

## ~Supplementary material~

### PTMoreR: a motif-centric analysis enabling cross-species PTM mapping and comparative phosphoproteomics across mammals

#### Supplementary Notes

**Overview** *PTMoreR* (Post-translational modification ortholog aligner) is not merely a P-site BLAST tool; instead, it considers the surrounding amino acid sequence of PTM sites during BLAST, enabling a motif-centric analysis across species. Particularly, *PTMoreR* supports: 1. Mapping the PTM sites and protein sequences between any species and Human; 2. Calculating sequence window similarity and allowing filtering thresholds of sequence similarity during the mapping; 3. Processing PTM site-specific enrichment analysis and offering flexible annotation based on kinase-substrate database and network plots; 4. Visualizing the regulation of modification sites on the basis of protein-protein interaction data. Here we present the detailed introduction and operation of *PTMoreR*, by which users can follow to analyze their own data freely and conveniently.

Users can visit this site: <https://yslproteomics.shinyapps.io/PTMoreR>. The whole source codes: <https://github.com/wangshisheng/PTMoreR>. Then the website homepage can be shown like this:

**Dear Users, Welcome to PTMoreR**

PTMoreR is a web-based tool, which possesses the core functions, including:

1. Mapping the PTM sites and protein sequences between any species and Human;
2. Calculating sequence window similarity and allowing filtering thresholds of sequence similarity during the mapping;
3. Performing PTM site-specific enrichment analysis and offering flexible annotations based on kinase-substrate database and network plots;
4. Visualizing the regulation of modification sites on the basis of protein-protein interaction data.

In addition, this tool supports both online access and local installation. The source codes and installation instructions can be available in the GitHub repository: <https://github.com/wangshisheng/PTMoreR> under an MIT license.

Finally, PTMoreR is developed by R shiny (Version 1.6.0), and is free and open to all users with no login requirement. It can be readily accessed by all popular web browsers including Google Chrome, Mozilla Firefox, Safari and Internet Explorer 10 (or later), and so on. We would highly appreciate that if you could send your feedback about any bug or feature request to Shisheng Wang at [shishengwang@whscu.cn](mailto:shishengwang@whscu.cn).

**Friendly suggestions:**

- a) Open PTMoreR with Chrome, Mozilla Firefox, Safari or Firefox;
- b) The minimum operating system specifications are: RAM 4GB, Hard drive 500 GB;
- c) The monitor resolution ( $\geq 1920 \times 1080$ ) is better.

© 2024 Yansheng Liu's Group and Hao Yang's Group. All Rights Reserved.  
Created by Shisheng Wang. E-mail: [shishengwang@whscu.cn](mailto:shishengwang@whscu.cn).

## 1. Data Preparation

*PTMoreR* supports four basic file formats (.csv, .txt, .xlsx, .xls). Before analysis, users should prepare their peptide sequences with modification and protein background data. The modified peptide sequence data required here could be readily generated based on results of several popular tools such as MaxQuant<sup>1</sup>, Spectronaut<sup>2</sup> and so on. The users then can upload the two data into *PTMoreR* with right formats respectively and start subsequent analysis.

### 1.1. Modified peptide sequences

#### 1.1.1. Modified peptide sequences with normal type

Herein, the first row is the column name (e.g. AnnotatedPeps) and each of the other rows is a modified peptide sequence. Users need to mark those modified residues (e.g. S, T, Y with phosphorylation) with some label they like (such as '#' or '@') in advance. The peptide sequences can be like:

AnnotatedPeps
GIGT#PPNTTPIK
GIGT#PPNT#T#PIK
NGS#PEIK
KS#ERGMAAK
MNGHS#DEESVR
RQIDS#S#EDEDEDYDNDKR
RYS#GS#DS#DS#ISER
RY#S#GS#DS#DS#ISER
RY#S#GS#DS#DS#IS#ER
KRPY#S#S#FS#NGK

In this situation, *PTMoreR* will search all proteins that these peptides belong to in the “Step 2. Pre-alignment”.

On the other hand, users could also prepare two columns: one column contains protein ids (i.e. UniProt IDs), the other column contains modified peptide sequences. Users need to mark those modified residues (e.g. S, T, Y with phosphorylation) with some label they like (such as '#' or '@') in advance, as below:

UniProt.ID	Pep.upload
D4A9J4	GIGT#PPNTTPIK
D4A9J4	GIGT#PPNT#T#PIK
D4A9J4	NGS#PEIK
D4A9J4	KS#ERGMAAK
D4AAG9	MNGHS#DEESVR
D4AAG9	RQIDS#S#EDEDEDYDNDKR
D4AAG9	RYS#GS#DS#DS#ISER
D4AAG9	RY#S#GS#DS#DS#ISER
D4AAG9	RY#S#GS#DS#DS#IS#ER
D4AAG9	KRPY#S#S#FS#NGK
D4AAG9	AASSGPRS#PLDQR
D4AAG9	S#PYGS#RS#PFEHSAEHR
D4AAG9	S#PFEHSAEHR
D4AAG9	S#T#PEHT#WSSR
M0RBE8	VIHS#S#DEGEDQTGEDEEDDEWDD
A0A0G2K130	DEEDT#S#FESLSK
B5DF91	NMALLS#QLY#HS#PAR
B5DF91	NMALLSQLYHS#PAR

In this situation, PTMoreR will only pre-align peptide sequences to the proteins in the first column.

### 1.1.2. Modified peptide sequences from MaxQuant

If the sequence data are obtained from MaxQuant, then users can find the modified peptide sequences in the modification txt file, for example, the Phospho (STY)Sites.txt file in the output tables from MaxQuant. The peptide sequences are like this:

MaxQuant_Outputs
LFLDGEEKEWAFEE(1)K
FDEGEDGEGS(0.996)NY(0.004)KKLC
ALVADEPEDLDT(1)EDEGLISFEEER
TYS(0.98)S(0.02)SGSSGGSHPPSSR
ELILGS(0.002)ET(0.052)PS(0.779)S(0.167)PR
S(0.008)KS(0.992)PS(0.999)PPRLT(0.001)EDR
AAKLS(1)EGS(1)QPAEEEEEDQETPSR
AAKLS(1)EGS(1)QPAEEEEEDQETPSR
QEPT(1)QEHKQEEGQKQEEQEEQEEEGK
NIGFKVNS(1)K

### 1.1.3. Modified peptide sequences from Spectronaut

If the sequence data are obtained from Spectronaut, then users can find the modified peptide sequences in the Standard Report part of Spectronaut, for example, export the Peptide Quant results from the Pivot Report and extract the modified peptide sequences from the EG.ModifiedSequence column. The peptide sequences are like this:

Spectronaut_Outputs
_INS[Phospho (STY)]APSS[Phospho (STY)]PIK_
_MLISAVS[Phospho (STY)]PEIR_
_KINS[Phospho (STY)]APSS[Phospho (STY)]PIK_
_KINS[Phospho (STY)]APS[Phospho (STY)]SPIK_
_INSAPSS[Phospho (STY)]PIK_
_EGSQGEPWT[Phospho (STY)]PTANLK_
_EGSQGEPWPT[Phospho (STY)]ANLK_
_SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEEK_
_SHMSGSPGPGGSNT[Phospho (STY)]APSTPVIGGSDKPGMEEK_
_SS[Phospho (STY)]SS[Phospho (STY)]LLASPGHISVK_

## 1.2. Background data

Background data here means the protein sequences of a species (.fasta format). Users should use the same protein sequence file as the background database. For example, users can download the protein sequences from UniProt (<https://www.uniprot.org/>)<sup>3</sup>. The protein sequences like this:

```
>sp|Q64578|AT2A1_RAT Sarcoplasmic/endoplasmic reticulum calcium ATPase 1 OS=Rattus norvegicus OX=10116 GN=Atp2a1 PE=1 SV=1
MEAAHKSSTEECLSYFGVSETTGLTPDQVKRHLEKYGNELPAEEGKSLWELVVEQFEDL
LVRILLAAACISFVLAWFEEGETVTAFFVEPFVILLILIANAIVGVWQERNAENAEALK
EYEPFMGKVVYRADRSVQRIKARDIVPGDIVEVAVGDKVPADIRLSIKSTTLRVDSIL
TGESVSVIKHTDPVDPRAVNQDKNMFLSGTNIAAGKAVGIVATTGVSTEIGKIRDQMA
ATEQKTPQLQKLDFFGEQLSKVISLICVAVWLINIGHFNDPVMHGGSNFRGAIVYFKIAV
ALAVAAIPEGLPAVITTCALGTRRMMAKNNAIVRSLSVETLGCTSVICSDKTGTLTTNQ
MSVCKHFIIDKVDGICSLNEFSITGSTYAPEGEVLKNDKPVRAQQYDGLVELATICALC
NDSSLDNETKGVYKVGGEATETALTTLVEKMNVFNTVRSLSKVERANACNSVIRQLMK
KEFTLEFSRDRKSMVYCSPAKSSRAAVGNKMFVKGAPEGVIDRCNIVRVGTTTRVPLTGP
VKEKIMSIVKEWGTGRDTRLCLALATRDTPPKREEMVLDSSAKFMIEYEMDLTFVGVVGM
DPPRKVTGSIQLCRDAGIRVIMITGDNKGTAIAICRRIGIFSENEVADRAYTGREFDD
LPLAEQREACRACCFARVEPSHKSKIVEYLQSYDEITAMTGDGVNDAPALKKAEIGIAM
GGGTAVAKTASEMVLADDFNSTIVAAVEEGRAIYNMVKQFIRYLISNVGEVVCIFLTAA
LGLPEALIPVQLLWNLVTDGLPATALGFNPPDLIDMRPPSPKPELISGWLFFRYMAI
GGYVGAATVGAAAWFLYAEDGPHVSYHQLTHFMQCTEHNPEFDGLDCEVFEAPEPHMTA
LSVLVTIEMCNALNSLENQSLLRMPHVNINWLLGSIKLSMSLHFLILYVDPLPMIFKLR
ALDFTQHLWMLKISLPVIGLDELKFIARNYLEG
>sp|Q64568|AT2B3_RAT Plasma membrane calcium-transporting ATPase 3 OS=Rattus norvegicus OX=10116 GN=Atp2b3 PE=1 SV=2
MGDMANSSIEFHPKPQQREVPVHVGFGCTLAELRSLMELRGAEALQIQEAYGDVSGLC
RRLKTSPTGLADNTNLEKRRQIYQGNFIPPKQPKTFLQLVWEALQDVTLIILEVAAIV
SLGLSFYAPPGESEACGNVSGGADEGEAEAGWIEGAAILLSVICVVLVTAFFNDWSKEK
QFRGLQSRIEQEQKFTVIRNGQLLQVPVAAALVVGDIQVYKQDGLLPADGVLIGNDLKID
ESSLTGESDHRKSADKDPMLLSGTHVMESGSRMVVTAAGVNSQTGIIFTLTGAGGEEEE
KKDKKKGQDQGAHESSQTKAKKQDGAHMEIQLKSAEGGEMEEREKKANVPKKEKSVL
QGKLTKLAVQIGKAGLVMSAITVILLVLYFVETFVVDGRVNLAECTPVYVYQVFKFFII
GVTVLVAVPEGLPLAVTISLAYSVKKMMKDNLLVRHLDACETHGNATAICSDKTGTLTT
NRMTVVQSYLGDTHYKEIPAPSALTPKILDLVHAISINSAYTTKILPPEKEGALPRQVG
NKTCEALLGFIIDLKRDQPVREQIPEDQLYKVYTFNSVRKSMSTVIRMPDGGFRLFSKG
ASEILLKKCTNILLNSNGELRGFRPRDRDDHVKKIIIEPMACDGLRTICIAVYRDFSAIQEPD
WDNENEVVDLTCIAVVGIEDPVREVPPEAIRKQCRAGITVRHVTGDNINTARAIAAKCG
IIQPGEDFLCLEGKEFNRRIRNEKEIEQERLDKVMPLRLVARSPTDKHTLVKGIIDS
TTGEQRQVAVTGDGTNDGPALKKADVGFAHGIAGTDAKEASDIILTDNFTSIVKAVM
WGRNVYDSISKFLQFQLTVNVVAVIVAFGTACITQDSPLKAVQMLWNLIMDTFASLALA
TEPPTESLLLRKPYGRDKPLISRTHMKNILGHAVYQLTIIFTLLFVGLFFDIDSGRNP
LHSPPEHYTIIIFNTFVMQLFNEINARKIHGERNVFDGIFSNPIFCTIVLGTGFIQIVI
VQFGGKPFSCSPLSTEQLWLCLFVGVGELVWGQVIATIPTSQKCLKEAGHGPDKDEMTD
EELAEGLLEIDHAERELRGQILWFRGLNRIQTQMEVVSTFKRSQSGFQGAARRSSVLSQ
LHDVTNLSTPTHIRVVKAFRSSLYEGLEKPEKSCIHNFMATPEFLINDYTHNIPLIDDT
DVDENEERLRAPPPPPQNNNAIDSGIYLTTHATKSATSSAFSSRPGSPLHSMETSL
```

### 1.3. Download example datasets

If users want to download the example datasets to their own computer and check the data format locally, they can download them from here:

The screenshot displays the PTMoreR web application interface. On the left is a dark sidebar with navigation links: Home, Assembled Functions, and Help. The main content area features a 'User Guide' panel at the top, which lists six steps for processing data. Below this is an email input field with the placeholder text 'Enter your email here (Optional, please also check junk mail if possible):' and the email 'wukongomics@163.com'. The 'Step 1. Import Sequence Data.' panel is highlighted with a red arrow. It contains two radio buttons: 'Import modified sequences' (unselected) and 'Load example data' (selected). Below these is a blue button labeled '1. Download example data from Rat'. Further down, there are two dropdown menus: 'Data type:' set to 'Normal' and 'Central amino acid:' set to 'ST'. To the right of the 'Input Parameters' panel is the 'Results' panel, which shows a table of 'AnnotatedPeps' with five entries.

