFind. For example, in a SILAC labeling experiment with light: heavy at 1:1, the value of q is 2, and 0, 1, 2 denotes none (which is the same as light), light, and heavy, respectively. For triple SILAC labeling, these numbers are 1, 2, and 3 for the light, medium, and heavy label, respectively
Target/Decoy	The peptide is a target/decoy sequence
Miss.Clv.Sites	Numbers of missed cleavage sites in the peptide
Avg.Frag.Mass.Shift	Averaged mass deviation of the experimental fragmentation ions from the theoretical fragmentation ions predicted based on the best peptide-spectrum match, in ppm
File_Name	Title of the best scoring MS2 spectrum of the peptide. It contains the name of the MS data file, the scan number, and the charge state of the precursor
Charge	The charge state of a peptide with the best score
Spec_Num	Total number of spectra that identify this same peptide

Table S9 Annotation of the header in the file ‘pFind_protein_contrast_result.txt’

Annotation of the header in the file ‘pFind_protein_contrast_result.txt’	
First row	
Rep1	Name of a data file or sample to be compared
Rep2	Name of another data file or sample to be compared
Second row	
Protein	The name of an identified protein
Total_pep_num@pro (Total_unique_pep_ num@pro)	Total_pep_num@pro: the total number of peptides identified for a protein; Total_unique_pep_num@pro: the total number of unique peptides identified for a protein
Total_spec_num@pro (Total_unique_spec_ num@pro)	Total_spec_num@pro: the total number of PSMs identified for a protein; Total_unique_pep_num@pro: the total number of PSMs of the uniquely identified peptides for the protein in question
Have unique peptide?	Whether or not a protein has unique peptides identified, if yes, it is marked 'have unique peptide'; if not, 'not unique'
Third row	
Peptide	Peptide sequence
Modification	The modification on the identified peptide
Total_spec_num@pep	The total number of identified spectra of a peptide
best-score@pep	The best score in all PMS of the peptide
Is unique?	Whether or not a peptide can be mapped back to a single protein, *i.e.*, whether or not it is a unique peptide. If yes, it is marked 'is unique'; if not, 'not unique'
Shared proteins	If a peptide is not a unique peptide, *i.e.*, it can be mapped back to other proteins in addition to the one in question, then the other proteins are listed here

Table S10 Annotation of the header in the file ‘pFind_PTM_contrast_result.txt’

Annotation of the header in the file ‘pFind_PTM_contrast_result.txt’	
First row	
Rep1	Name of a data file or sample to be compared
Rep2	Name of another data file or sample to be compared
Second row	
Protein	The name of a protein found to carry a particular PTM
Total_site_num@pro	The total number of modification sites identified in a protein
Total_pep_num@pro	The total number of peptides found in a protein that carry a specified modification
Total_spec_num@pro	The total number of MS2 spectra supporting the identification of the peptides of a protein that carry a specified modification
Have unique peptide?	If any of the PTM sites found in a protein is associated with a uniquely identified peptide, it will be marked 'have unique peptide'; if not, ' no unique peptide'
Third row	
Site	List of identified modification sites in a protein
Total_pep_num@site (Total_unique_pep_ num@site)	Total_pep_num@site: the total number of the peptide with the modification identified in this site.; Total_unique_pep_num@site: the total number of unique peptide with the modification identified in this protein
Total_spec_num@site (Total_unique_spec_ num@site)	Total_spec_num@pro: the total PSM number of peptides with the modification identified in this site; Total_unique_pep_num@site: the total PSM number of unique peptides with the modification identified in this site
Best-score@site	The best score in all PMS of sites
Have unique peptide?	Whether or not the identified modification sites are found in one or more unique peptides of a protein. If yes, it is marked 'have unique peptide'; if not, 'not unique'
Fourth row	
Peptide	An identified peptide that supports the finding of a particular modification on a particular site of a protein
Modification	The modification on the identified peptide
Total_spec_num@pep	The total number of MS2 spectra supporting the identification of the peptide with the modification
Shared proteins	If the identified peptide may originate from other proteins in addition to the one in question, these other ones are listed here
Is unique?	If the identified peptide with a particular modification is a unique peptide, it is marked 'is unique'; if not, 'not unique'

FAQs

Q1: Can I use pFind 3 to analyze ETD, EThcD, or ETciD data?

ANSWER: Yes, pFind 3 can be hacked to search these data types. If you have such needs, please contact pfind@ict.ac.cn.

