# **Supplementary Materials**

## **Supplementary tables**

**Table S1** The datasets used in this protocol — sample preparation

Data type	Non-labeling	<sup>14</sup> N/ <sup>15</sup> 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	<sup>14</sup> N/ <sup>15</sup> N-labeled <i>C. elegans</i> lysates	Phosphopeptides enriched from SILAC-labeled human cell lysates	TMT-labeled peptides from human cell lysates
Light/Heavy mixing ratio	Not applied	1:1 <sup>14</sup> N/ <sup>15</sup> N	1:1 SILAC KOR0/K8R10	1:1:1:1:1:1 of MS2 reporter ions 126/127/128/129/130/131
Reduction	ТСЕР			
Alkylation	IAA			
Enzyme digestion	Trypsin			

Data type	Non-labeling	<sup>14</sup> N/ <sup>15</sup> N-labeling	SILAC-labeling	TMT-labeling
RAW files	ID_data_HeLa_QE_	Quant_15Ndata_	Quant_SILACdata_	Quant_TMTdata_
	HF_120min_rep1.raw	Celegans_QE_	HeLa_QE_	RMS_QE_
	ID_data_HeLa_QE_	105min.raw	110min.raw	105min. raw
	HF_120min_rep2.raw			
Effective LC gradie	ent			
Running time	2–100 min	2–85 min	0–80 min	2–58 min
Gradient	7 20	2 27	2 27	5 29
(% of acetonitrile)	7–30	3–27	3–27	5–28
MS data acquisition	1			
Instrument type	Q Exactive HF	Q-Exactive	Q-Exactive	Q-Exactive
MS AGC target	$3 \times 10^{6}$	$3 \times 10^{6}$	$1 \times 10^{6}$	$3 \times 10^{6}$
MS Resolution	60,000	70,000	70,000	70,000
MS Maximum IT	60 ms	60 ms	60 ms	60 ms
MS1 Scan range	200–2000	300-2000	400-2000	300-2000
(m/z)	200-2000	300-2000	400-2000	300-2000
MS2 AGC target	$1 \times 10^{5}$	$1 \times 10^{5}$	$5 \times 10^4$	$1 \times 10^{5}$
MS2 Resolution	15,000	17,500	17,500	17,500
MS2 Maximum IT	60 ms	60 ms	250 ms	60 ms
Isolation Window	2 <i>m</i> / <i>z</i>	2 <i>m/z</i>	2 <i>m</i> / <i>z</i>	2 <i>m</i> / <i>z</i>
NCE (HCD)	27	27	27	32
Intensity threshold	$3 \times 10^4$	$5 \times 10^4$	$4 \times 10^{3}$	$5 \times 10^{4}$
Dynamic exclusion	45 s	30 s	60 s	30 s

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

Annotation of pFin	d 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	
<i>M</i> / <i>Z</i> : 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 <sup>15</sup> 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
Ĩ	manual checking or file sharing between database search engines

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

Annotation of pFind param	neters 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 <i>in silico</i> 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	
$FDR \leq 1\%$	FDR at the peptide or spectrum level. The peptide level is set by default
(peptide/spectra)	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 $\ge 1$	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 ≤ 1%	FDR at the protein level
	*

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

Annotation of the <i>Summary</i> par	el 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

#### **Table S5** Annotation of the Summary panel in pBuild

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 <i>in silico</i>	
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	Which protease is used to digest samples? Trypsin in	
	KR_C	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 labelin quant.ini	ng type can be changed in the configuration file named	
Chrom tolerance	15 ppm	This is the maximal allowed mass deviation specified	
	ie pp	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		1	
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	