**User Guide**

There are 6 steps to process data in PTMoreR, please do it step by step:

- Step 1. Import Sequence Data.** In this part, users can upload their own peptide sequences with modification (e.g. phosphorylation). The example data were obtained from Rat and can be found when users click 'Load example data' below.
- Step 2. Pre-alignment.** This step aligns those peptide sequences with the background database (protein sequences) and force the modified sites/residues to be central sites, then users can get the standard peptide window sequences (e.g. 15 amino acid length by default).
- Step 3. Blast to Human.** This step will map the PTM site and protein sequences and identifiers between two species (One is that you chose in Step 1 and the other is that you want to blast to, which is Human by default).
- Step 4. Motif Enrichment and Plot.** This step will find overrepresented sequence motifs for uploaded peptides and blasted peptides respectively, then visualize them. Uploaded peptides here means those modified peptides uploaded directly by users. Blasted peptides here means those modified peptides mapped to human after blasting.
- Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.** This step will offer more flexible annotation based on kinase-substrate databases (e.g. PhosphoSitePlus) and network plots.
- Step 6. Interaction Plot.** This step will visualize the expression of modification sites on interacting proteins on the basis of protein-protein interaction data.

Enter your email here (Optional, please also check junk mail if possible):  
wukongomics@163.com

**Step 1. Import Sequence Data.**

**Input Parameters**

☐ Import modified sequences ☒ Load example data

1. Download example data from Rat

2. Data type:  
Normal

3. Central amino acid:  
ST

**Results**

Sequence data:  
Show 10 entries Search:

AnnotatedPeps	
1	GIGTAPPNTTPK
2	GIGTAPPNTATPIK
3	NGSAPK
4	KSERGMAAK
5	MINGHSADEESVR

First, select “Load example data” and the example data will be shown on the right panel interactively. Second, users can download the example data (the modified peptide sequences) with relative format (2. Data type: Normal, MaxQuant, Spectronaut) by clicking the corresponding button. The data sets are saved as .csv format and users can open them in other software, such as Excel.

## 2. Data processing

### Step 1. Import Sequence Data

After preparing proper data, users could click “Data processing” part, shown as below. Firstly, users could find a brief user guide (detailed manual can be found in the Help part), which describes six basic steps in this tool. In addition, users can click the top-right corner (-/+) to collapse or expand the contents. Secondly, users can type in their e-mails and PTMoreR will send the blasted results to the e-mail. This is optional. And if users run this tool locally, this function is removed. Thirdly, users can process their data step by step. There are two main panels: right, *parameters panel*, users can adjust parameters here; left, *results panel*, many results after users set the parameters will be shown here and users can also download these results. users can also click the top-right corner (-/+) to collapse or expand the contents.

The screenshot displays the PTMoreR web application interface. On the left is a dark sidebar with navigation links: Home, Assembled Functions, and Help. The main content area features a 'User Guide' panel at the top, which lists six steps for data processing. Below the guide is an email input field with the placeholder text 'Enter your email here (Optional, please also check junk mail if possible):' and a text box containing 'wukongomics@163.com'. The interface then presents a list of six processing steps, each with an 'Input Parameters' button and a 'Results' button. Red arrows point to the expand/collapse icons (+/-) on the 'Input Parameters' and 'Results' buttons for Step 1. The steps are:

- Step 1. Import Sequence Data.
- Step 2. Pre-alignment.
- Step 3. Blast to Human.
- Step 4. Motif Enrichment and Plot.
- Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.
- Step 6. Interaction Plot.

In the first step, users should upload data here or load the example data with the above data formats. By default, we use the example data to show results of every step.

### Step 1.1. Uploading data.

When users prepare their data (the modified peptide sequences and protein background data), they can upload these data from here:

Step 1. Import Sequence Data.

The figure displays two panels from the PTMoreR web application. The left panel, titled 'Input Parameters', contains two main sections: 'Import modified sequences' (selected) and 'Load example data'. Under 'Import modified sequences', there are radio buttons for 'A. Upload' (selected) and 'B. Paste'. Below this, '1. File format:' has radio buttons for '.xlsx', '.xls', and '.csv/txt' (selected). '1.1. Import your data:' features a 'Browse...' button and the text 'No file selected'. '2. Data type:' is a dropdown menu set to 'Normal'. '3. Central amino acid:' is a text input field containing 'STY'. '4. Label of modification:' is a text input field containing '#'. '5. Width:' is a text input field containing '7'. '6. Select or upload the protein sequences (.fasta file):' has radio buttons for '6.1. Select' (selected) and '6.2. Upload'. At the bottom, 'Species:' is a dropdown menu. The right panel, titled 'Results', shows 'Sequence data:' with a 'Show 10 entries' dropdown and a 'Search:' input field. Below this is a table with one entry: '1 PTMoreR detects that you did not upload your data. Please upload the sequence data, or load the example data to check first.' The table has a 'Showing 1 to 1 of 1 entries' footer and 'Previous', '1', and 'Next' navigation buttons.

In the *parameters panel* of “*Step 1: Import Sequence Data*”, there are two choices:

*a. Import modified sequences.* When users choose this option, they can choose “*A. Upload*” to upload their own data (the modified peptide sequences) here. Users should select the right format (*1. File format: .csv, .txt, .xlsx, .xls*) based on their data and then click “Browse” button (*1.1. Import your data*) to import the data;

In the *results panel* of “*Step 1: Import Sequence Data*”, if users don’t upload their data, here will show “*PTMoreR detects that you did not upload your data. Please upload the sequence data, or load the example data to check first.*” to warn users.



Users can also directly paste their sequences into *PTMoreR* by choosing “*B. Paste*” like below:

Step 1. Import Sequence Data.

Input Parameters

☒ Import modified sequences

☐ Load example data

A. Upload

B. Paste

1. Paste your data here:

PPQLTPVLLSPAPUSIPSSUSFHK  
NTVNGTGT#PVHISTLQVGETR  
RT#S#PQQT#R  
S#PMPMPENAE#DALLQFTAEFSSR  
TPS#GEFLER

2. Data type:

Normal

3. Central amino acid:

STY

4. Label of modification:

#

5. Width:

7

6. Select or upload the protein sequences (.fasta file):

☒ 6.1. Select

☐ 6.2. Upload

Species:

Results

Sequence data:

Show

10

entries

Search:

	Input_Seqs
1	GIGT#PPNTTPIK
2	GIGT#PPNT#T#PIK
3	NGS#PEIK
4	KS#ERGMAAK
5	MNGHS#DEESVR
6	RQIDS#S#EDEDEDYDNDKR
7	RY#S#GS#DS#DS#ISER
8	RY#S#GS#DS#DS#ISER
9	RY#S#GS#DS#DS#ISER
10	KRPY#S#S#FS#NGK

Showing 1 to 10 of 30 entries

Previous

1

2

3

Next

*b. Load example data.* As described in part 1.3 above, users can choose this option and download the example data to check them locally.

Step 1. Import Sequence Data.

Input Parameters

☐ Import modified sequences

☒ Load example data

1. Download example data from Rat

2. Data type:

Normal

3. Central amino acid:

STY

4. Label of modification:

#

5. Width:

7

Example species:

10116-Rattus norvegicus (Rat)

Results

Sequence data:

Show

10

entries

Search:

	AnnotatedPeps
1	GIGT#PPNTTPIK
2	GIGT#PPNT#T#PIK
3	NGS#PEIK
4	KS#ERGMAAK
5	MNGHS#DEESVR
6	RQIDS#S#EDEDEDYDNDKR
7	RY#S#GS#DS#DS#ISER
8	RY#S#GS#DS#DS#ISER
9	RY#S#GS#DS#DS#ISER
10	KRPY#S#S#FS#NGK

Showing 1 to 10 of 520 entries

Previous

1

2

3

4

5

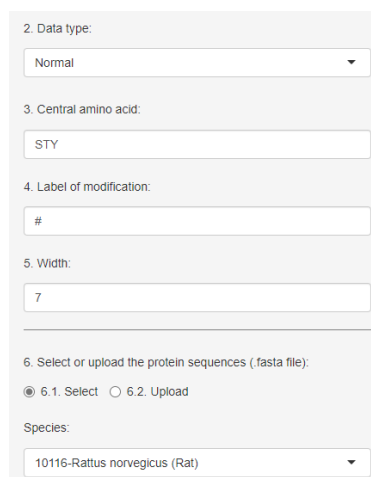
...

52

Next

## Step 1.2. Parameters

There are some basic parameters that users can change based on their own data, shown as below:



The screenshot shows a web form with the following fields and options:

- 2. Data type: A dropdown menu with "Normal" selected.
- 3. Central amino acid: A text input field containing "STY".
- 4. Label of modification: A text input field containing "#".
- 5. Width: A text input field containing "7".
- 6. Select or upload the protein sequences (.fasta file): Radio buttons for "6.1. Select" (selected) and "6.2. Upload".
- Species: A dropdown menu with "10116-Rattus norvegicus (Rat)" selected.

2. *Data type*: The original post-translational modification data obtained from which kind of search software. If you have processed the PTM data with standard format (e.g. NPT#Y#GSWFTEK), you should choose the "Normal", otherwise, if your PTM data are obtained from MaxQuant or Spectronaut, you should choose the relative type. Shown also as part 1.3 above.

3. *Central amino acid*: The central residue that users want to analyze, for example, phosphorylation motif analysis, can center on phosphorylated S, T or Y residues. If they want to analyze multi motif sites, here should be "STY".

4. *Label of modification*: The label represents modification, users can use some label they like, such as "#", "@", in which "#" is recommended. Here is an example:



The diagram illustrates the transformation of a protein motif. On the left, the motif is "EGSQGEPWT[Phospho (STY)]PTANLK". A red arrow points to the right, where the motif is shown as "EGSQGEPWT#PTANLK". This demonstrates how the PTM label is replaced by a specific character (#) in the final motif.

5. *Width*: It is the number of left/right side characters of the central residue.

6. *Select or upload the protein sequences (.fasta file)*: If users want to use the default database, they just select relative species. By default, PTMoreR integrates 27074 species and download automatically the protein sequences (.fasta format) from UniProt database (<https://www.uniprot.org>).

6. Select or upload the protein sequences (.fasta file):

☒ 6.1. Select ☐ 6.2. Upload

Species:

- 10116-Rattus norvegicus (Rat)
- 9606-Homo sapiens (Human)
- 10090-Mus musculus (Mouse)
- 60711-Chlorocebus sabaeus (Green monkey) (Cercopithecus sabaeus)
- 41-Stigmatella aurantiaca
- 52-Chondromyces crocatus
- 63-Vitreoscilla filiformis

Optionally, if users want to upload their own .fasta file (choose 6.2. *Upload*), they should type in the taxonomic identifier in “6.2.1. *Please type in taxonomic identifier of uploaded species*”, for example, rat’s taxonomic identifier is 10116.

6. Select or upload the protein sequences (.fasta file):

☐ 6.1. Select ☒ 6.2. Upload

6.2.1. Please type in taxonomic identifier of uploaded species:

6.2.2. Please upload your fasta file:

No file selected

After typing in the taxonomic identifier, users can also upload their own .fasta file (6.2.2. *Please upload your fasta file*), no species limits here, but the calculation time would be longer.


## Step 2. Pre-alignment

This step means align those peptide sequences with the background database (protein sequences) and force the modified sites/residues to be central sites, then users can get the standard peptide window sequences.

### Step 2. Pre-alignment.



**Input Parameters**

☐ 1. Check if containing some regular sequence?



**Results**

Pre-alignment table:

 Download  Result description

Show  entries Search:

	Uploaded Peptides	Protein ID from Database	Center amino acid	Modified Amino Acid Site in Protein Sequence	Combined Protein Identifi
1	GIGT#PPNTTPIK	D4A9J4	T	110	D4A9J4_T110
2	GIGT#PPNT#T#PIK	D4A9J4	T;T;T	110;114;115	D4A9J4_T110;D4A9J4_T114
3	NGS#PEIK	D4A9J4	S	121	D4A9J4_S121
4	KS#ERGMAAK	D4A9J4	S	659	D4A9J4_S659
5	MNGHS#DEESVR	D4AAG9	S	5	D4AAG9_S5
6	RQIDS#S#EDEDEDYDNDKR	D4AAG9	S;S	214;215	D4AAG9_S214;D4AAG9_S215
7	RYS#GS#DS#DS#ISER	D4AAG9	S;S;S;S	1094;1096;1098;1100	D4AAG9_S1094;D4AAG9_S1096;D4AAG9_S1098;D4AAG9_S1100
8	RY#S#GS#DS#DS#ISER	D4AAG9	Y;S;S;S;S	1093;1094;1096;1098;1100	D4AAG9_Y1093;D4AAG9_S1094;D4AAG9_S1096;D4AAG9_S1098;D4AAG9_S1100
9	RY#S#GS#DS#DS#ISER	D4AAG9	Y;S;S;S;S	1093;1094;1096;1098;1100;1102	D4AAG9_Y1093;D4AAG9_S1094;D4AAG9_S1096;D4AAG9_S1098;D4AAG9_S1100;D4AAG9_S1102
10	KRPY#S#S#F#WNGK	D4AAG9	Y;S;S;S	1577;1578;1579;1581	D4AAG9_Y1577;D4AAG9_S1578;D4AAG9_S1579;D4AAG9_S1581

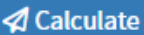
Showing 1 to 10 of 496 entries Previous 1 2 3 4 5 ... 50 Next

## Step 2.1. Parameters

### Step 2. Pre-alignment.

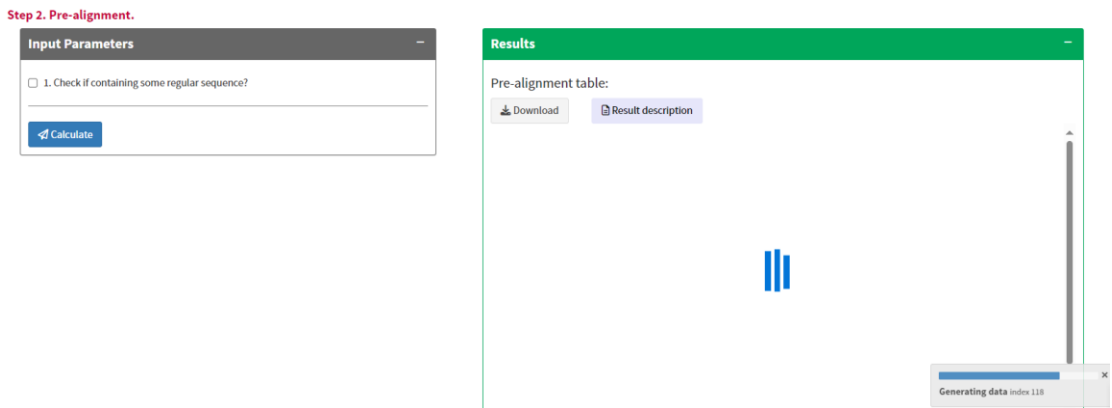
**Input Parameters**

☐ 1. Check if containing some regular sequence?