Q2: I am looking for MBR (match between run) for quantitation. Where is it?

ANSWER: Noted and added to the wish list for the next upgrade.

Q3: I wish to find novel modifications that may be outside the Unimod collection. Can pFind 3 perform a blind search and report unexpected amino acid modifications as mass shifts?

ANSWER: Yes, pFind 3 can perform a blind search. Please contact pfind@ict.ac.cn if you need this function.

Q4: I wonder why some of the *q*-values reported by "pFind.spectra" are greater than 1, for example, 512 or 1024? What is the meaning of a *q*-value being 512 or 1024?

ANSWER: If a *q*-value is 1024, it means that there is no identification for that spectrum. Likewise, a *q*-value of 512 denotes a low-confidence identification. Both are meant to be filtered out.

Q5: Does pFind 3 support quantitation based on 16-plex TMT labeling?

ANSWER: Not yet. At the moment, pFind 3 works with 6-plex or 10-plex TMT labeling.

Q6: How to open the result files like ".summary" or ".spectra"?

ANSWER: Use a text editor like "Notepad" or "Notepad++" software to open ".summary" file. Drag a ".spectra" file to an excel window to open it.

Q7: How to add a new modification?

ANSWER: Firstly, you must know the exact mass shift, the corresponding chemical formula and the modified amino acid, then follow the instruction below. Here we give an example (name: testMod; chemical formula: C(4)H(8)O(2); modified amino acid: K).

a. From the pFind menu bar, click "Options" and select "Meta Data Configuration" (Fig. S1).

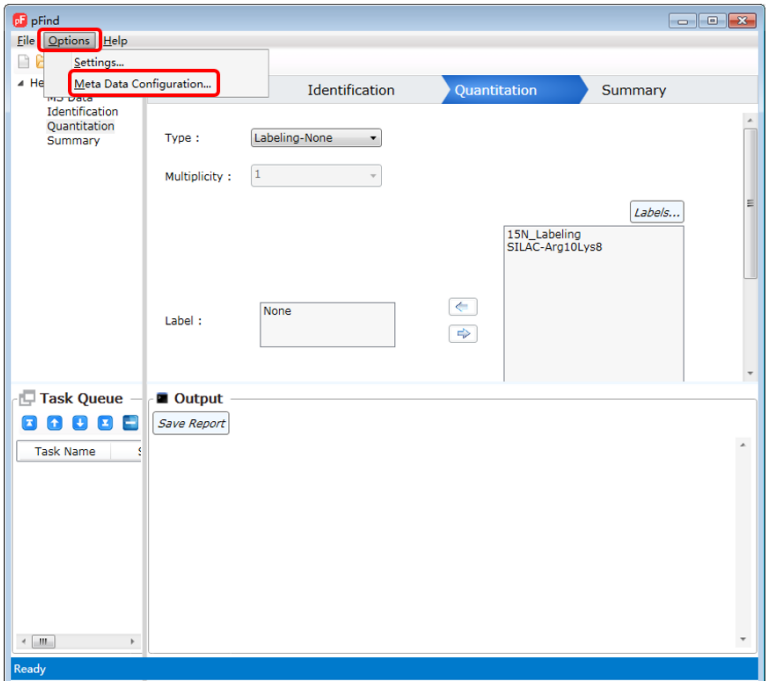

Fig. S1 Open "Option" and select "Meta Data Configuration"

b. In the "pConfig" panel that pops up, pick up the tab of "Modifications", and click "Add" (Fig. S2).

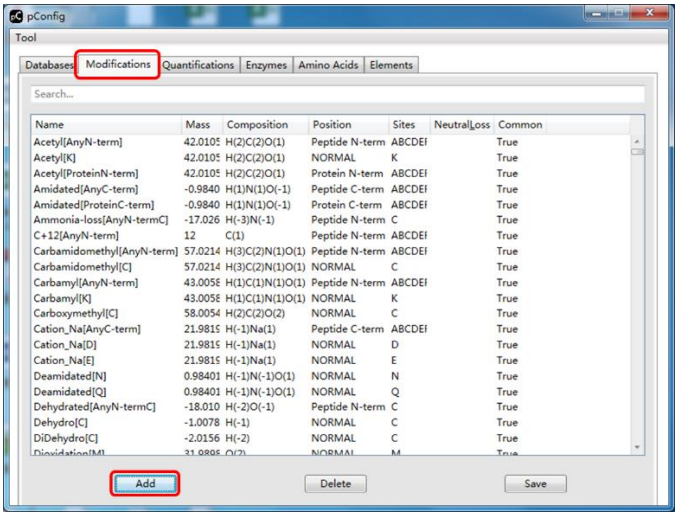

Fig. S2 Select the tab of "Modifications" and click "Add"

c. Setting the chemical formula (Fig. S3).

