

Supporting Information for:

motifeR: An integrated web software for identification and visualization of protein post-translational modification motifs

Shisheng Wang,[†] Yue Cai,[‡] Jingqiu Cheng,[†] Wenxue Li,[§] Hao Yang^{*†} and
Yansheng Liu,^{*§}

[†] West China-Washington Mitochondria and Metabolism Research Center; Key Lab of Transplant Engineering and Immunology, MOH, West China Hospital, Sichuan University, Chengdu, China.

[‡] Ministry of Science and Technology, West China Hospital, Sichuan University, Chengdu, China.

[§] Department of Pharmacology, Yale University School of Medicine, New Haven, CT 06520, USA.

Corresponding Author

*Email address: yanghao@scu.edu.cn; yansheng.liu@yale.edu.

Table of Contents

I. Supplementary Figures and Tables.

Figure S1. The current status of some published tools.

Figure S2. Distribution of original peptide modification site number.

Figure S3. Acetylated peptide motif visualization after enrichment.

Table S1. The example of pre-aligned peptide results. (Please see the online supplementary file).

Table S2. The example of kinase-substrate network analysis. (Please see the online supplementary file).

II. Supplementary notes.

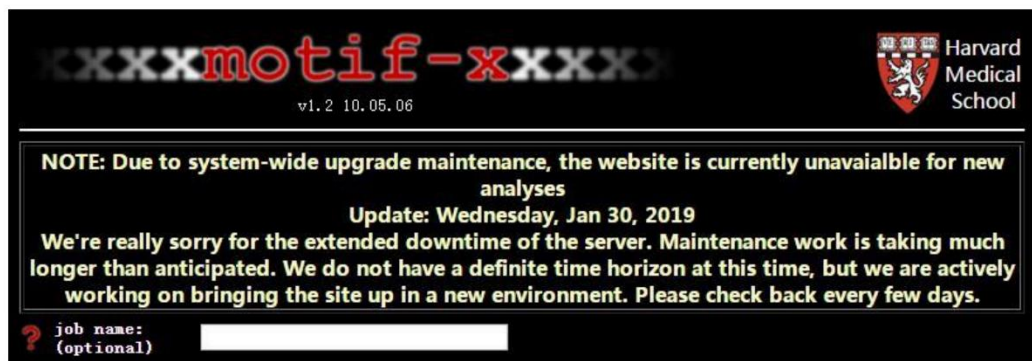
1. Data Preparation.
2. Import data.
3. Pre-alignment.
4. Own Background.
5. Motif Enrichment.
6. Motif Plot.
7. Kinase-Substrate Analysis.
8. Building Species Database.

III. Case Study

A real example for introduction of the operation of this software to help users understand it better when they process their own data sets.

I. Supplementary Figures

A



B Discovery of biological sequence motifs in R

The screenshot shows the GitHub repository page for 'omarwagih/rmotifx'. The repository has 62 commits, 1 branch, 0 releases, and 3 contributors. The latest commit is #13775F on Dec 20, 2018. The file list shows the following files and their commit dates:

File	Commit Message	Commit Date
R	added citation	4 years ago
build	added citation	4 years ago
inst/extdata	added citation	4 years ago
man	package name changed	a year ago
.gitignore	DS_store banished!	5 years ago
DESCRIPTION	update name to rmotifx	5 years ago
NAMESPACE	first commit	5 years ago
README.md	Update README.md	4 months ago

Figure S1. The current status of some published tools. A. motif-x has been in maintenance for a while and users don't know when it can be back. (Link: <http://motif-x.med.harvard.edu/motif-x.html>). B. rmotifx package has not been updated in a long time. (Link: <https://github.com/omarwagih/rmotifx/>).