1. *Check if containing some regular sequence*: if users want to check whether the aligned peptides contain some specific sequences, for example, you want to find those peptides whose 3th and 5th position are R (arginine), then you can select this parameter and type in a simple regular expression, like “`^\\w{2}R\\w{1}R`” (more details can be found here: [https://en.wikipedia.org/wiki/Regular\\_expression](https://en.wikipedia.org/wiki/Regular_expression)). Otherwise, you just unselect it.

Then, you can click the “Calculate” button, it should be like this:



A process bar will appear in the bottom right corner to tell users where it goes.

## Step 2.2 Results

### Step 2.2.1. Alignment results

Step 2. Pre-alignment.

Input Parameters

☐ 1. Check if containing some regular sequence?

Calculate

Results

Pre-alignment table:

Download Result description

Show 10 entries Search:

	Uploaded Peptides	Protein ID from Database	Center amino acid	Modified Amino Acid Site in Protein Sequence	Combined Protein Identifi
1	GIGT#PPNTTPIK	D4A9J4	T	110	D4A9J4_T110
2	GIGT#PPNT#T#PIK	D4A9J4	T,T;T	110;114;115	D4A9J4_T110;D4A9J4_T114
3	NGS#PEIK	D4A9J4	S	121	D4A9J4_S121
4	KS#ERGMAAK	D4A9J4	S	659	D4A9J4_S659
5	MNGHS#DEESVR	D4AAG9	S	5	D4AAG9_S5
6	RQIDS#S#EDEDDYDNDKR	D4AAG9	S;S	214;215	D4AAG9_S214;D4AAG9_S215
7	RY#SGS#DS#DS#ISER	D4AAG9	S;S;S;S	1094;1096;1098;1100	D4AAG9_S1094;D4AAG9_S1096;D4AAG9_S1098;D4AAG9_S1100
8	RY#SGS#DS#DS#ISER	D4AAG9	Y;S;S;S;S	1093;1094;1096;1098;1100	D4AAG9_Y1093;D4AAG9_S1094;D4AAG9_S1096;D4AAG9_S1098;D4AAG9_S1100
9	RY#SGS#DS#DS#ISER	D4AAG9	Y;S;S;S;S	1093;1094;1096;1098;1100;1102	D4AAG9_Y1093;D4AAG9_S1094;D4AAG9_S1096;D4AAG9_S1098;D4AAG9_S1100;D4AAG9_S1102
10	KRPY#S#S#F#HNGK	D4AAG9	Y;S;S;S	1577;1578;1579;1581	D4AAG9_Y1577;D4AAG9_S1578;D4AAG9_S1579;D4AAG9_S1581

Showing 1 to 10 of 496 entries Previous 1 2 3 4 5 ... 50 Next

Users can click “Result description” button to read the introduction about every column, shown as below:

**Step 2. Pre-alignment**

**Input Parameters**

☐ 1. Check if containing regular sequence

**Calculate**

Pre-alignment result description:

1. Uploaded Peptides: this column contains those peptides users upload.
2. Protein ID from Database: provide the protein name containing this peptide from the fasta file the user uploaded.
3. Center amino acid: the central amino acid in the aligned peptide.
4. Modified Amino Acid Position in Protein Sequence: the position of modified amino acid in the protein sequence.
5. Combined Protein Identifier: Combining the protein ID, Center.amino.acid and PROindex.from.Database together with '\_\_'.
6. Aligned Standard Peptides: the aligned standard peptides. Note for multiple modification sites or types, the column provides peptides with all the sites respectively centered.
7. Peptide Skeleton: the peptide skeleton.
8. Main Modification Site in Uploaded Peptides: the position of the main modified amino acid in the peptide, for example, if users upload their peptides containing Class I phosphorylation sites with high confidence, such as 'TSLWNPT#Y@GSWFTEK', then this software will recognize '#' as Class I phosphorylation site and '@' as non-Class I phosphorylation site by default, so the Pep.main.index will be 7.
9. All Modification Site in Uploaded Peptides: the position of all modified amino acid in the peptide. As the example in Pep.main.index, the Pep.all.index will be 7;8.
10. Regular Sequence Inclusion Check: whether containing the sequences that match the regular expression (see above), if true, marked with 'Yes', otherwise, 'No'. This column only appears when users choose the parameter--- Check if containing some regular sequence.

**Cancel**

4	KS#ERGMMAAK	D4A9J4	S	659
5	MNGHS#DEESVR	D4AAG9	S	5
6	RQIDS#S#EDEDEDYDNDKR	D4AAG9	S,S	214;215
7	RY#S#GS#DS#DS#ISER	D4AAG9	S,S,S,S	1094;1096;1098;1100
8	RY#S#GS#DS#DS#ISER	D4AAG9	Y,S,S,S,S	1093;1094;1096;1098;1100
9	RY#S#GS#DS#DS#ISER	D4AAG9	Y,S,S,S,S,S	1093;1094;1096;1098;1100;1102

1. **Uploaded Peptides:** this column contains those peptides users upload.
2. **Protein ID from Database:** provide the protein name containing this peptide from the fasta file the user uploaded.
3. **Center amino acid:** the central amino acid in the aligned peptide.
4. **Modified Amino Acid Position in Protein Sequence:** the position of modified amino acid in the protein sequence.
5. **Combined Protein Identifier:** Combining the protein ID, Center.amino.acid and PROindex.from.Database together with '\_\_'.
6. **Aligned Standard Peptides:** the aligned standard peptides. Note for multiple modification sites or types, the column provides peptides with all the sites respectively centered.
7. **Peptide Skeleton:** the peptide skeleton.
8. **Main Modification Site in Uploaded Peptides:** the position of the main modified amino acid in the peptide, for example, if users upload their peptides containing Class I phosphorylation sites with high confidence, such as 'TSLWNPT#Y@GSWFTEK', then this software will recognize '#' as Class I phosphorylation site and '@' as non-Class I phosphorylation site by default, so the Pep.main.index will be 7.
9. **All Modification Site in Uploaded Peptides:** the position of all modified amino acid in the peptide. As the example in Pep.main.index, the Pep.all.index will be 7;8.
10. **Regular Sequence Inclusion Check:** whether containing the sequences that match the regular expression (see above), if true, marked with 'Yes', otherwise, 'No'. This column only appears when users choose the parameter--- Check if containing some regular sequence.

### Step 3. Blast to Human (or other species)

This step will map the PTM site and protein sequences and identifiers between the uploaded species and the other species.

#### Step 3.1. Parameters

##### Step 3. Blast to Human (or other species).

**Input Parameters**

1. Expectation value (E) threshold:  
0.00001

2. The criterion for the best BLAST hit:  
Percentage

3. Central amino acid matching degree:  
Fuzzy matching

☒ 4. Whether setting sequence windows similarity (SWS)?

4.1. SWS threshold:  
8

☐ 5. Whether setting BLOSUM50 Score?

☐ 6. Whether blast to other species?

[Calculate](#)

- 1. Expectation value (E) threshold:** Expectation value (E) threshold for saving hits.
- 2. The criterion for best last hit:** This tool performs a BLAST search between query and subject sequences and returns only the best hit based on the selected criterion. "Percentage" means If e-values are identical then the hit with the largest matching percentage is chosen. "Longest alignment length" means If e-values are identical then the hit with the longest alignment length is chosen.
- 3. Central amino acid matching degree:** The matching degree of central amino acids (CAAs) when the uploaded peptides are blasted to Human protein sequences.
  - 1. Exact matching:** The CAAs are same, for example, the CAA is "S" in the uploaded peptides and the CAA is also "S" in the blasted sequence.
  - 2. Fuzzy matching:** Only for phosphorylation, not for other modification type. For example, the CAA is "S" in the uploaded peptides and the CAA could be "S", "T", or "Y" in the blasted sequence.
  - All:** Reporting all results.
- 4. Whether setting sequence windows similarity (SWS)?** This parameter means whether users want to set the threshold of SWS. If true, users can set **4.1. SWS threshold:** The similarity of sequence windows between the uploaded peptides and the blasted peptides. For example, there are 15 amino acids in one sequence window, 8 here means there are 8 amino acids are exactly same (amino acids names and positions in both windows are all the same).
- 5. Whether setting BLOSUM50 Score?** This parameter means whether users want to set the threshold of BLOSUM50 Score. If true, users can set **5.1. BLOSUM50 Score threshold:**

BLOSUM50 means that the matrix was built using blocks of aligned sequences that had no more than 50% identity, which is used to score alignments between evolutionarily divergent protein sequences. The default BLOSUM50 Score is 0, which means PTMoreR filters the blasted results with score  $\geq 0$ .

6. *Whether blast to other species?* If false (by default), this tool will blast to Human automatically. Otherwise, users can select another species below they want to blast to. Please note, this may take a quite long time, we suggest to use the local version.

## Step 3.2. Results

After setting proper parameters, users can click “Calculate” button and the results will be shown in the right panel.

### Step 3.2.3. Final blast results

When users click “3. Final blast results” here, this tool will calculate the final result, like below:

Step 3. Blast to Human (or other species).

Input Parameters

1. Expectation value (E) threshold:

2. The criterion for the best BLAST hit:

3. Central amino acid matching degree:

☒ 4. Whether setting sequence windows similarity (SWS)?

4.1. SWS threshold:

☐ 5. Whether setting BLOSUM50 Score?

☐ 6. Whether blast to other species?

Results

Final blasted table:

Show  entries

Search:

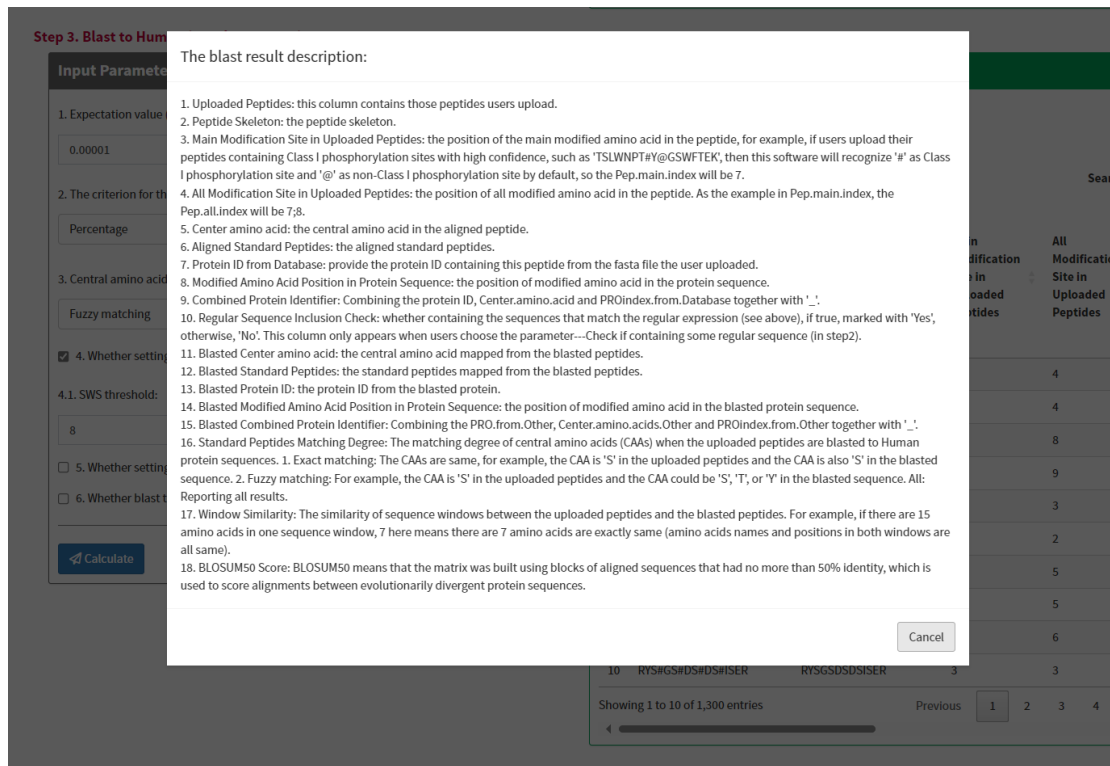
	Uploaded Peptides	Peptide Skeleton	Main Modification Site in Uploaded Peptides	All Modification Site in Uploaded Peptides	Center amino acid	Aligned Standard Peptides
1	GIGT#PPNNTTPIK	GIGTPPNNTTPIK	4	4	T	QEVKGIGTP
2	GIGT#PPNT#T#PIK	GIGTPPNNTTPIK	4	4	T	QEVKGIGTP
3	GIGT#PPNT#T#PIK	GIGTPPNNTTPIK	8	8	T	GIGTPPNNTT
4	GIGT#PPNT#T#PIK	GIGTPPNNTTPIK	9	9	T	IGTTPNNTTP
5	NGS#PEIK	NGSPEIK	3	3	S	TTPIKNGSP
6	KS#ERGMAAK	KSERGMAAK	2	2	S	ASVSSKKSE
7	MNGHS#DEESVR	MNGHSDEESVR	5	5	S	___MNGHSI
8	RQIDS#S#EDEDEDYDNDKR	RQIDSSSEDEDYDNDKR	5	5	S	QKKRQIDSS
9	RQIDS#S#EDEDEDYDNDKR	RQIDSSSEDEDYDNDKR	6	6	S	KKRQIDSS
10	RYSGS#DS#DS#ISER	RYSGSDSDSISER	3	3	S	RSRSRRYSG

Showing 1 to 10 of 1,300 entries

Previous
2
3
4
5
...
130
Next

Users can click “Result description” button to read the introduction about every column, shown as below:





1. **Uploaded Peptides:** this column contains those peptides users upload.
2. **Peptide Skeleton:** the peptide skeleton.
3. **Main Modification Site in Uploaded Peptides:** the position of the main modified amino acid in the peptide, for example, if users upload their peptides containing Class I phosphorylation sites with high confidence, such as 'TSLWNPT#Y@GSWFTEK', then this software will recognize '#' as Class I phosphorylation site and '@' as non-Class I phosphorylation site by default, so the Pep.main.index will be 7.
4. **All Modification Site in Uploaded Peptides:** the position of all modified amino acid in the peptide. As the example in Pep.main.index, the Pep.all.index will be 7;8.
5. **Center amino acid:** the central amino acid in the aligned peptide.
6. **Aligned Standard Peptides:** the aligned standard peptides.
7. **Protein ID from Database:** provide the protein ID containing this peptide from the fasta file the user uploaded.
8. **Modified Amino Acid Position in Protein Sequence:** the position of modified amino acid in the protein sequence.
9. **Combined Protein Identifier:** Combining the protein ID, Center.amino.acid and PROindex.from.Database together with '.\_'.
10. **Regular Sequence Inclusion Check:** whether containing the sequences that match the regular expression (see above), if true, marked with 'Yes', otherwise, 'No'. This column only appears when users choose the parameter---Check if containing some regular sequence (in step2).
11. **Blasted Center amino acid:** the central amino acid mapped from the blasted peptides.