c.1 Type the name of this modification;

c.2 Click "Edit…" to edit the chemical formula;

c.3 Click "Element" box to select "H" and click " + ", fill the space with the number of H

atoms;

c.4 Repeat c.3 to add “C” and “O” element;

c.5 Click “Apply” and the monoisotopic mass will be calculated automatically.

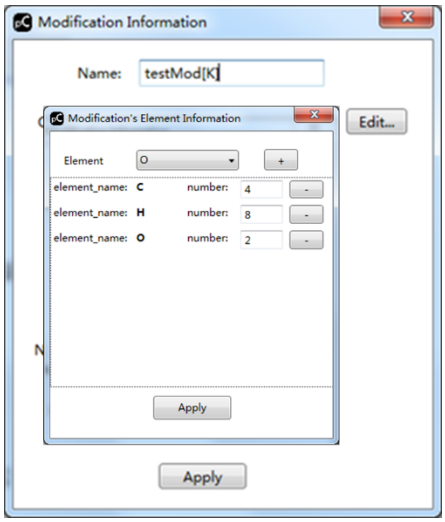

Fig. S3 Setting the chemical formula

d. Setting the modification sites (Fig. S4).

d.1 Select “Anywhere” in the “Position” box.

d.2 Fill the “Sites” with “K” for lysine.

d.3 Fill the “Neutral Loss” with the exact monoisotopic mass of the neutral loss. Leave it blank if no neutral loss is expected from this modification.

d.4 Click “Apply”.

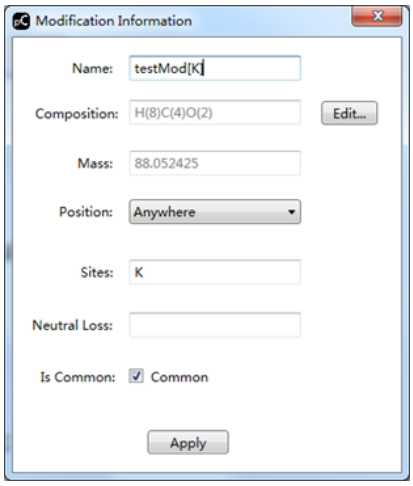

Fig. S4 Setting the modification sites

e. Click “Save” in the “pConfig” panel to finish the setting.

Q8: How to create a new SILAC label?

ANSWER: follow the very long instruction below and you will become a pro. As an example, we demonstrate how to add a medium SILAC label of Arg6Lys4 (Arg-$^{13}C_6$, Lys-$^{2}H_4$) and how to set up the parameters for triple SILAC-labeling (Arg0Lys0, Arg6Lys4, and Arg10Lys8).

a. Define a new element "1H" (Later, four "1H" atoms of a lysine residue will be labeled with either hydrogen or deuterium, and eight "H" atoms will remain hydrogen).

a.1 Open the "bin" folder of pFind 3 and find "element.ini" (Fig.S5).

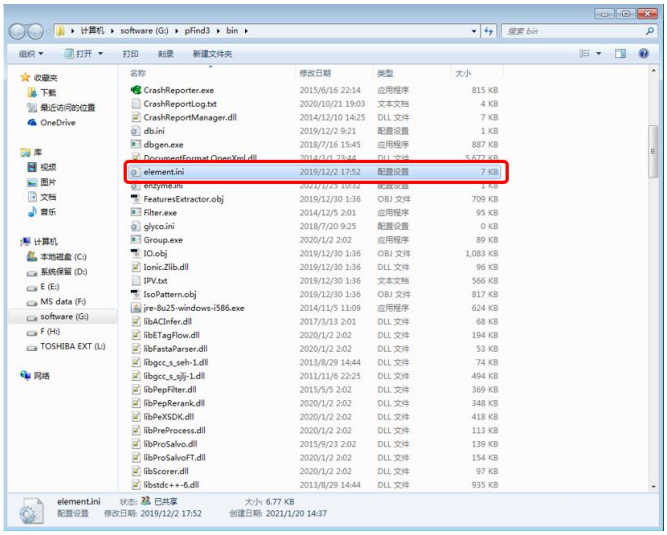

Fig. S5 Find the file 'element.ini'

a.2 Open "element.ini" using a text editor and copy the information of "H" (Fig. S6). Note that on this list, hydrogen "H" is the second entry or E2 and deuterium "2H" is the 115th entry or E115.