Annotation of pFind parameter	ers in the <i>Quantitation</i> panel
MS1 quantitation	
Туре	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 <sup>1</sup> H, <sup>12</sup> C, and <sup>14</sup> N
Heavy Label	The heavy stable isotope label, $e.g.$ deuterium or <sup>15</sup> 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	The number of MS1 scans along half of the chromatogram peak.
Per Half CMTG: 200	Generally, it is 100–200
Number of Holes	If the peptide (or reporter ion) intensity in two consecutive MS1 (or
in CMTG: 2	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	Similar to precursor tolerance in the identification panell. pQuant will
Tolerance in ppm: 15	find the experimental $m/z$ of precursor to reconstruct the
	chromatographic peak. A value of 15 means pQuant will allow $\pm$ 15
	ppm mass accuracy when it constructs a chromatographic peak
TYPE_SAME_START_END_	For 1:1 mixed sample, the length of chromatograph of light and heavy
BETWEEN_EVIDENCE (For	isotope are the same. For the 10:1 or 1:10 mixed sample, the low
1:1 Mixed Samples /For10:1	abundance isotope has the shorter chromatograph
or 1:10 Mixed Samples/	
Independent)	
ELEMENT_ENRICHMENT_	If 15N or 13C is entered here, pQuant will estimate the atomic
CALIBRIATION	enrichment ratio of <sup>15</sup> N or <sup>13</sup> C
(none/15N/13C)	
MS2	Since quantitation here is based on MS1, make sure the option 'MS2
	Quantitation' is unchecked. It is off by default
MS2 quantitation	

 Table S6 Annotation of pFind parameters in the Quantitation panel

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

Annotation of	the <i>Protein</i> 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 S7** Annotation of the *Protein* panel in pBuild

Annotation of the he	eader 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

Table S8 Annotation of the header in the file 'pFind.protein'

	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

Annotation of the head	ler 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_pep_num@pro: the total number of peptides identified for a protein;
(Total_unique_pep_	Total_unique_pep_num@pro: the total number of unique peptides identified
num@pro)	for a protein
Total_spec_num@pro	Total_spec_num@pro: the total number of PSMs identified for a protein;
(Total_unique_spec_	Total_unique_pep_num@pro: the total number of PSMs of the uniquely
num@pro)	identified peptides for the protein in question
	Whether or not a protein has unique peptides identified, if yes, it is marked
Have unique peptide?	'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
	Whether or not a peptide can be mapped back to a single protein, <i>i.e.</i> ,
	whether or not it is a unique peptide. If yes, it is marked 'is unique'; if not,
Is unique?	'not unique'
	If a peptide is not a unique peptide, <i>i.e.</i> , it can be mapped back to other
	proteins in addition to the one in question, then the other proteins are listed
Shared proteins	here

 Table S9
 Annotation of the header in the file 'pFind\_protein\_contrast\_result.txt'

Annotation of the head	er 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_spec_num@pro: the total PSM number of peptides with the
(Total_unique_spec_	modification identified in this site; Total_unique_pep_num@site: the total
num@site)	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'

 $Table \ S10 \ {\rm Annotation \ of \ the \ header \ in \ the \ file \ `pFind_PTM_contrast_result.txt'}$ 

### FAQs

- Q1: Can I use pFind 3 to analyze ETD, EThcD, or ETciD data? <u>ANSWER</u>: 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? <u>ANSWER</u>: 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?
   <u>ANSWER</u>: 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?
  <u>ANSWER</u>: 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? <u>ANSWER</u>: 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"?
   <u>ANSWER</u>: 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?

<u>ANSWER</u>: 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).

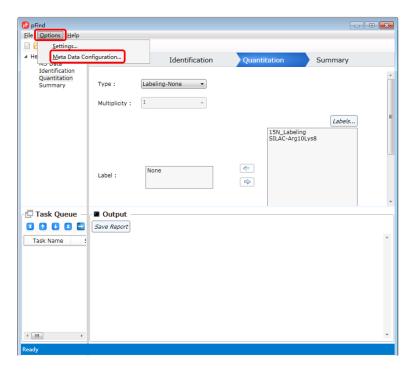


Fig. S1 Open "Option" and select "Meta Data Configuration"