12. *Blasted Standard Peptides*: the standard peptides mapped from the blasted peptides.
13. *Blasted Protein ID*: the protein ID from the blasted protein.
14. *Blasted Modified Amino Acid Position in Protein Sequence*: the position of modified amino acid in the blasted protein sequence.
15. *Blasted Combined Protein Identifier*: Combining the PRO.from.Other, Center.amino.acids.Other and PROindex.from.Other together with '\_ '.
16. *Standard Peptides Matching Degree*: The matching degree of central amino acids (CAAs) when the uploaded peptides are blasted to Human protein sequences. 1. Exact matching: The CAAs are same, for example, the CAA is 'S' in the uploaded peptides and the CAA is also 'S' in the blasted sequence. 2. Fuzzy matching: For example, the CAA is 'S' in the uploaded peptides and the CAA could be 'S', 'T', or 'Y' in the blasted sequence. All: Reporting all results.
17. *Window Similarity*: The similarity of sequence windows between the uploaded peptides and the blasted peptides. For example, if there are 15 amino acids in one sequence window, 7 here means there are 7 amino acids are exactly same (amino acids names and positions in both windows are all same).
18. *BLOSUM50 Score*: BLOSUM50 means that the matrix was built using blocks of aligned sequences that had no more than 50% identity, which is used to score alignments between evolutionarily divergent protein sequences.

## Step 4. Motif Enrichment and Plot

This step will find overrepresented sequence motifs as we previous described for single specie<sup>4</sup>.

### Step 4.1. Parameters

#### Step 4. Motif Enrichment and Plot.

**Input Parameters**

☐ 1. Only analyze regular sequences?


2. Minimum number:

3. P-value threshold:

4. Motif index for plot:

☒ 5. Equal height or not?

6. Figure height:

 Calculate

1. *Only analyze regular sequences?* If true, this tool will only analyze the peptides contain some specific sequences (regular sequences) obtain from “3. Pre-alignment” part above.

2. *Minimum number:* This threshold refers to the minimum number of times you wish each of your extracted motifs to occur in the data set.

3. *P-value threshold:* The p-value threshold for the binomial probability. This is used for the selection of significant residue/position in the motif.

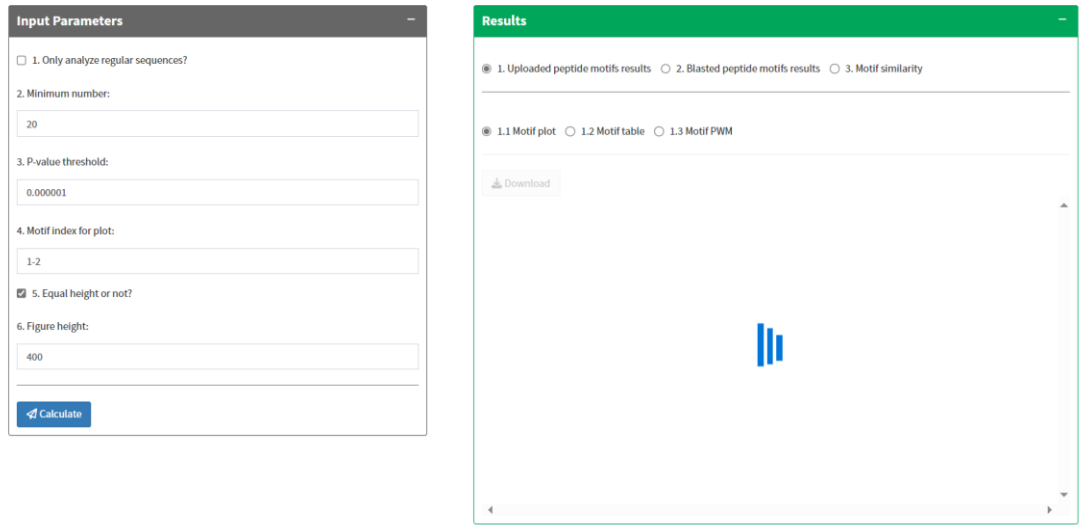
4. *Motif index for plot:* Which motif would be plotted. If users only type in one number, it will plot the relative motif. If users type in "1-10", it will plot the 1th to 10th motifs.

5. *Equal height or not?* Whether all residues in the figure have equal height. Default is false.

6. *Figure height:* The height of the figure.

Then, users can click the “Calculate” button, this tool will process motif enrichment and plot those motifs:

Step 4. Motif Enrichment and Plot.

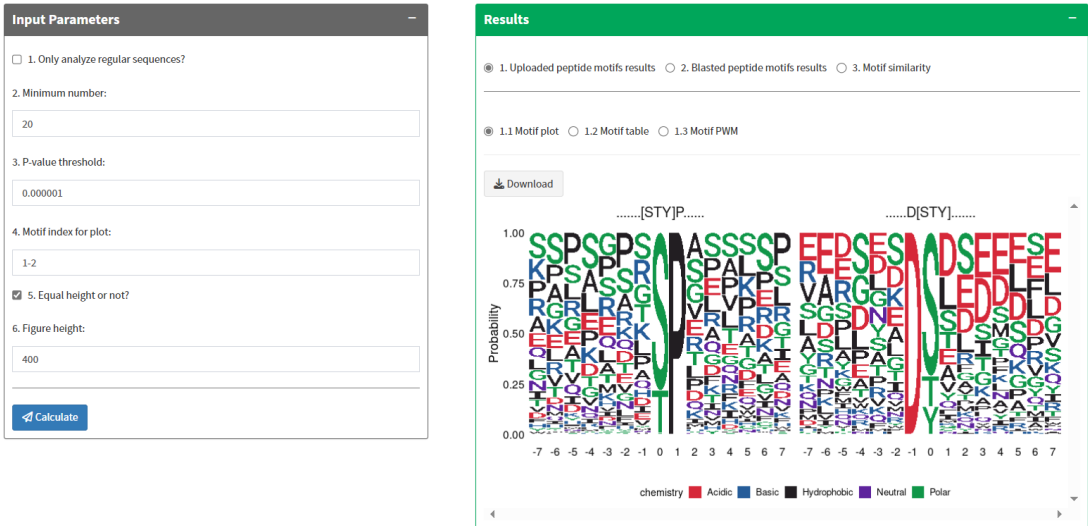


Step 4.2. Results

Step 4.2.1. Uploaded peptide motifs results

All results here are obtained from the uploaded peptides. When users click “1.1. Motif plot”, the motif plot is shown as below:

Step 4. Motif Enrichment and Plot.



When users click “1.2. Motif table”, the motif table is shown as below:

**Step 4. Motif Enrichment and Plot.**

### Input Parameters

☐ 1. Only analyze regular sequences?

2. Minimum number:  
20

3. P-value threshold:  
0.000001

4. Motif index for plot:  
1-2

☒ 5. Equal height or not?

6. Figure height:  
400

Calculate

### Results

☒ 1. Uploaded peptide motifs results ☐ 2. Blasted peptide motifs results ☐ 3. Motif similarity

☐ 1.1 Motif plot ☒ 1.2 Motif table ☐ 1.3 Motif PWM

Download Result description

Show 10 entries Search:

	motif	score	fg.matches	fg.size	bg.matches	bg.size	fold.increase	
1	.....[STY]P.....	307.652655568589	172	834	119500	1819055	3.13935422373398	(
2	.....D[STY].....	9.43207016118089	73	662	84730	1699555	2.21188675383652	(
3	...R...[STY].....	8.35777190019163	70	589	90201	1614825	2.12763368375427	I

Showing 1 to 3 of 3 entries Previous 1 Next

Every column in this table means (Users can also check this by clicking “Result description” button):

1. *motif*: the overrepresented motif.
2. *score*: the motif score, which is calculated by taking the sum of the negative log probabilities used to fix each position of the motif. Higher motif scores typically correspond to motifs that are more statistically significant as well as more specific.
3. *fg.matches*: frequency of sequences matching this motif in the foreground set.
4. *fg.size*: total number of foreground sequences.
5. *bg.matches*: frequency of sequences matching this motif in the background set.
6. *bg.size*: total number of background sequences.
7. *fold.increase*: An indicator of the enrichment level of the extracted motifs. Specifically, it is calculated as (foreground matches/foreground size)/(background matches/background size).
8. *Enrich.seq*: those peptides are overrepresented in this motif.
9. *Enrich.pro*: those proteins in which the peptides exist from Enrich.seq.

When users click “1.3. Motif PWM”, PWM here means position weight matrix, the motif table is shown as below:

Step 4. Motif Enrichment and Plot.

Input Parameters

☐ 1. Only analyze regular sequences?

2. Minimum number:

3. P-value threshold:

4. Motif index for plot:

☒ 5. Equal height or not?

6. Figure height:

Calculate

Results

1. Uploaded peptide motifs results

2. Blasted peptide motifs results

3. Motif similarity

1.1 Motif plot

1.2 Motif table

1.3 Motif PWM

Please select one motif to view its PWM (Position weight matrix):  

.....[STY]P.....

Download

Show 10 entries

Search:

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
A	0.0828402	0.0941176	0.0705882	0.1169591	0.0581395	0.0872093	0.0523256	0	0	0.128655
C	0	0.0117647	0.0058824	0	0.0116279	0.0116279	0	0	0	0
D	0.0295858	0.0352941	0.0411765	0.0643275	0.0232558	0.0581395	0.0406977	0	0	0.0526316
E	0.0710059	0.0705882	0.0705882	0.0877193	0.0930233	0.0523256	0.0348837	0	0	0.0818713
F	0.0236686	0.0058824	0.0117647	0.0116959	0.0174419	0.0290698	0	0	0	0.0116959
G	0.0532544	0.0823529	0.0705882	0.0526316	0.1162791	0.0406977	0.0988372	0	0	0.0994152
H	0.0236686	0.0176471	0.0176471	0.005848	0.0290698	0.0174419	0.0406977	0	0	0.0116959
I	0.0414201	0.0176471	0.0411765	0.0233918	0.0232558	0.0174419	0.0348837	0	0	0.0233918
K	0.1065089	0.0764706	0.0235294	0.0818713	0.0406977	0.0639535	0.0872093	0	0	0.0409357
L	0.0532544	0.0647059	0.0823529	0.0994152	0.0639535	0.0348837	0.0755814	0	0	0.0643275

Showing 1 to 10 of 20 entries

Previous12Next

Users can select one motif to view its PWM. The motif here are obtained from “1.2. Motif table”, like below:

Please select one motif to view its PWM (Position weight matrix):  

.....[STY]P.....

.....[STY]P.....  
.....D[STY].....  
.....R...[STY].....

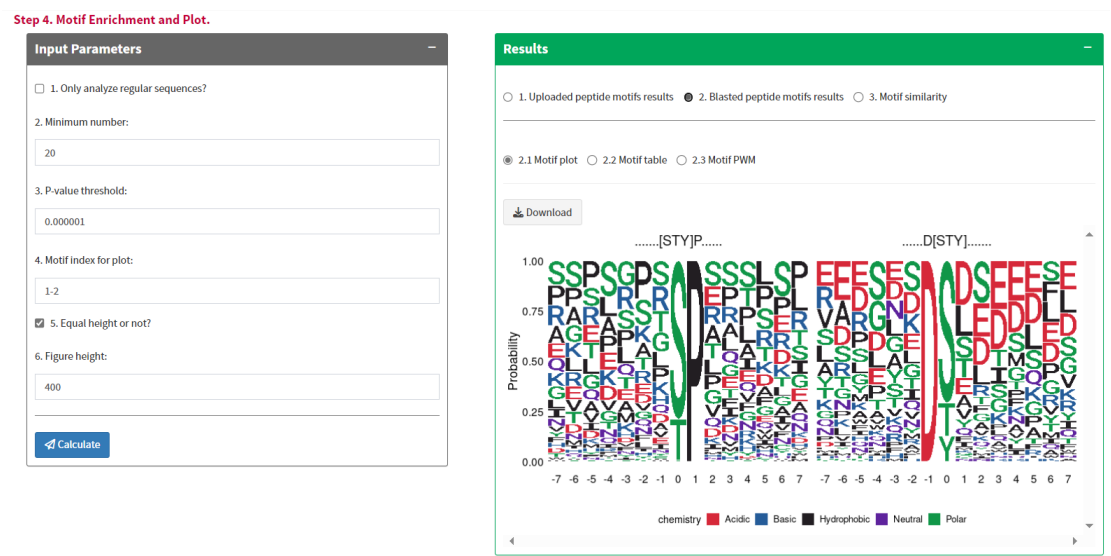
	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
A	0.0828402	0.0941176	0.0705882	0.1169591	0.0581395	0.0872093	0.0523256	0	0	0.128655
C	0	0.0117647	0.0058824	0	0.0116279	0.0116279	0	0	0	0
D	0.0295858	0.0352941	0.0411765	0.0643275	0.0232558	0.0581395	0.0406977	0	0	0.0526316
E	0.0710059	0.0705882	0.0705882	0.0877193	0.0930233	0.0523256	0.0348837	0	0	0.0818713
F	0.0236686	0.0058824	0.0117647	0.0116959	0.0174419	0.0290698	0	0	0	0.0116959
G	0.0532544	0.0823529	0.0705882	0.0526316	0.1162791	0.0406977	0.0988372	0	0	0.0994152
H	0.0236686	0.0176471	0.0176471	0.005848	0.0290698	0.0174419	0.0406977	0	0	0.0116959
I	0.0414201	0.0176471	0.0411765	0.0233918	0.0232558	0.0174419	0.0348837	0	0	0.0233918
K	0.1065089	0.0764706	0.0235294	0.0818713	0.0406977	0.0639535	0.0872093	0	0	0.0409357
L	0.0532544	0.0647059	0.0823529	0.0994152	0.0639535	0.0348837	0.0755814	0	0	0.0643275

Showing 1 to 10 of 20 entries

Previous12Next

Step 4.2.2. Blasted peptide motifs results

All results here are obtained from the blasted peptides. Similar to the above, when users click “2.1. Motif plot”, the motif plot is shown as below:



When users click “2.2. Motif table”, the motif table is shown as below:

Step 4. Motif Enrichment and Plot.