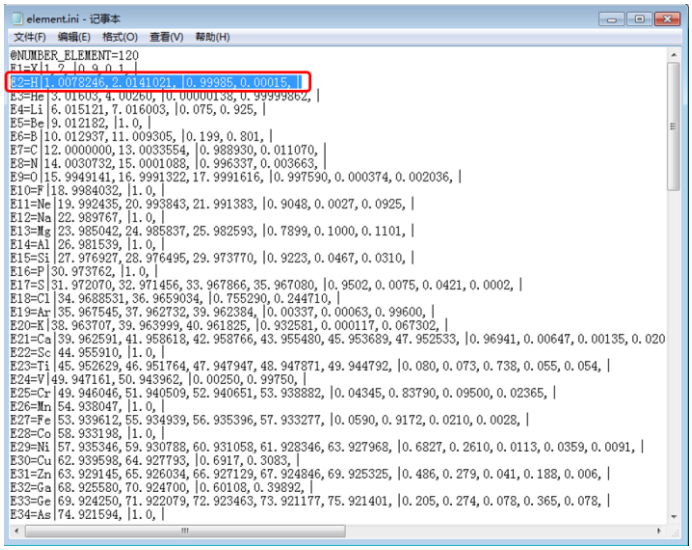

Fig. S6 Copy the information of "H"

a.3 Scroll to the bottom of the list, paste “H” and change element name “H” to “1H” (Fig. S7).

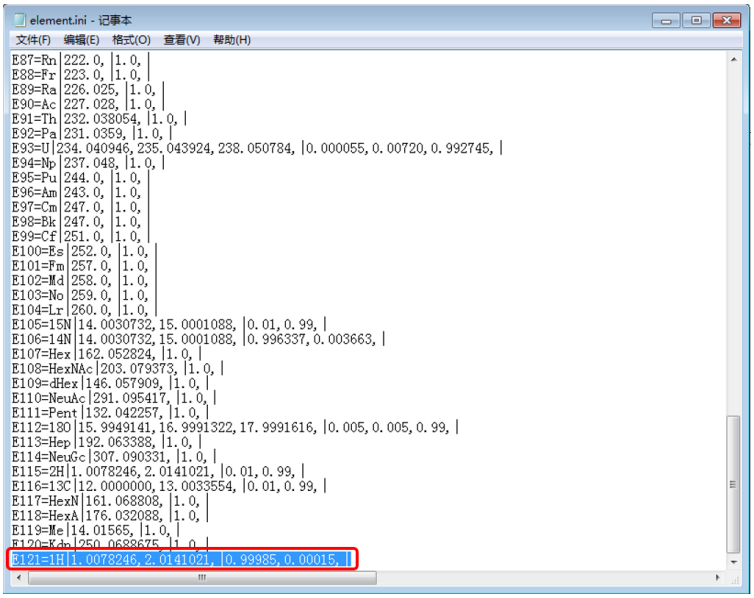

Fig. S7 Change element name “H” into “1H”

a.4 Back to the top and update the total number of elements from 120 to 121 (Fig. S8).

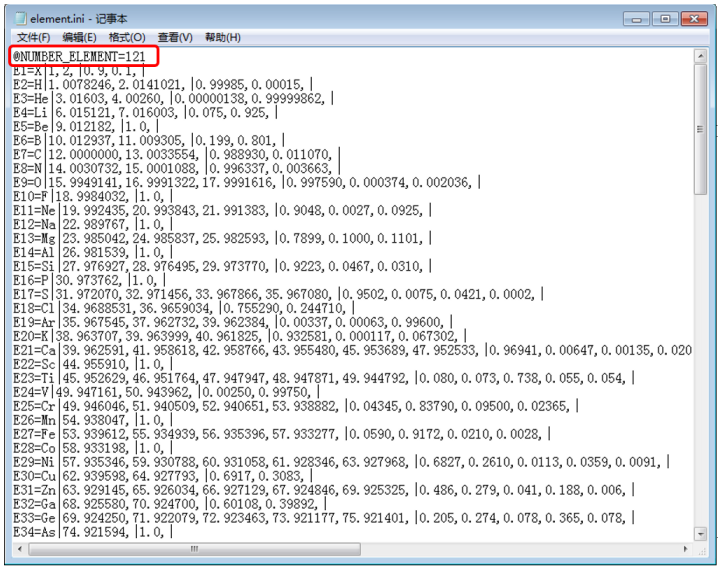

Fig. S8 Change the total number of elements

b. Change the elemental composition of Lys from C(6)H(12)N(2)O(1) to C(6)H(8)1H(4)N(2)O(1).