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

atabases Modifications Qua	intificatio	ins Enzymes A	mino Acids Eler	ments			
Search							
Search							
Name	Mass	Composition	Position	Sites	NeutralLoss	Common	
Acetyl[AnyN-term]	42.0105	H(2)C(2)O(1)	Peptide N-term	ABCDEF		True	
Acetyl[K]	42.0105	H(2)C(2)O(1)	NORMAL	К		True	
Acetyl[ProteinN-term]	42.0105	H(2)C(2)O(1)	Protein N-term	ABCDEF		True	
Amidated[AnyC-term]	-0.9840	H(1)N(1)O(-1)	Peptide C-term	ABCDEF		True	
Amidated[ProteinC-term]	-0.9840	H(1)N(1)O(-1)	Protein C-term	ABCDEF		True	
Ammonia-loss[AnyN-termC]	-17.026	H(-3)N(-1)	Peptide N-term	С		True	
C+12[AnyN-term]	12	C(1)	Peptide N-term	ABCDEF		True	
Carbamidomethyl[AnyN-term]	57.0214	H(3)C(2)N(1)O(1)	Peptide N-term	ABCDEF		True	
Carbamidomethyl[C]	57.0214	H(3)C(2)N(1)O(1)	NORMAL	С		True	
Carbamyl[AnyN-term]	43.0058	H(1)C(1)N(1)O(1)	Peptide N-term	ABCDEF		True	
Carbamyl[K]	43.0058	H(1)C(1)N(1)O(1)	NORMAL	К		True	
Carboxymethyl[C]	58.0054	H(2)C(2)O(2)	NORMAL	С		True	
Cation_Na[AnyC-term]	21.9819	H(-1)Na(1)	Peptide C-term	ABCDEF		True	
Cation_Na[D]	21.9819	H(-1)Na(1)	NORMAL	D		True	
Cation_Na[E]	21.9819	H(-1)Na(1)	NORMAL	E		True	
Deamidated[N]	0.98401	H(-1)N(-1)O(1)	NORMAL	N		True	
Deamidated[Q]	0.98401	H(-1)N(-1)O(1)	NORMAL	Q		True	
Dehydrated[AnyN-termC]	-18.010	H(-2)O(-1)	Peptide N-term	C		True	
Dehydro[C]	-1.0078	H(-1)	NORMAL	C		True	
DiDehydro[C]	-2.0156	H(-2)	NORMAL	С		True	
DiovidationIMI	21 0905	0(2)	NORMAL	м		True	*

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.

1	∎C N	lodification I	nformation		×
		Name:	testMod[K]		
	4	Modification	s Element Information	×	Edit
		Element	•	+	
	N	element_name: element_name: element_name:		4 - 8 - 2 -	
			Apply		
			Apply	)	-

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".

Name:	testMod[K]	
Composition:	H(8)C(4)O(2)	Edit
Mass:	88.052425	
Position:	Anywhere	T
Sites:	К	
Neutral Loss:		
Is Common:	Common	

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?

<u>ANSWER</u>: 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-<sup>13</sup>C<sub>6</sub>, Lys-<sup>2</sup>H<sub>4</sub>) 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).

编织 🔹 🗐 打开 🔹	打印 刻景 新建文件夹				H • 🔲 🕯
☆ 約藤平	名称	修改日期	类型	大小	
1 TSE	CrashReporter.exe	2015/6/16 22:14	应用程序	815 KB	
3. 景近访问的位置	CrashReportLog.txt	2020/10/21 19:03	文本文档	4 KB	
	CrashReportManager.dll	2014/12/10 14:25	DLL 文件	7 KB	
Chebrive	a db.ini	2019/12/2 9:21	配置设置	1 KB	
	📧 dbgen.exe	2018/7/16 15:45	应用程序	887 KB	
二月 (二)	DocumentFormat OpenXml dll	2014/3/1 23:44	DU 文件	5.677.KB	
📷 视频	(a) element.ini	2019/12/2 17:52	配置设置	7 KB	
📓 图片	g_ enzyme.ini	2021/1/25 10:32	INCOME OF CHIL	1 KB	
文档	🐔 FeaturesExtractor.obj	2019/12/30 1:36	OBJ 文件	709 KB	
2) 音乐	Filter.exe	2014/12/5 2:01	应用程序	95 KB	
	glyco.ini	2018/7/20 9:25	配置设置	0 KB	
🌉 计算机	Group.exe	2020/1/2 2:02	应用程序	89 KB	
🏭 本地磁盘 (C:)	📑 IO.obj	2019/12/30 1:36	OBJ 文件	1,083 KB	
	🖉 Ionic.Zlib.dll	2019/12/30 1:36	DLL 文件	96 KB	
E (E:)	IPV.txt	2019/12/30 1:36	文本文档	566 KB	
MS data (F:)	📑 IsoPattern.obj	2019/12/30 1:36	<b>OBJ</b> 文件	817 KB	
software (G:)	jre-8u25-windows-i586.exe	2014/11/5 11:09	应用程序	624 KB	
	ibACInfer.dl	2017/3/13 2:01	DLL 文件	68 KB	
🕞 F (H:)	ibETagFlow.dll	2020/1/2 2:02	DLL 文件	194 KB	
TOSHIBA EXT (L:)	ibFastaParser.dll	2020/1/2 2:02	DLL 文件	53 KB	
	libgcc_s_seh-1.dl	2013/8/29 14:44	DLL 文件	74 KB	
9 网络	libgcc_s_sjij-1.dll	2011/11/6 22:25	DLL 文件	494 KB	
	ibPepFilter.dll	2015/5/5 2:02	DLL 文件	369 KB	
	libPepRerank.dll	2020/1/2 2:02	DLL 文件	348 KB	
	libPeXSDK.dll	2020/1/2 2:02	DLL 文件	418 KB	
	libPreProcess.dll	2020/1/2 2:02	DLL 文件	113 KB	
	libProSalvo.dll	2015/9/23 2:02	DLL 文件	139 KB	
	libProSalvoFT.dll	2020/1/2 2:02	DLL 文件	154 KB	
	ibScorer.dll	2020/1/2 2:02	DLL 文件	97 KB	
	ibstdc++-6.dll	2013/8/29 14:44	DLL 文件	935 KB	