**Input Parameters**

☐ 1. Only analyze regular sequences?

2. Minimum number:  
20

3. P-value threshold:  
0.000001

4. Motif index for plot:  
1-2

☒ 5. Equal height or not?

6. Figure height:  
400

Calculate

**Results**

☐ 1. Uploaded peptide motifs results ☒ 2. Blasted peptide motifs results ☐ 3. Motif similarity

☐ 2.1 Motif plot ☒ 2.2 Motif table ☐ 2.3 Motif PWM

Download Result description

Show 10 entries Search:

	motif	score	fg.matches	fg.size	bg.matches	bg.size	fold.increase
1	.....[STY]P.....	307.652655568589	152	757	118694	1769061	2.99269012717649
2	.....D[STY].....	10.9322397688032	72	605	81754	1650367	2.40241838193476
3	....R..[STY].....	6.09512944444443	59	533	87948	1568613	1.9743068157877

Showing 1 to 3 of 3 entries Previous 1 Next

Every column means the same thing as above (Users can also check this by clicking “Result description” button).

Step 4. Motif Enrichment

Input Parameters

☐ 1. Only analyze regular sequences?

2. Minimum number:  
20

3. P-value threshold:  
0.000001

4. Motif index for plot:  
1-2

☒ 5. Equal height or not?

6. Figure height:  
400

Calculate

Blasted motif Enrichment result description:

1. motif: the overrepresented motif.

2. score: the motif score, which is calculated by taking the sum of the negative log probabilities used to fix each position of the motif. Higher motif scores typically correspond to motifs that are more statistically significant as well as more specific.

3. fg.matches: frequency of sequences matching this motif in the foreground set.

4. fg.size: total number of foreground sequences.

5. bg.matches: frequency of sequences matching this motif in the background set.

6. bg.size: total number of background sequences.

7. fold.increase: An indicator of the enrichment level of the extracted motifs. Specifically, it is calculated as (foreground matches/foreground size)/(background matches/background size).

8. Enrich.seq: those peptides are overrepresented in this motif.

9. Enrich.pro: those proteins in which the peptides exist from Enrich.seq.

Cancel

Results

hg.matches

hg.size

fold.increase

1

2

3

Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.

Input Parameters

Results

Step 6. Interaction Plot.

Input Parameters

Results

Just as above, when users click “2.3. Motif PWM”, PWM here means position weight matrix, the motif table is shown as below:

Step 4. Motif Enrichment and Plot.

Input Parameters

☐ 1. Only analyze regular sequences?

2. Minimum number:  
20

3. P-value threshold:  
0.000001

4. Motif index for plot:  
1-2

☒ 5. Equal height or not?

6. Figure height:  
400

Calculate

Results

☐ 1. Uploaded peptide motifs results

☒ 2. Blasted peptide motifs results

☐ 3. Motif similarity

☐ 2.1 Motif plot

☐ 2.2 Motif table

☒ 2.3 Motif PWM

Please select one motif to view its PWM (Position weight matrix):  
.....[STY]P.....

Download

Show 10 entries

Search:

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
A	0.0872483	0.0933333	0.06	0.0927152	0.0592105	0.0723684	0.0394737	0	0	0.0927152
C	0	0.0066667	0.0066667	0.0066225	0.0065789	0.0131579	0.0131579	0	0	0
D	0.0268456	0.0466667	0.04	0.0728477	0.0263158	0.0460526	0.0460526	0	0	0.0463576
E	0.0738255	0.0666667	0.0866667	0.0860927	0.0592105	0.0526316	0.0328947	0	0	0.0993377
F	0.0268456	0	0.0133333	0.013245	0.0197368	0.0328947	0	0	0	0.013245
G	0.0671141	0.0866667	0.0733333	0.0529801	0.125	0.0394737	0.0986842	0	0	0.0728477
H	0.0268456	0.02	0.02	0.0066225	0.0263158	0.0197368	0.0460526	0	0	0.013245
I	0.0469799	0.02	0.04	0.0198675	0.0131579	0.0197368	0.0263158	0	0	0.0264901
K	0.0671141	0.0733333	0.0266667	0.0794702	0.0460526	0.0855263	0.0657895	0	0	0.0331126
L	0.0469799	0.0733333	0.08	0.0993377	0.0855263	0.0328947	0.0789474	0	0	0.0794702

Showing 1 to 10 of 20 entries

Previous 1 2 Next

Users can select one motif to view its PWM. The motif here are obtained from “2.2. Motif table”, like below:



Please select one motif to view its PWM (Position weight matrix):

.....[STY]P.....  
.....D[STY].....  
.....R..[STY].....

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
A	0.0872483	0.0933333	0.06	0.0927152	0.0592105	0.0723684	0.0394737	0	0	0.0927152
C	0	0.0066667	0.0066667	0.0066225	0.0065789	0.0131579	0.0131579	0	0	0
D	0.0268456	0.0466667	0.04	0.0728477	0.0263158	0.0460526	0.0460526	0	0	0.0463576
E	0.0738255	0.0666667	0.0866667	0.0860927	0.0592105	0.0526316	0.0328947	0	0	0.0993377
F	0.0268456	0	0.0133333	0.013245	0.0197368	0.0328947	0	0	0	0.013245
G	0.0671141	0.0866667	0.0733333	0.0529801	0.125	0.0394737	0.0986842	0	0	0.0728477
H	0.0268456	0.02	0.02	0.0066225	0.0263158	0.0197368	0.0460526	0	0	0.013245
I	0.0469799	0.02	0.04	0.0198675	0.0131579	0.0197368	0.0263158	0	0	0.0264901
K	0.0671141	0.0733333	0.0266667	0.0794702	0.0460526	0.0855263	0.0657895	0	0	0.0331126
L	0.0469799	0.0733333	0.08	0.0993377	0.0855263	0.0328947	0.0789474	0	0	0.0794702

Showing 1 to 10 of 20 entries

Previous
1
2
Next

### Step 4.2.3. Motif similarity

Herein, users could evaluate how similar a motif PWM calculated from those human kinase substrates is to a motif PWM calculated from sequences in 129 mammalian species corresponding to one particular human substrate.

Users should type in a protein name/uniprot id with a phosphosite (e.g. AKTS1\_T246 or Q96B36\_T246) below. Then PTMoreR will search the sequences across mammalian species aligned and calculate PWM (named PWM\_input). After that, this tool will calculate PWMs for each human kinase in PhosphoSitePlus database (named PWM\_PsP). Finally, this tool will evaluate the cosine similarity between PWM\_input and PWM\_PsP. For example, here we type in AKT1\_S126, it shows as below:

#### Step 4. Motif Enrichment and Plot.

**Input Parameters**

☐ 1. Only analyze regular sequences?

2. Minimum number:

3. P-value threshold:

4. Motif index for plot:

☒ 5. Equal height or not?

6. Figure height:

[Calculate](#)

**Results**

☐ 1. Uploaded peptide motifs results ☐ 2. Blasted peptide motifs results ☒ 3. Motif similarity


Hints: Herein, users should type in a protein name/uniprot id with a phosphosite (e.g. AKTS1\_T246 or Q96B36\_T246) below. Then PTMoreR will search the sequences across mammalian species aligned and calculate PWM (named PWM\_input). After that, this tool will calculate PWMs for each human kinase in PhosphoSitePlus database (named PWM\_PsP). Finally, this tool will evaluate the cosine similarity between PWM\_input and PWM\_PsP.

Please type in a protein with a phosphosite:

[Search](#)

☒ 3.1 Motif plot based on aligned results across 129 mammalian species ☐ 3.2 Aligned table ☐ 3.3 PWM similarity

[Download](#)



Loading database...

The results in “3.1 Motif plot based on aligned results across 129 mammalian species” show the motif plot based on the 15-mer peptide sequences mapped across 129 mammalian species.

**Step 4. Motif Enrichment and Plot.**

Input Parameters

☐ 1. Only analyze regular sequences?

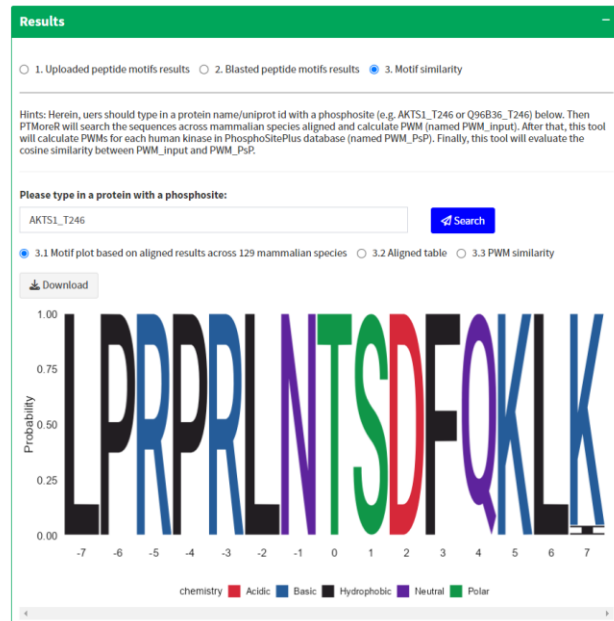
2. Minimum number:

3. P-value threshold:

4. Motif index for plot:

☒ 5. Equal height or not?

6. Figure height:



The “3.2 Aligned table” shows all mapping information about the phosphosite users type in across 129 mammalian species.

Results

1. Uploaded peptide motifs results

2. Blasted peptide motifs results

3. Motif similarity

Hints: Herein, users should type in a protein name/uniprot id with a phosphosite (e.g. AKT1\_S126 or P31749\_S126) below. Then PTMoreR will search the sequences across mammalian species aligned and calculate PWM (named PWM\_input). After that, this tool will calculate PWMs for each human kinase in PhosphoSitePlus database (named PWM\_PsP). Finally, this tool will evaluate the cosine similarity between PWM\_input and PWM\_PsP.

Please type in a protein with a phosphosite:

AKT1\_S126

Search

3.1 Aligned results across mammalian species

3.2 PWM similarity

Download

Show10entries

Search:

	Uniprot.Human	Position.Human	Center.amino.acids.Human	Seqwindows.Human	Center.amino.
1	P31749	126	S	DFRSGSPSDNSGAEE	S
2	P31749	126	S	DFRSGSPSDNSGAEE	S
3	P31749	126	S	DFRSGSPSDNSGAEE	S
4	P31749	126	S	DFRSGSPSDNSGAEE	S
5	P31749	126	S	DFRSGSPSDNSGAEE	S
6	P31749	126	S	DFRSGSPSDNSGAEE	S
7	P31749	126	S	DFRSGSPSDNSGAEE	S
8	P31749	126	S	DFRSGSPSDNSGAEE	S
9	P31749	126	S	DFRSGSPSDNSGAEE	S
10	P31749	126	S	DFRSGSPSDNSGAEE	S

Showing 1 to 10 of 90 entries

Previous

1

2

3

4

5

...