b.1 From the pFind menu bar, click “Options” and select “Meta Data Configuration” (same as Fig. S1).

b.2 In the “pConfig” panel that pops up, pick up the tab of Amino Acids, and double click on amino acid “K” (Fig. S9).

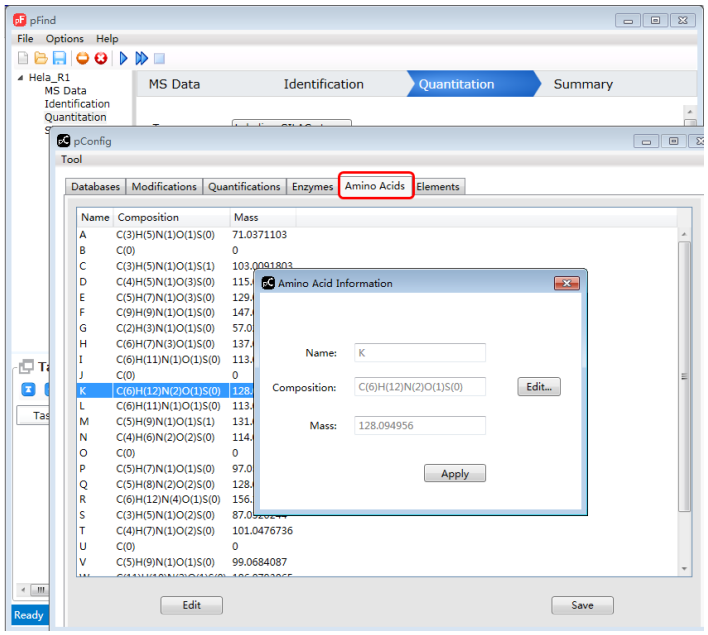

Fig. S9 Select amino acid ‘K’

b.3 Define the elemental composition of “K”, change H(12) to H(8)1H(4) (Fig. S10–S11).