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.

📃 eleme	ent.ini - 记事	本										
文件(F)	编辑(E) 格	(C) 查看	(V) 帮助	I(H)								
	LELEMENT											
F1=X 1	2 10 9 0	1										
82=H 1.	0078246,	2.0141021		985,0.00								
E3=He 3	3.01603,4	. 00260, 10	.00000	138, 0. 99	aaaasez,							
		7.016003,	0.075	, 0. 925,	I							
	9. 012182,		10.00									
		11.009305										
					0.011070,							
					0.003663,			o. 1				
			22, 17.	9991616,	10. 331230	), 0. 000374	, 0. 0020	36,				
E10=F   1	8.998403	2, 11. 0, 1	40.01	001202	0.0040.0	0007 0 00	or I					
	22. 98976		43, 21.	991383,	10. 9048, 0.	0027, 0. 09	25, [					
			97 OF	000502	0 7900 0	1000.0.11	01 I					
	26, 98153		51,25.	982595,	10. 1899, 0.	1000, 0. 11	01, 1					
			OF 20	072770	0 0222 0	0467.0.03	10.1					
	30. 973762		50, 25.	515110,	10. 9220, 0.	0401, 0. 03	10, 1					
			0 22 3	67266 39	967090	0.9502.0.	0075 0	0421 0 00	102 I			
					0.0.244710		0010, 0.	0421, 0, 00	102,1			
						, 00063, O.	1_00866					
20=K	38. 963707	39, 96399	9.40.9	61825 10	932581	0000117.0	067302	. 1				
						45. 953689			941.0.0	0647.0.0	0135.0.0	020
	44.95591		,	,			,	,	,	,	,	
			64.47.	947947.4	18.947871.	49.944792	. 10, 080	. 0. 073. 0.	738.0.0	55.0.054	.	
		, 50. 94396					,,	,,	,	,	, ,	
25=Cr	49.94604	6, 51, 9405	09, 52.	940651.5	53. 938882.	0.04345.	0.83790	.0.09500.	0.02365	. I		
	54.93804											
			39, 56.	935396,5	57.933277,	0.0590,0	. 9172, 0.	. 0210, 0. 0	028,			
	58.93319											
						63.927968	,  0.682	7, 0. 2610,	0.0113,	0.0359,0	.0091,	
		8,64.9277										
						69.925325	, [0. 486	, 0. 279, 0.	041, 0.1	88, 0. 006	,	
		0,70.9247										
			79, 72.	923463,	13. 921177,	75.921401	, 10.205	, 0. 274, 0.	078, 0. 3	65, U. 078,	,	
	74.92159	94,  1. U,										
٠												•

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).