9

Next

Results

+

28

The table is like below:

Uniprot Human	Position Human	Center amino acids Human	Seqwindows Human	Center amino acids Other	Seqwindows Other	Uniprot Other	Position Other	Taxonomy ID	Taxonomy
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A133EV04	125	10020	Dipodomys ordii (Ord's kangaroo rat)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	G3H397	126	10029	Cricetulus griseus (Chinese hamster) (Cricetulus barabensis griseus)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A1U7QILZ	126	10038	Mesocricetus auratus (Golden hamster)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	P31750	126	10090	Mus musculus (Mouse)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A8C6IDB1	126	10103	Mus spicilegus (Steppe mouse)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	P47196	126	10116	Rattus norvegicus (Rat)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A296XK38	126	10141	Cavia porcellus (Guinea pig)
P31749	126 S		DFRSGSPSDNSGAE	S	EFQSGSPSDNSGAE	A0A8P6QPS2	126	10160	Oryzomys latirostris (Florida manatee)
P31749	126 S		DFRSGSPSDNSGAE	S	EFQSGSPSDNSGAE	GS4W17	266	10181	Heterocephalus glaber (Naked mole rat)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A8C6RGIS	127	1026970	Nannospalax galili (Northern Israeli blind subterranean mole rat) (Spalax galili)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	S7N156	111	109478	Myotis brandtii (Brandt's bat)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A897A493	126	1230840	Oryzomys afer afer
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A2Y9DM80	126	127502	Trichechus manatus latirostris (Florida manatee)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A8C8ZL01	126	1328070	Pteropus vampyrus (Greater bamboo lemur) (Hapalemur simus)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A6P6BRD4	126	132908	Pteropus vampyrus (Large flying fox)
P31749	126 S		DFRSGSPSDNSGAE	S	DFQSGSPSDNSGAE	F6QP11	126	13616	Monodelphis domestica (Gray short-tailed opossum)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A89B716	126	185453	Chrysochloris asiatica (Cape golden mole)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A3Q0DU04	126	1868482	Carollia sylvatica (Philippine tarsier) (Tarsius sylvatica)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A8B7R1K3	126	186990	Hipposideros amiger (Great Himalayan leaf-nosed bat)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A485NG58	126	191816	Lynx pardinus (Iberian lynx) (Felis pardina)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A89M129	126	230844	Peromyscus maniculatus bairdi (Prairie deer mouse)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A2KSS2L0	135	2715852	Cebus imitator (Panamanian white-faced capuchin) (Cebus capucinus imitator)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A73URH0	126	27622	Molossus molossus (Pallas' mastiff bat) (Vespertilio molossus)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A8C0K2Y4	126	286419	Canis lupus dingo (dingo)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A8M1GN04	126	29073	Ursus maritimus (Polar bear) (Thalarcos maritimus)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A4C2CT4	126	29139	Vombatus ursinus (Common wombat)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A83AQ03	126	30538	Vicugna pacos (Alpaca) (Lama pacos)
P31749	126 S		DFRSGSPSDNSGAE	S	DFRSGSPSDNSGAE	A0A8C5V171	126	30608	Microcebus murinus (Gray mouse lemur) (Lemur murinus)

Each column in this table means:

1. Uniprot.Human: the human UniProt ID of the protein name that users type in.
2. Position.Human: the phosphosite position in the human protein sequence.
3. Center.amino.acids.Human: the central amino acid mapped from the human protein sequence.
4. Seqwindows.Human: the standard peptides mapped from the human protein sequence.
5. Center.amino.acids.Other: the central amino acid mapped from other species protein sequence.
6. Seqwindows.Other: the standard peptides mapped from other species protein sequence.
7. Uniprot.Other: the UniProt ID from other species protein.
8. Position.Other: the phosphosite position in other protein sequence.
9. Taxonomy.ID: taxonomy IDs of other species.
10. Taxonomy: Taxonomy names of other species.

When users click “3.3 PWM similarity”, this tool will evaluate the cosine similarity between PWM\_input and PWM\_PsP:

Step 4. Motif Enrichment and Plot.

Input Parameters

☐ 1. Only analyze regular sequences?

2. Minimum number:

20

3. P-value threshold:

0.000001

4. Motif index for plot:

1-2

☒ 5. Equal height or not?

6. Figure height:

400

Calculate

Results

☐ 1. Uploaded peptide motifs results

☐ 2. Blasted peptide motifs results

☒ 3. Motif similarity

Hints: Herein, users should type in a protein name/uniprot id with a phosphosite (e.g. AKTS1\_T246 or G96B36\_T246) below. Then PTMoreR will search the sequences across mammalian species aligned and calculate PWM (named PWM\_input). After that, this tool will calculate PWMs for each human kinase in PhosphoSitePlus database (named PWM\_PsP). Finally, this tool will evaluate the cosine similarity between PWM\_input and PWM\_PsP.

Please type in a protein with a phosphosite:

AKTS1\_T246

Search

☐ 3.1 Motif plot based on aligned results across 129 mammalian species

☐ 3.2 Aligned table

☒ 3.3 PWM similarity

Download

Show

10

 entries

Search:

	PWM_input	PWM_PsP	Cosine.Similarity
1	AKTS1_T246	AKT1	0.4155067510381872
2	AKTS1_T246	AKT3	0.3897702147108681
3	AKTS1_T246	AKT2	0.3784497408502874
4	AKTS1_T246	SGK1	0.3613961165730667
5	AKTS1_T246	PIM2	0.3438977991856739
6	AKTS1_T246	LRRK2	0.3428924372807644
7	AKTS1_T246	CAMK2D	0.342590204447749
8	AKTS1_T246	DAPK3	0.3416843256841278
9	AKTS1_T246	P90RSK	0.3412563085876141
10	AKTS1_T246	LKB1	0.3407952175047307

Showing 1 to 10 of 189 entries

Previous

1

2

3

4

5

...

19

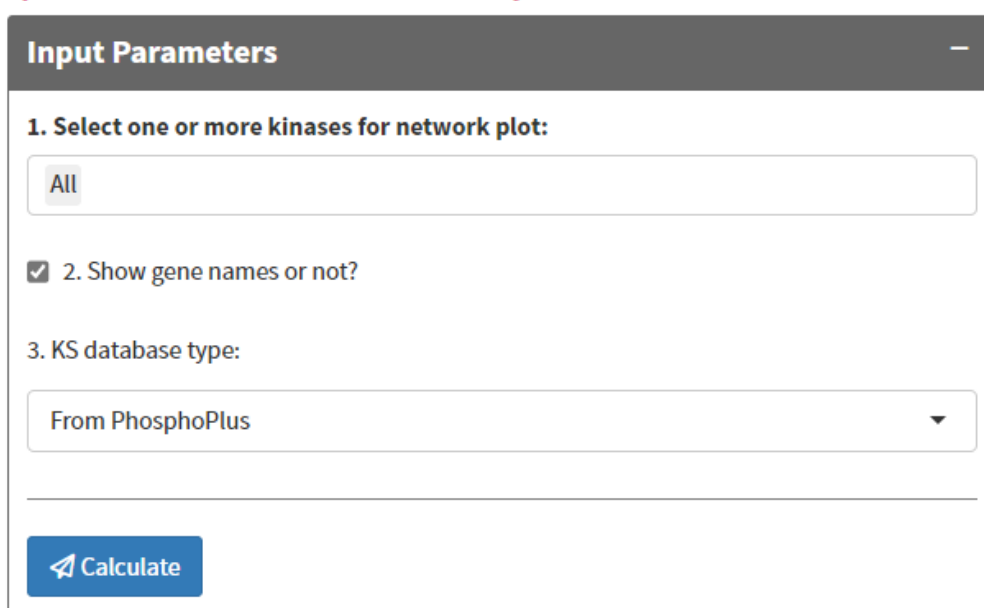
Next

## Step 5. Annotation and enrichment analysis based on Kinase-Substrate database

This step will perform kinase-substrate annotation and enrichment for every kinase using Fisher test based on PhosphoSitePlus<sup>5</sup>, to facilitate a site-specific analysis on phosphorylation regulation. Users should note that here is only for phosphoproteomics data, other modification data are not inappropriate here.

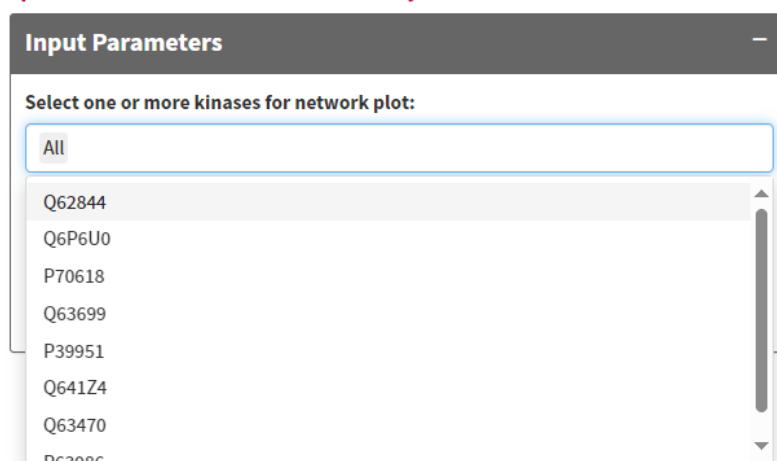
### Step 5.1. Parameters

#### Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.



1. *Select one or more kinases for network plot:* This means which kinase users want to select to show a network plot. If selecting “All”, all kinases identified in users’ data will be used to plot network. User can also just select one or more kinases, like below:

#### Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.



2. *Show gene names or not?* If true, the node table, the edge table, and the network plot will

show the gene names, otherwise, it shows the IDs (e.g. UniProt IDs by default).

3. *KS database type*: Which kind of Kinase-Substrate database users want to choose. From PhosphoPlus means the KS database was downloaded from the PhosphoPlus database. From the Kinase library (top 1) means the KS database was produced from the Kinase Library (Lew Cantley et al. 2024)<sup>6,7</sup> with the topmost percentile. From the Kinase library (above 0.99) means the KS database was produced from the Kinase Library with percentile > 0.99..

**Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.**

**Input Parameters**

1. Select one or more kinases for network plot:

All

☒ 2. Show gene names or not?

3. KS database type:

From PhosphoPlus

From PhosphoPlus

From the Kinase library (top 1)

From the Kinase library (above 0.99)

**Results**

[Annotation tables](#)

Which kind of Kinase-Substrate (KS) database users want to choose. From PhosphoPlus means the KS database was downloaded from the PhosphoPlus database, and From the Kinase library (top 1) means the KS database was produced from the Kinase Library (Lew Cantley et al. 2024) with the topmost percentile. The Kinase library (above 0.99) means the KS database was produced from the Kinase Library with percentile > 0.99.

Then, users can click the “Calculate” button.



## Step 5.2. Results

Here shows the annotation tables for the uploaded peptides (1. *Uploaded peptide annotation*) and the blasted peptides (2. *Blasted peptide annotation*) based on the kinase-substrate database from PhosphoSitePlus <sup>5</sup>.

### Step 5.2.1. Uploaded peptide annotation

When users click “1. *Uploaded peptide annotation*”, they will obtain the annotation table for the uploaded peptides:

Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.

#### Input Parameters

1. Select one or more kinases for network plot:

All

☒ 2. Show gene names or not?

3. KS database type:

From PhosphoPlus

Calculate

#### Results

Annotation tables:

☒ 1. Uploaded peptide annotation ☐ 2. Blasted peptide annotation ☐ 3. KS enrichment analysis

Download Result description

Show 10 entries Search:

	KIN_ACC_ID	SUB_ACC_ID	Pep.upload	Pep.all.index	Center.amino.acid	Seqwindows
1	Q62844	Q9R044	GAY#SLSIR	3	Y	SETTKGAYSLIRDW
2	Q62844	Q05764	GAY#SLSIR	3	Y	SETTKGAYSLIRDW
3	Q62844	Q8CGX0	GAY#SLSIR	3	Y	SETTKGAYSLIRDW
4	Q62844	Q03114	GAY#SLSIR	3	Y	SETTKGAYSLIRDW
5	Q62844	Q9R044	GAYS#LSIR	4	S	ETTKGAYSLIRDW
6	Q62844	Q05764	GAYS#LSIR	4	S	ETTKGAYSLIRDW
7	Q62844	Q8CGX0	GAYS#LSIR	4	S	ETTKGAYSLIRDW
8	Q62844	Q03114	GAYS#LSIR	4	S	ETTKGAYSLIRDW
9	Q62844	Q9R044	LDNGGY#Y#ITTR	7	Y	KLDNGGY#ITTRAQF
10	Q62844	Q05764	LDNGGY#Y#ITTR	7	Y	KLDNGGY#ITTRAQF

Showing 1 to 10 of 278 entries Previous 1 2 3 4 5 ... 28 Next

Network Plot

☐ Change figure size?

Download

Every column in this table means (Users can also check this by clicking “Result description” button):

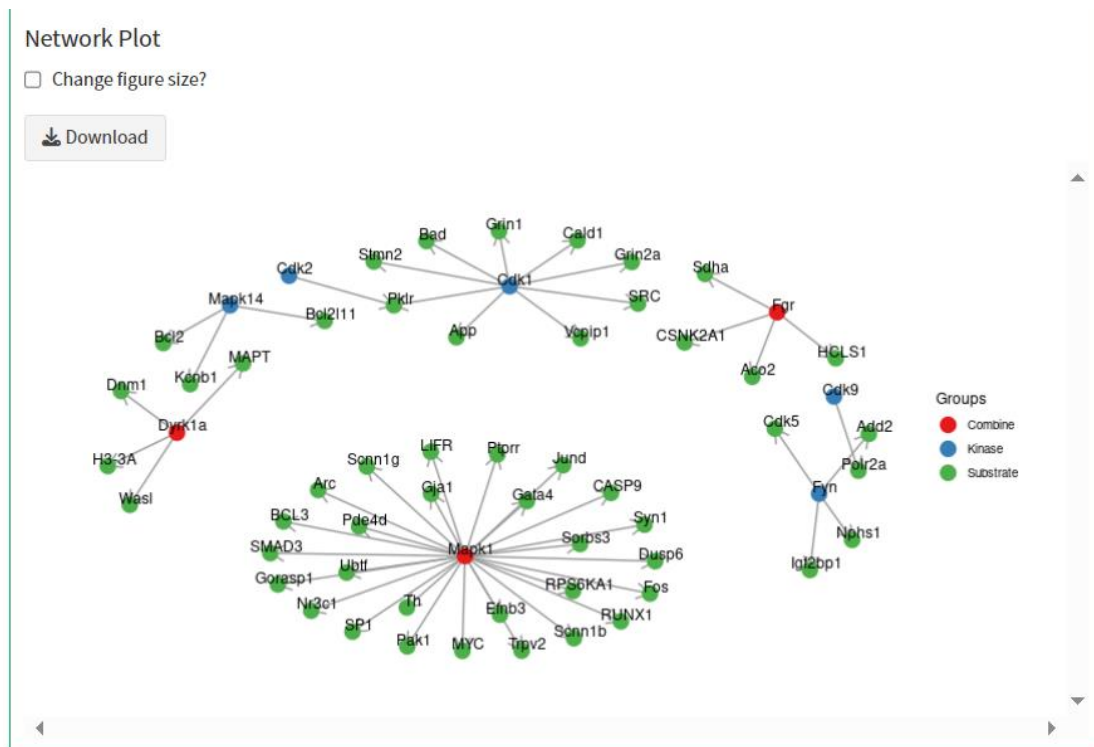
1. *KIN\_ACC\_ID*: kinase uniprot id.
2. *SUB\_ACC\_ID*: substrate uniprot id.
3. *Pep.upload*: the original peptide.
4. *Pep.all.index*: the position of all modified amino acid in the peptide.
5. *Center.amino.acid*: the central amino acid in the aligned peptide. Or, Center.amino.acids.Other: the central amino acid mapped from the human peptides.
6. *Seqwindows*: the aligned standard peptides. Or, Seqwindows.Other: the standard peptides mapped from the human peptides.