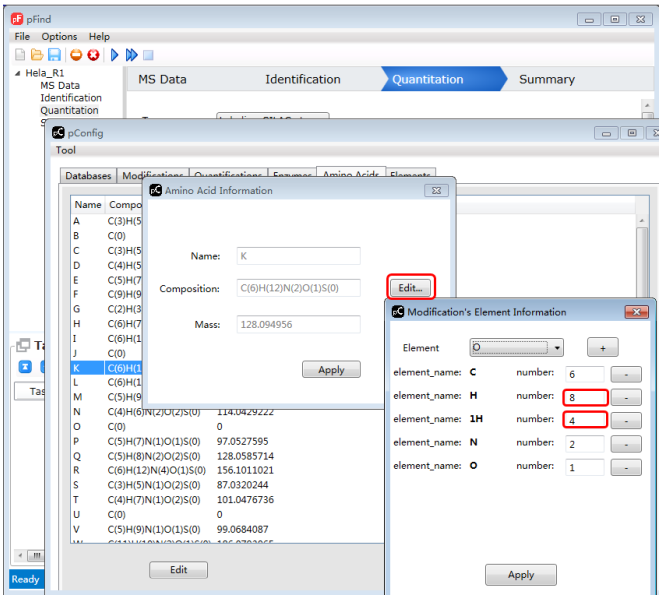

Fig. S10 Define the element of Lys

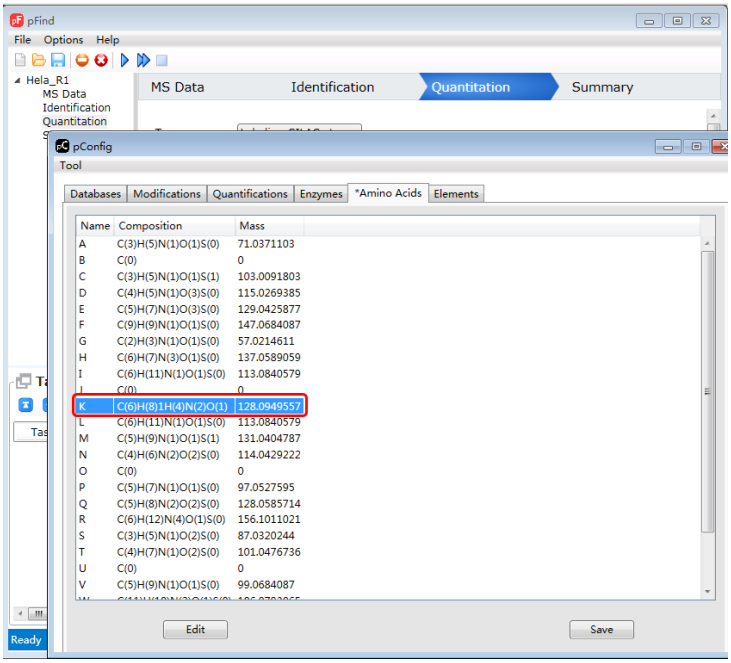

Fig. S11 View the element of Lys after modifying it

c. Set up the SILAC parameters.

c.1 Repeat step b.1. Pick up the tab of "Quantifications" and click "Add" to make a new SILAC Labeled type "SILAC-Arg6Lys4" (Fig. S12).

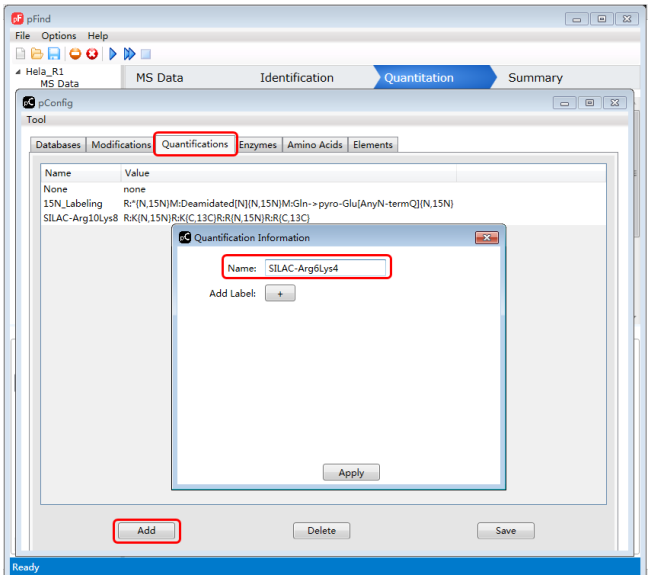

Fig. S12 Add a new SILAC label

c.2 Specify which element is to be labeled by what in which amino acid. For arginine in "SILAC-Arg6Lys4" the element "C" is to be labeled by ^{13}C. For lysine, the element "1H" is to be labeled by deuterium (^{2}H), as shown in Fig. S13. Click "Apply".

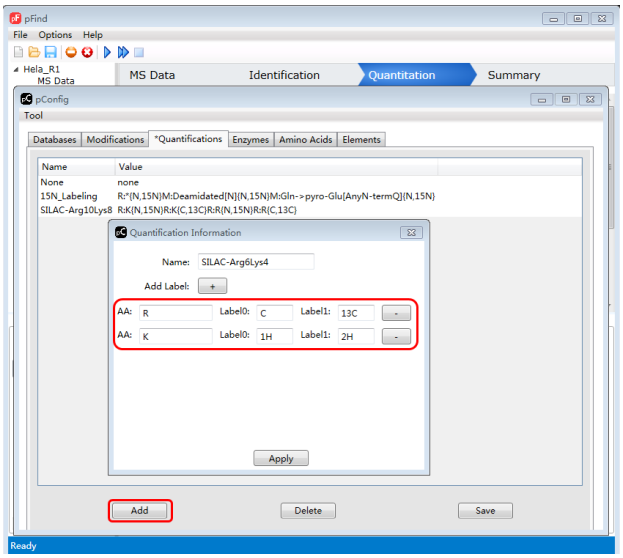

Fig. S13 Add the light and the heavy elements

c.3 Save the new SILAC-label “SILAC-Arg6Lys4” (Fig. S14).