element.ini - 记事本	- • •
文件(F) 編輯(E) 格式(O) 查看(V) 帮助(H)	
B37=Rn [22:0, [1.0,] B37=Rn [22:0, [1.0,] B39=Rn [22:0, 25, [1.0,] B90=Rn [22:038054, [1.0,] B91=Th [23:038054, [1.0,] B92=Pn [231.0359, [1.0,] B93=U1[234, 004946, 235:043924, 238.050784, [0.000055, 0.00720, 0.992745, ] B93=U1[24, 0.0140, [1.0,] B93=Dn [247, 0, 1.0,] B93=Dn [247, 0, 1.0,] B93=Dn [247, 0, 1.0,] B100=En [252, 0, [1.0,] B100=En [252, 0, [1.0,] B101=Fm [257, 0, 1.0,] B102=W1 [258, 0, 1.0,] B103=W0 [258, 0, 1.0,] B104=Lr [260, 0, 1.0,] B105=HN [14, 0030732, 15.0001088, [0.01, 0.99,] B106=I4W1 [14, 0030732, 15.0001088, [0.996337, 0.003663,] B107=Har [162, 052824, [1.0,] B103=HexNAc [203.073373, [1.0,]]	
E110=Neuke [291, 095417, [1, 0, ] E111=Pent 132, 042257, [1, 0, ] E112=Hei 132, 042257, [1, 0, ] E113=Hei 192, 063388, [1, 0, ] E144=HeinGe [307, 090331, [1, 0, ] E145=E2H[1, 0078246, 2, 0141021, [0, 01, 0, 99, ] E116=13C [12, 0000000, 13, 0033554, [0, 01, 0, 99, ] E117=HeixN [161, 068808, [1, 0, ] E118=HeixN [160, 068808, [1, 0, ] E119=Hei 14, 01565, [1, 0, ] E129=Hei 14, 01565, [1, 0, ] E129=Hei 14, 01565, [1, 0, ] E121=HH1, 0078246, 2, 0141021, [0, 99985, 0, 00015, ]]	=

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).

elementini - 记事本
文件(F) 編撮(E) 格式(O) 查看(V) 帮助(H)
ONUMBER ELEMENT=121
B1=X[1, 2, [0, 9, 0, 1, ]
E2=H1, 1, 2078246, 2, 0141021, [0. 99985, 0. 00015, ]
D=111:0:01603, 4. 00260, 10. 00000138, 0. 99999862,
E4=Li 6. 015121, 7. 016003, [0, 075, 0, 925, ]
E5=Be 9. 012182, 11.0,
E6=B 10. 012937, 11. 009305, [0. 199, 0. 801, ]
E7=C 12. 0000000, 13. 0033554, 0. 988930, 0. 011070,
E8=N 14. 0030732, 15. 0001088, 0. 996337, 0. 003663,
E9=0 15. 9949141, 16. 9991322, 17. 9991616, 0. 997590, 0. 000374, 0. 002036,
E10=F   18. 9984032,   1. 0,
E11=Ne 19. 992435, 20. 993843, 21. 991383, 0. 9048, 0. 0027, 0. 0925,
E12=Na 22. 989767, 11.0,
E13=Mg 23. 985042, 24. 985837, 25. 982593, 0. 7899, 0. 1000, 0. 1101,
E14=A1 26, 981539, 11, 0,
E15=S1 27. 976927, 28. 976495, 29. 973770, 0. 9223, 0. 0467, 0. 0310,
E16=P   30. 973762,   1. 0,   E17=S   31. 972070, 32. 971456, 33. 967866, 35. 967080,   0. 9502, 0. 0075, 0. 0421, 0. 0002,
E1-5151. 912010, 52. 911456, 55. 961866, 55. 961866, 10. 9502, 0. 0015, 0. 0421, 0. 0002, 1 E18=C1   34. 9688531, 36. 9659034,   0. 755290, 0. 244710,
E13-C1 34, 9668551, 56, 9659054, 10, 755290, 0, 244110, 1 E19=Ar 35, 967545, 37, 962732, 39, 962384, 10, 00337, 0, 00063, 0, 99600, 1
E19-A1 135. 501343, 31. 502132, 35. 502384, 10. 00331, 0. 00003, 0. 55000, 1 E20=K 38. 963707, 39. 963999, 40. 961825, 0. 932581, 0. 000117, 0. 067302,
$E_2 = -6[39, 962591, 41, 958618, 42, 958766, 43, 955480, 45, 953689, 47, 952533, [0, 96941, 0, 00647, 0, 00135, 0, 020]$
E23=Ti 45. 952629, 46. 951764, 47. 947947, 48. 947871, 49. 944792, 0. 080, 0. 073, 0. 738, 0. 055, 0. 054, 0.
E24=V 49. 947161, 50. 943962, 0. 00250, 0. 99750,
E25=Cr   49. 946046, 51. 940509, 52. 940651, 53. 938882,   0. 04345, 0. 83790, 0. 09500, 0. 02365,
E26=Mn 54. 938047, 1.0,
E27=Fe 53. 939612, 55. 934939, 56. 935396, 57. 933277, 0. 0590, 0. 9172, 0. 0210, 0. 0028,
E28=Co 58. 933198, 11.0,
E29=Ni 57. 935346, 59. 930788, 60. 931058, 61. 928346, 63. 927968, 0. 6827, 0. 2610, 0. 0113, 0. 0359, 0. 0091,
E30=Cu 62. 939598, 64. 927793, 0. 6917, 0. 3083,
E31=Zn 63. 929145, 65. 926034, 66. 927129, 67. 924846, 69. 925325, 0. 486, 0. 279, 0. 041, 0. 188, 0. 006,
E32=6a [68, 925580, 70, 924700, 10, 60108, 0, 39892, ]
E33=Ge 69. 924250, 71. 922079, 72. 923463, 73. 921177, 75. 921401, 0. 205, 0. 274, 0. 078, 0. 365, 0. 078, 0.
E34=As 74. 921594, 1.0,
< m h