7. *PROindex.from.Database*: the position of modified amino acid in the protein sequence. Or, *PROindex.from.Other*: the position of modified amino acid in the mapped human protein sequence.

8. *GENE*: kinase gene name.

9. *SUB\_GENE*: substrate gene name.

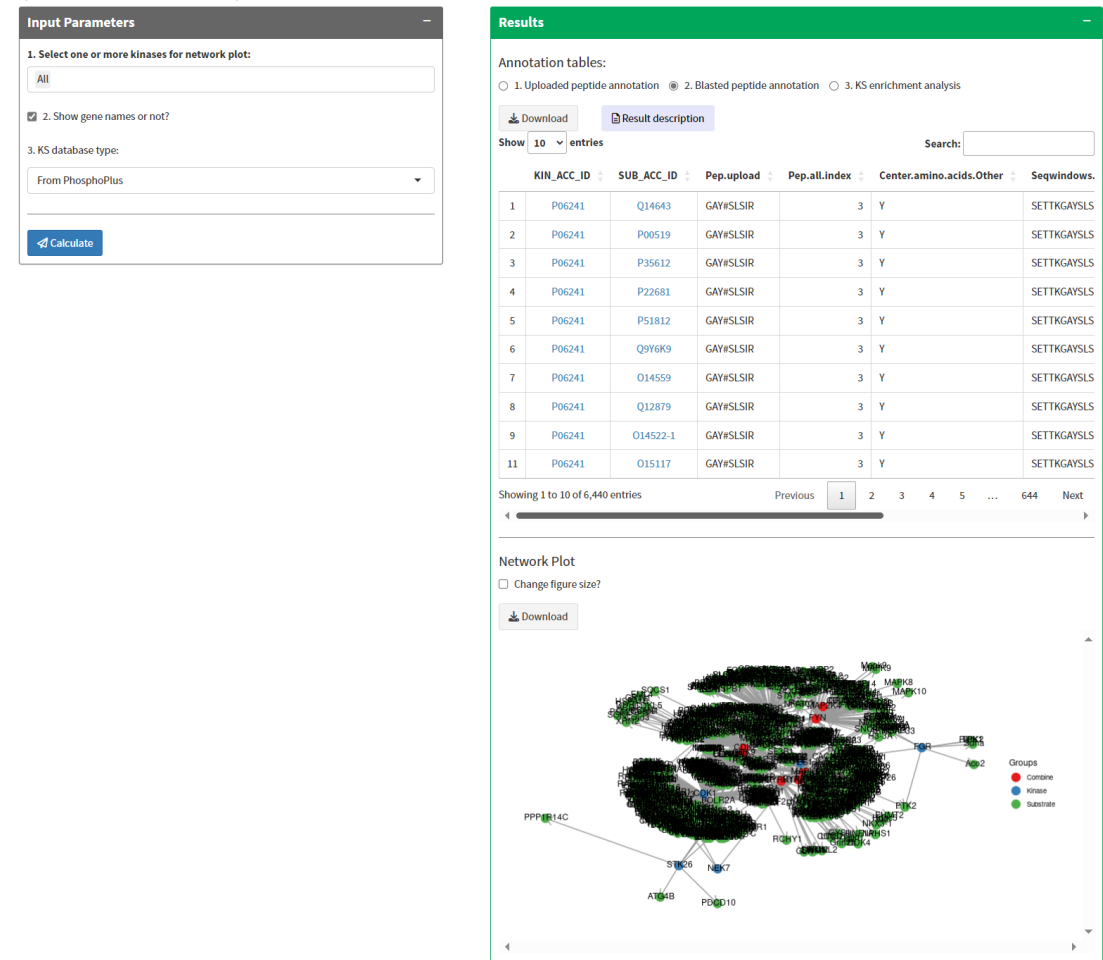
Then they can also obtain the network plot:



Step 5.2.2. Blasted peptide annotation

When users click “2. *Blasted peptide annotation*”, they will obtain the annotation table for the blasted peptides:

Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.

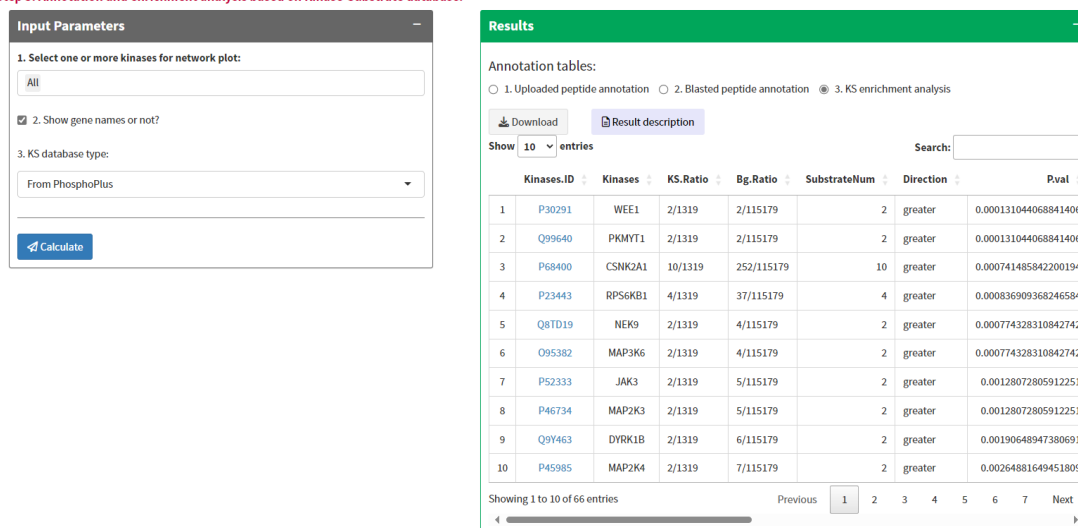


The annotation table and network plot are similar to the above.

### Step 5.2.3. KS enrichment analysis

When users click “3. *KS enrichment analysis*”, this function will process the enrichment analysis at phosphosite levels for every kinases. Then, the results are shown as below:

Step 5. Annotation and enrichment analysis based on Kinase-Substrate database.



Every column in this table means (Users can also check this by clicking “Result description” button):

1. *Kinases.ID*: Kinase uniprot ids.
2. *Kinases*: Kinase gene names.
3. *KS.Ratio*: k/n, k means the overlap between phosphosites-of-interest and the uploaded phosphosite set, n means the number of all unique phosphosites-of-interest.
4. *Bg.Ratio*: M/N, M means the number of substrate phosphosites of each kinase in the whole phosphosite set, N means the number of phosphosites in the whole phosphosite set.
5. *SubstrateNum*: Same as k.
6. *Direction*: If  $KS.Ratio \geq Bg.Ratio$ , the value is 'greater', otherwise, 'less'.
7. *P.val*: Original P value obtained from Fisher test.
8. *P.adj*: Adjusted P value based on the BH method.
9. *Substrates*: Substrate phosphosites-of-interest.

## Step 6. Interaction Plot

In some cases, it may not be convenient for the users to display a network diagram of protein-protein-modification sites relationships. This step mainly shows the plot of the interaction between the uploaded and the blasted peptides/proteins. In this step, users need to upload two kinds of data (Expression data and Interaction data) and type in the samples information, shown as below:

Step 6. Interaction Plot.

Input Parameters

☒ Upload experimental data ☐ Load example data

1. Expression data:  
1.1 Import your expression data:  

Browse... No file selected

☒ 1.1.1 First row as column names?  
☐ 1.1.2 First column as row names?

2. Samples information:  
2.1 Group and replicate number:  

2.2 Group names:

3. Interaction data:  
3.1 Import interaction data:  

Browse... No file selected

☒ 3.1.1 First row as column names?  
☐ 3.1.2 First column as row names?

Results

☒ 1. Original Expression data ☐ 2. Processed Expression data ☐ 3. Interaction plot

1.1. Expression Data:  
Show 10 entries Search:

Description

1 PTMoreR detects that you do not upload your data. Please upload the expression data, or load the example data to check first.

Showing 1 to 1 of 1 entries Previous 1 Next

1.2. Interaction Data:  
Show 10 entries Search:

Description

1 PTMoreR detects that you do not upload your data. Please upload the interaction data, or load the example data to check first.

Showing 1 to 1 of 1 entries Previous 1 Next

37

Users can click “Load example data”. And shown in the “1. Original Expression data” part below, the example data are from the African green monkey (*Chlorocebus sabaeus*) cell phosphoproteome quantified at 6 time points after SARS-CoV-2 infection (0, 2, 4, 8, 12, or 24 h) using Spectronaut (18.0.230605.50606)<sup>8</sup>.

#### Step 6. Interaction Plot.

Input Parameters

☐ Upload experimental data
 ☒ Load example data

1. Download example expression data

2. Samples information:

2.1 Group and replicate number:

6-2-2-3-3-3-3

2.2 Group names:

0h;2h;4h;8h;12h;24h

3. Download example interaction data

Results

☒ 1. Original Expression data
 ☐ 2. Processed Expression data
 ☐ 3. Interaction plot

1.1. Expression Data:

Show 10 entries

Search:

	PTM.ProteinId	PTM.SiteAA	PTM.SiteLocation	Infect0h_1	Infect0h_3	Infect2h_1	Infect2h_2
1	AA0A09QUI5	S	203	716860	796557.3125	465183.9688	45876
2	AA0A09QUI5	S	43	115649.1484	111377.6016	118711.6406	13093
3	AA0A09QUI5	T	44	73666.78906	69151.78125	77679.3125	86795
4	AA0A09QUI5	S	205	607172.125	673837.9375	354726.5625	36755
5	AA0A09QUI5	S	206	572782.875	556242.375	293836.2188	26047
6	AA0A09QUI5	S	36				
7	AA0A09QUI8	S	47	73103.36719	98003.66406	82692.94531	13106
8	AA0A09QUI9	S	82	11173.93555	9107.095703	32361.55664	17032
9	AA0A09QUI9	T	84	9019.821289	9600.298828	32226.09961	16501
10	AA0A09QUI9	S	252	13969.50977	20955.19727	7962.896973	8761

Showing 1 to 10 of 13,191 entries

Previous 1 2 3 4 5 ... 1,320 Next

1.2. Interaction Data:

Show 10 entries

Search:

	Bait	Preys	PreyGene
1	nsp1	P09884	POLA1
2	nsp1	P49642	PRIM1
3	nsp1	P49643	PRIM2
4	nsp1	Q14181	POLA2
5	nsp1	Q8NBJ5	COLGALT1
6	nsp1	Q99959	PKP2
7	E	O00203	AP3B1
8	E	O60885	BRD4
9	E	P25440	BRD2
10	E	Q6LUX04	CWC27

Showing 1 to 10 of 332 entries

Previous 1 2 3 4 5 ... 34 Next

## Step 6.1. Data preparation

Here users should prepare two kinds of data: The modification site expression data and the interaction data.

### Step 6.1.1. The modification site expression data

Users should note that the modification site expression data should be obtained from the same species as processed in the previous steps. As shown in the example data below, the first three columns are protein ids (e.g. PTM.ProteinId), modification site amino acid (e.g. PTM.SiteAA), modification site position (e.g. PTM.SiteLocation). And the other columns are samples (e.g. The African green monkey cell samples with 6 time points after SARS-CoV-2 infection (0, 2, 4, 8, 12, or 24 h)). These data can be extracted from many popular software, such as MaxQuant, Spectronaut etc., and saved in a .csv file. The missing values are NA and shown as blank after uploaded into *PTMoreR*.

38

PTM.ProteinId	PTM.SiteAA	PTM.SiteLocation	Infect0h_1	Infect0h_3	Infect2h_1	Infect2h_3	Infect4h_1	Infect4h_2	Infect4h_3
A0A0D9QUI5	S	203	716860	796557.3125	465183.9688	458769.1875	467674.1875	517208.6875	563238.9375
A0A0D9QUI5	S	43	115649.1484	111377.6016	118711.6406	130935.4453	91683.72656	106065.1563	109223.6016
A0A0D9QUI5	T	44	73666.78906	69151.78125	77679.3125	86795.76563	60218.52344	69550.14063	71446.70313
A0A0D9QUI5	S	205	607172.125	673837.9375	354726.5625	367551.6875	375132.9375	433767.8125	444802.3125
A0A0D9QUI5	S	206	572782.875	556242.375	293836.2188	260473.8125	358492.0313	304453.4063	395302.4063
A0A0D9QUI5	S	36						418.7579041	
A0A0D9QUI8	S	47	73103.36719	98003.66406	82692.94531	131065.3516	88582.30469	156939.7969	121457.4609
A0A0D9QUI9	S	82	11173.93555	9107.095703	32361.55664	17032.79297	37902.76563	34069.21094	28763.04102
A0A0D9QUI9	T	84	9019.821289	9600.298828	32226.09961	16501.47852	38630.11328	35007.21875	29710.75586
A0A0D9QUI9	S	252	13969.50977	20955.19727	7962.896973	8761.28125	11051.95215	10059.21973	9189.84668

### Step 6.1.2. The interaction data

Users should know the interaction data in advance, which could be protein-protein interaction data, or drug-protein interaction data and so on. There are three columns in this table: The first one is protein ids/names from one species (e.g. SARS-CoV-2 protein names), or drug names; the second one is protein ids from the species that users blast to in the previous steps (e.g. human protein ids, as users blast the phosphopeptides from green monkey to human); the third one is gene names relative to the protein ids in the second column.

Bait	Preys	PreyGene
nsp1	P09884	POLA1
nsp1	P49642	PRIM1
nsp1	P49643	PRIM2
nsp1	Q14181	POLA2
nsp1	Q8NBJS	COLGALT1
nsp1	Q99959	PKP2
E	O00203	AP3B1
E	O60885	BRD4
E	P25440	BRD2
E	Q6UX04	CWC27

## Step 6.2. Parameters

### Step 6. Interaction Plot.

Input Parameters

☒ Upload experimental data ☐ Load example data

1. Expression data:

1.1 Import your expression data:

Browse...