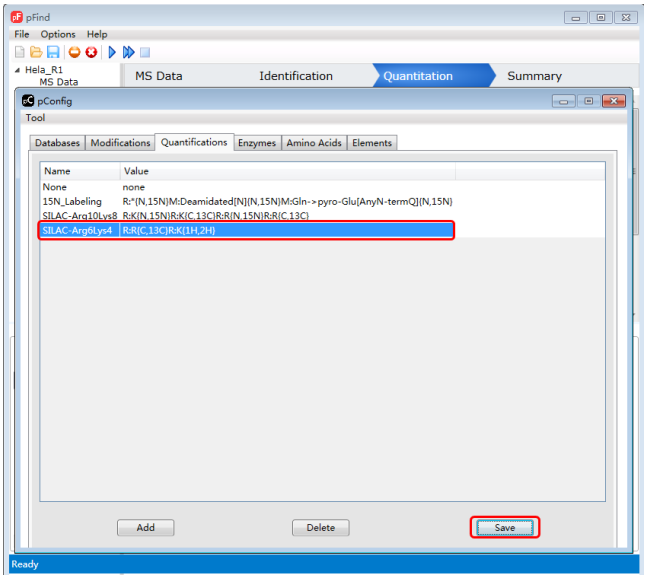

Fig. S14 Save the new SILAC label

d. Switch to the Quantitation panel and set up the parameters as shown in Fig. S15. Importantly, select “3” in “Multiplicity” and set the “Light Label”, “Medium Label”, and “Heavy Label” to “None”, “SILAC-Arg6Lys4”, and “SILAC-Arg10Lys8”, respectively.

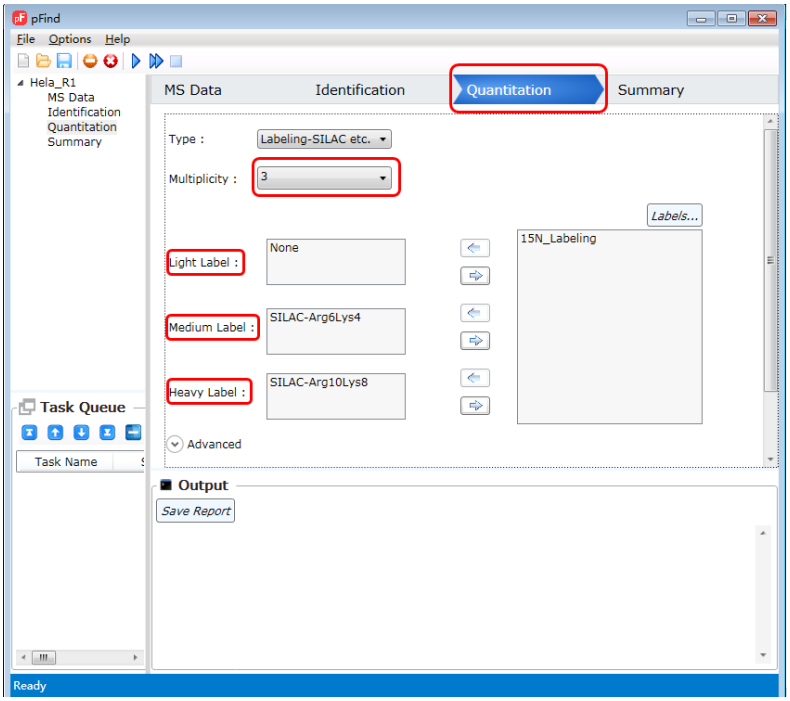

Fig. S15 Set up the quantitation parameters

Q9: For a phosphopeptide containing multiple serine and threonine residues, does pFind 3 provide a score to indicate the confidence level with which the phosphate group is assigned to the reported site? This is important for a follow-up mutagenesis experiment.

ANSWER: The confidence level associated with the localization of a PTM is not calculated in the pFind 3 program but can be obtained by feeding the pFind 3 search result to another program called PTMiner (An *et al*. 2019). We recommend PTMiner for its outstanding performance and smooth handling of pFind 3 search results. PTMiner is available at http://fugroup.amss.ac.cn/software/ptminer/ptminer.html.

Reference:

Chi H, Liu C, Yang H, Zeng W-F, Wu L, Zhou W-J, Wang R-M, Niu X-N, Ding Y-H, Zhang Y (2018) Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. Nat Biotechnol 36(11):1059-1061

An Z, Zhai L, Ying W, Qian X, Gong F, Tan M, Fu Y (2019) 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 18(2):391-405