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).

	tions Hel									
✓ Hela_F MS Ide		MS Data		Identifica	ition	Quantitatio	n	Summa	ary	*
Ĩ	Database	es Modifications Qu	antifications		Amino Acids	Elements				
	Database	es   Modifications   Qu	antifications	Enzymes	Amino Acids	Elements				
	Name	Composition	Mass							
	A	C(3)H(5)N(1)O(1)S(0)	71.037110	3						~
	В	C(0)	0							
		C(3)H(5)N(1)O(1)S(1)	103.00918							
	DE	C(4)H(5)N(1)O(3)S(0) C(5)H(7)N(1)O(3)S(0)	115. 💽 A 129.	mino Acid I	Information			<b>X</b>		
	F	C(9)H(9)N(1)O(1)S(0)	129.							
	G	C(2)H(3)N(1)O(1)S(0)	57.0							
	H	C(6)H(7)N(3)O(1)S(0)	137.							
dQ Ta	I	C(6)H(11)N(1)O(1)S(0)	113.	Name:	K					
	J	C(0)	0							=
		C(6)H(12)N(2)O(1)S(0)	120.	omposition:	C(6)H(12)	N(2)O(1)S(0)	Edi	t		
Tas		C(6)H(11)N(1)O(1)S(0)	113.							
	M	C(5)H(9)N(1)O(1)S(1)	131.	Mass:	128.0949	6				
	N	C(4)H(6)N(2)O(2)S(0)	114.							
	O P	C(0) C(5)H(7)N(1)O(1)S(0)	0 97.0							
	Q	C(5)H(7)N(1)O(1)S(0) C(5)H(8)N(2)O(2)S(0)	128.			Apply				
	R	C(6)H(12)N(4)O(1)S(0)	156.							
	S	C(3)H(5)N(1)O(2)S(0)	87.0	-						
	т	C(4)H(7)N(1)O(2)S(0)	101.04767	36						
	U	C(0)	0							
	V	C(5)H(9)N(1)O(1)S(0)	99.068408							-
< III		C(14) U(10) N(2) O(1) C(0)	100.07000							
		Edit						Save		
Ready		con						0000		

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).

	00						
Hela_R MS	1 Data	MS I	Data	Identification	Quantitation	Summar	y
	ntification						
Qua	intitation						
	C pConfig						
	Tool						
	Databases			ations Ensures Amino /	at the set of the		
	Databases		Amino Acid Infe		Elamante I		
	Name		Amino Acid Inio	ormauon			
		C(3)H(5					
		C(0)					
		C(3)H(5		К			
		C(4)H(5	Name:	K			
		C(5)H(7		C(6)11(12)N(2)O(1)C(0)	Edit		
	E O	C(9)H(9	Composition:	C(6)H(12)N(2)O(1)S(0)	Edit		
		C(2)H(3			Modification's	s Element Information	
		C(6)H(7	Mass:	128.094956			
QТ.		C(6)H(1			Element	•	+
		C(0) C(6)H(1					<u> </u>
		C(6)H(1		Apply	element_name:	C number:	6 -
Tas		C(5)H(9			element name:	H number:	8 .
		C(4)H(6)N(2)	0(2)5(0) 114	.0429222	-		
		C(0)	0		element_name:	1H number:	4 -
	P	C(5)H(7)N(1)	O(1)S(0) 97.0	0527595	element_name:	N number:	2 -
		C(5)H(8)N(2)		.0585714		•	
		C(6)H(12)N(4		.1011021	element_name:	o number:	1 -
		C(3)H(5)N(1)		320244			
		C(4)H(7)N(1)		.0476736			
		C(0) C(5)H(9)N(1)	0	0684087			
	V V	C(2)H(3)M(T)	O(T)2(0) 33'(	1084087			