No file selected

☒ 1.1.1 First row as column names?

☐ 1.1.2 First column as row names?

2. Samples information:

2.1 Group and replicate number:

2.2 Group names:

3. Interaction data:

3.1 Import interaction data:

Browse...

No file selected

☒ 3.1.1 First row as column names?

☐ 3.1.2 First column as row names?

The parameters in the “1. *Expression data*” and the “3. *Interaction data*” parts are similar and easy. Users just click the “Browse” and find their data in a .csv file, then notice that whether the first row/column is used as row/column names, if true, they should select relative parameters.

For the “2. *Samples information*” part, the “2.1. *Group and replicate number*” means users can type in the group number and the biological replicate number here. Please note, the group number and replicate number are linked with “;”, and the replicate number of each group is linked with “-”. For example, if you have two groups, each group has three replicates, then you should type in “2;3-3” here. Similarly, if you have 3 groups with 5 replicates in every groups, you should type in “3;5-5-5”.

2.2 *Group names*: Type in the group names of your samples. Please note, the group names are linked with “;”. For example, there are 6 time points after SARS-CoV-2 infection (0, 2, 4, 8, 12, or 24 h), you can type in “0h;2h;4h;8h;12h;24h”.



### Step 6.3. Processed Expression data

After uploading the proper data and setting up the right parameters, users can click “2. Processed Expression data” (see below), and this tool will process the data automatically for users. By default, those sites with over 50% missing ratios across all samples were removed and missing values were imputed with the k-Nearest Neighbor algorithm provided in NAGuideR<sup>9</sup>.

Step 6. Interaction Plot.

**Input Parameters**

☐ Upload experimental data ☒ Load example data

[1. Download example expression data](#)

**2. Samples information:**

2.1 Group and replicate number:

6:2-2-3-3-3-3

2.2 Group names:

0h;2h;4h;8h;12h;24h

[3. Download example interaction data](#)

**Results**

☐ 1. Original Expression data ☒ 2. Processed Expression data ☐ 3. Interaction plot

☒ 2.1 Median normalization or not? ☒ 2.2 Log or not?

[Download](#)

Show 10 entries

Search:

	Infect0h_1	Infect0h_3	Infect2h_1	Infect2h_3
AA009QIIS_S_203	2.690996420064836	2.76432502302313	2.224753861278204	2.103783475176827
AA009QIIS_S_43	0.05905966892960317	-0.07399401357948819	0.2544138912500972	0.2948706764252863
AA009QIIS_T_44	-0.5916087077241277	-0.7616148260274211	-0.3574451807720421	-0.2982884561412918
AA009QIIS_S_205	2.451410585210866	2.522948510650268	1.833659847521451	1.783962122219042
AA009QIIS_S_206	2.367293386127361	2.246260509826919	1.561964717994748	1.287153311088361
AA009QIIS_S_47	-0.60268520926086	-0.25854555139053	-0.2672113573961805	0.2963013223803052
AA009QIIS_S_82	-3.312485664341492	-3.686318289583899	-1.620694607768741	-2.64759809617011
AA009QIIS_T_84	-3.62145231391303	-3.610230021283688	-1.626746026361837	-2.693317823361685
AA009QIIS_S_252	-2.990341675121951	-2.484073136711949	-3.643610316312978	-3.606699347880994
AA009QIIS_S_583	-0.1028138587762369	0.01416205470896102	-0.08937254376290947	0.02582638668141129

Showing 1 to 10 of 10,118 entries

Previous 1 2 3 4 5 ... 1,012 Next

2.1. *Median normalization or not*: if true, *PTMoreR* will process median normalization for original data. (Note, *PTMoreR* was not designed to perform sophisticated normalization analysis. Any normalized datasets with NA can be accepted for analysis).

2.2. *Log or not*: if true, the data will be transformed to the logarithmic scale with base 2.

## Step 6.4. Interaction plot

In this step, PTMoreR will show the final interaction plot based on the uploaded expression data and the interaction data.

### Step 6. Interaction Plot.

Input Parameters

☐ Upload experimental data ☒ Load example data

1. Download example expression data

2. Samples information:

2.1 Group and replicate number:

6;2-2;3-3-3-3

2.2 Group names:

0h;2h;4h;8h;12h;24h

3. Download example interaction data

Results

☐ 1. Original Expression data ☐ 2. Processed Expression data ☒ 3. Interaction plot

3.1 Node color for the protein in the first column of the interaction data:

red

3.2 Node color for the protein in the second column of the interaction data:

grey

3.3 Node color for expression data:

blue;white;darkred

☒ 3.4 Scaled expression data (Z-score) or not?

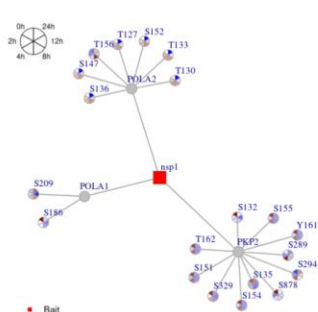
3.5 Select one interacting protein:

nsp1

3.6 Figure Height:

600

Download



3.1 Node color for the protein in the first column of the interaction data: Three are three columns in the interaction data (see “7.1.2. The interaction data” part), the node color for one protein in the first column (e.g. One SARS-CoV-2 protein in the *Bait* column).

3.2 Node color for the protein in the second column of the interaction data: Similar as above, the node color for one protein in the second column (e.g. One human protein in the *Preys* column).

3.3 Node color for expression data: Three color names which are linked with “;” for the expression data (e.g. blue;white;darkred). The first color (e.g. blue) indicates the lowest expression value, the second one (e.g. white) indicates the middle expression value, and the third one (e.g. darkred) indicates the largest expression value.

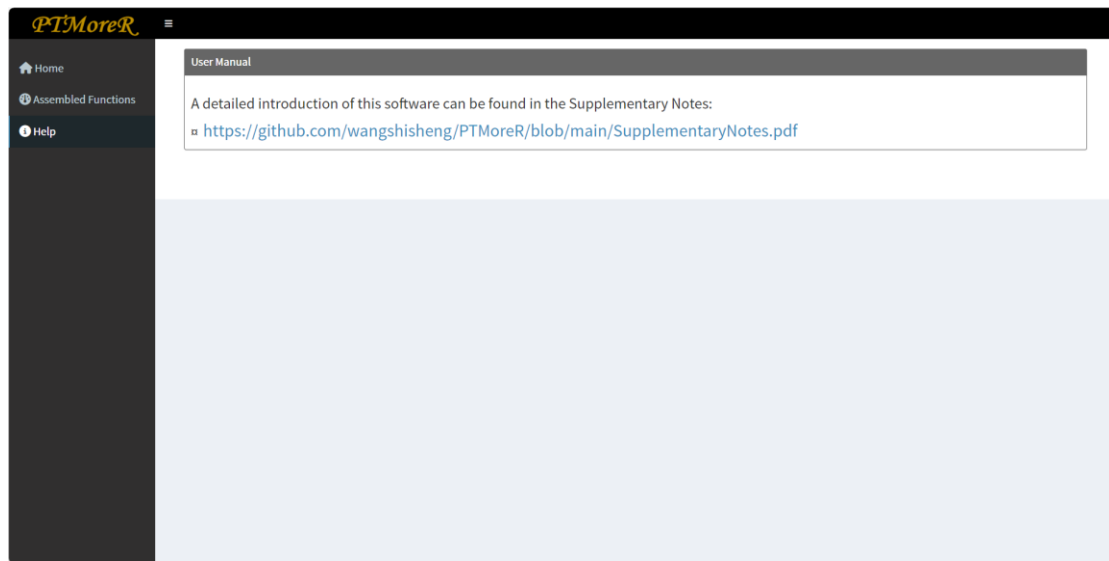
3.4 Scaled expression data (Z-score) or not? If true, the expression data will be scaled by rows, which means the mean of all of the values in each row is 0 and the standard deviation is 1.

3.5 Select one interacting protein: Users can change this parameter to select any protein in the first column (e.g. One SARS-CoV-2 protein in the *Bait* column) and the plot will show the selected protein.

42

### 3. Help

This part provides a detailed operation manual about *PTMoreR*, which is saved as a pdf file as below:



#### 4. How to run this tool locally?

*PTMoreR* is an open source software for non-commercial use and all codes can be obtained on our GitHub: <https://github.com/wangshisheng/PTMoreR>. If users want to run *PTMoreR* on their own computer independent of the internet speed, they should operate as below:

As this tool was developed with R, you may:

- a) Install R. You can download R from here: <https://www.r-project.org/>.
- b) Install RStudio. (Recommendatory but not necessary). You can download RStudio from here: <https://www.rstudio.com/>.
- c) Check packages. After installing R and RStudio, you should check whether you have installed these packages (shiny, shinyjs, shinyWidgets, shinyBS, shinydashboard, shinycssloaders, DT, data.table, openxlsx, Biostrings, GenomicFeatures, rtracklayer, stringi, stringr, ggsci, ggplot2, ggrepel, msa, tidyr, ggraph, graphlayouts, scales, impute, igraph, scatterpie, plotfunctions, mapplots, devtools, KinSwingR, rBLAST). You may run the codes below to check them:

```
if(!require(pacman)) install.packages("pacman")
pacman::p_load(shiny, shinyjs, shinyWidgets, shinyBS, shinydashboard, shinycssloaders,
DT, data.table, openxlsx, Biostrings, GenomicFeatures, rtracklayer, stringi, stringr, ggsci,
ggplot2, ggrepel, msa, tidyr, ggraph, graphlayouts, scales, impute, igraph, scatterpie,
plotfunctions, mapplots, devtools, KinSwingR, rBLAST)
```

Then install some packages from GitHub, as below:

```
devtools::install_github("drostlab/metablastr", build_vignettes = TRUE, dependencies =
TRUE)
devtools::install_github('omarwagih/rmotifx')
devtools::install_github("omarwagih/ggseqlogo")
```

- d) Run this tool locally

```
if(!require(PTMoreR)) devtools::install_github("wangshisheng/PTMoreR")
library(PTMoreR)
PTMoreR_app()
```

Then PTMoreR will be started as below, and the detailed operation about PTMoreR can be found in the Supplementary Notes part 1-4 above.

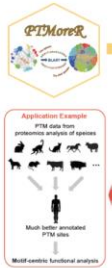
**PTMoreR**

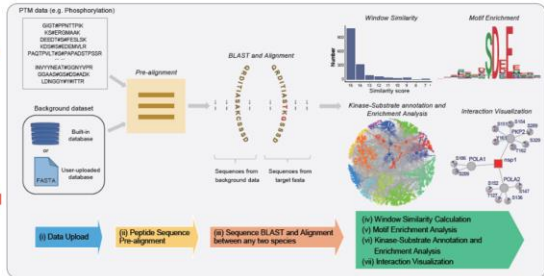
Home

Assembled Functions

Help

## Dear Users, Welcome to PTMoreR





**PTMoreR** is a web-based tool, which possesses the core functions, including:

- Mapping the PTM sites and protein sequences between any species and Human;
- Calculating sequence window similarity and allowing filtering thresholds of sequence similarity during the mapping;
- Performing PTM site-specific enrichment analysis and offering flexible annotations based on kinase-substrate database and network plots;
- Visualizing the regulation of modification sites on the basis of protein-protein interaction data.

In addition, this tool supports both online access and local installation. The source codes and installation instructions can be available in the GitHub repository: <https://github.com/wangshisheng/PTMoreR> under an MIT license.

Finally, PTMoreR is developed by **R shiny (Version 1.6.0)**, and is free and open to all users with no login requirement. It can be readily accessed by all popular web browsers including Google Chrome, Mozilla Firefox, Safari and Internet Explorer 10 (or later), and so on. We would highly appreciate that if you could send your feedback about any bug or feature request to Shisheng Wang at [shishengwang@wchscu.cn](mailto:shishengwang@wchscu.cn).

**Friendly suggestions:**

- Open PTMoreR with Chrome, Mozilla Firefox, Safari or Firefox;
- The minimum operating system specifications are: RAM 4GB, Hard drive 500 GB;
- The monitor resolution ( $\geq 1920 \times 1080$ ) is better.

^\_^ Enjoy yourself in PTMoreR ^\_^

© 2024 Yansheng Liu's Group and Hao Yang's Group. All Rights Reserved.  
Created by Shisheng Wang. E-mail: [shishengwang@wchscu.cn](mailto:shishengwang@wchscu.cn).

## 5. References

1. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319. 10.1038/nprot.2016.136.
2. Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., and Escher, C. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* **14**, 1400-1410.
3. UniProt, C. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480-D489. 10.1093/nar/gkaa1100.
4. Wang, S., Cai, Y., Cheng, J., Li, W., Liu, Y., and Yang, H. (2019). motifeR: An Integrated Web Software for Identification and Visualization of Protein Posttranslational Modification Motifs. *Proteomics* **19**, e1900245. 10.1002/pmic.201900245.
5. Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* **40**, D261-270. 10.1093/nar/gkr1122.
6. Yaron-Barir, T.M., Joughin, B.A., Huntsman, E.M., Kerelsky, A., Cizin, D.M., Cohen, B.M., Regev, A., Song, J., Vasan, N., Lin, T.Y., et al. (2024). The intrinsic substrate specificity of the human tyrosine kinome. *Nature* **629**, 1174-1181. 10.1038/s41586-024-07407-y.
7. Johnson, J.L., Yaron, T.M., Huntsman, E.M., Kerelsky, A., Song, J., Regev, A., Lin, T.Y., Liberatore, K., Cizin, D.M., Cohen, B.M., et al. (2023). An atlas of substrate specificities for the human serine/threonine kinome. *Nature* **613**, 759-766. 10.1038/s41586-022-05575-3.
8. Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinovic, S.M., Cheng, L.Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., et al. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* **14**, 1400-1410. 10.1074/mcp.M114.044305.
9. Wang, S., Li, W., Hu, L., Cheng, J., Yang, H., and Liu, Y. (2020). NAGuideR: performing and prioritizing missing value imputations for consistent bottom-up proteomic analyses. *Nucleic Acids Res* **48**, e83. 10.1093/nar/gkaa498.