Fig. S10 Define the element of Lys

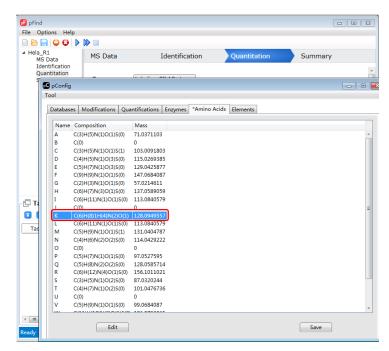


Fig. S11 View the element of Lys after modifying it

- 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).

🛃 pFind	X
File Options Help	
Hela_R1     MS Data     Identification     Quantitation     Summary	/
C pConfig	
Tool	
Databases Modifications Quantifications Enzymes Amino Acids Elements	
Name Value	1
None none 15N_Labeling R:*(N.15N)M:Deamidated[N]{N,15N}M:Gln->pyro-Glu[AnyN-termQ]{N,15N}	
SILAC-Arg10Lys8 R:K(N,15N)R:K(C,13C)R:R(N,15N)R:R(C,13C)	
Quantification Information	
Name: SILAC-Arg6Lys4	
Add Label: +	
	l h
Apply	
Add Delete Save	

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 <sup>13</sup>C. For lysine, the element "1H" is to be labeled by deuterium (<sup>2</sup>H), as shown in Fig. S13. Click "Apply".

ela_R1 MS Data	MS Data	Identification	Quantitation	Summary
pConfig				
pol				
Databases Mod	difications *Quantification	ns Enzymes Amino Acids	Elements	
Name	Value			
None	none			
15N_Labeling SILAC-Arg10Ly	R:*{N,15N}M:Deamida s8 R:K{N,15N}R:K{C,13C}F	ed[N]{N,15N}M:Gln->pyro-Glu :R{N,15N}R:R{C,13C}	[AnyN-termQ]{N,15N}	
	Quantification Infor	mation	8	
	Name: SI	AC-Arg6Lys4		
	Add Label:	•		
	AA: R	Label0: C Label1:	13C -	
	AA: K	Label0: 1H Label1:	2Н -	
		Apply		

Fig. S13 Add the light and the heavy elements

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

🗾 pFind								
File Options Help								
			_	_				
✓ Hela_R1 MS Data	MS Data	Identification	Quantitation	Summa	гу			
Config								
Tool								
Databases Modifi	cations Quantifications	Enzymes Amino Acids Ele	ements					
Name	Value							
None	none							
15N_Labeling								
	R:K{N,15N}R:K{C,13C}R:R{ R:R{C,13C}R:K{1H,2H}	N,15N}R:R{C,13C}						
SIENC Algocyse	nate, 150 partiti, 201							
1								
	Add	Delete		Save				
		. ,						
Ready								

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.

🕫 pFind				
<u>File Options H</u> elp				
🖻 🗁 🔜 🗢 😂 🕨				
<ul> <li>Hela_R1 MS Data Identification Quantitation Summary</li> </ul>	MS Data	Identification	Quantitation	Summary
	Type : Lat	peling-SILAC etc. 🔹		
	Multiplicity : 3	•		
	_		15N_Labelin	Labels
	Light Label :	lone		9
	Medium Label :	ILAC-Arg6Lys4		
🗖 Task Queue –	Heavy Label :	ILAC-Arg10Lys8	*	
Task Name				
	Output —			
	Save Report			
- <u>III</u>				

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.

<u>ANSWER</u>: 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. <u>PTMiner is available at http://fugroup.amss.ac.cn/software/ptminer/ptminer.html</u>.

#### **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