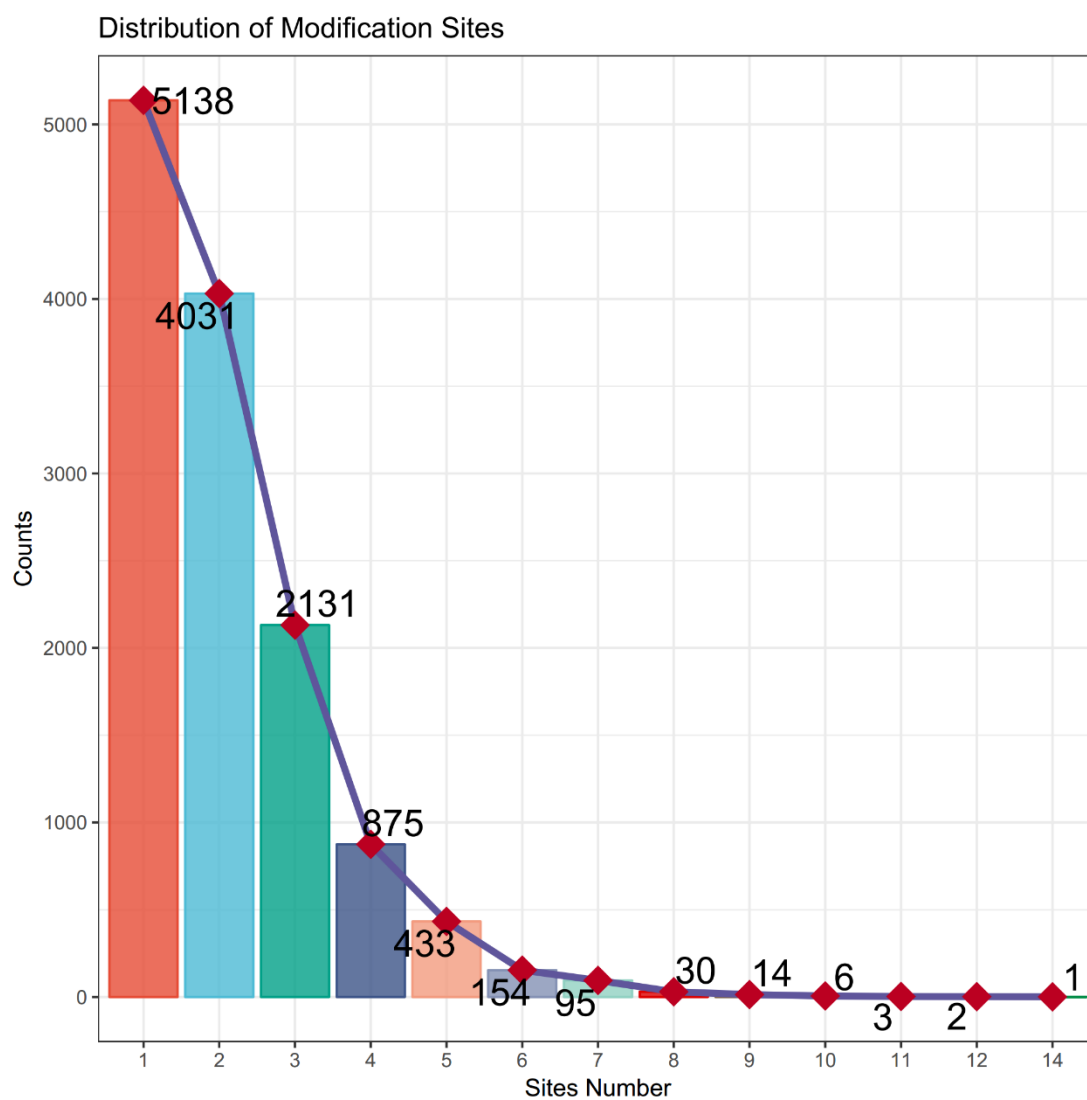


Figure S2. Distribution of original peptide modification site number. This plot can be obtained in “Pre-alignment” module. (This example data from Emdal *et al.*¹)

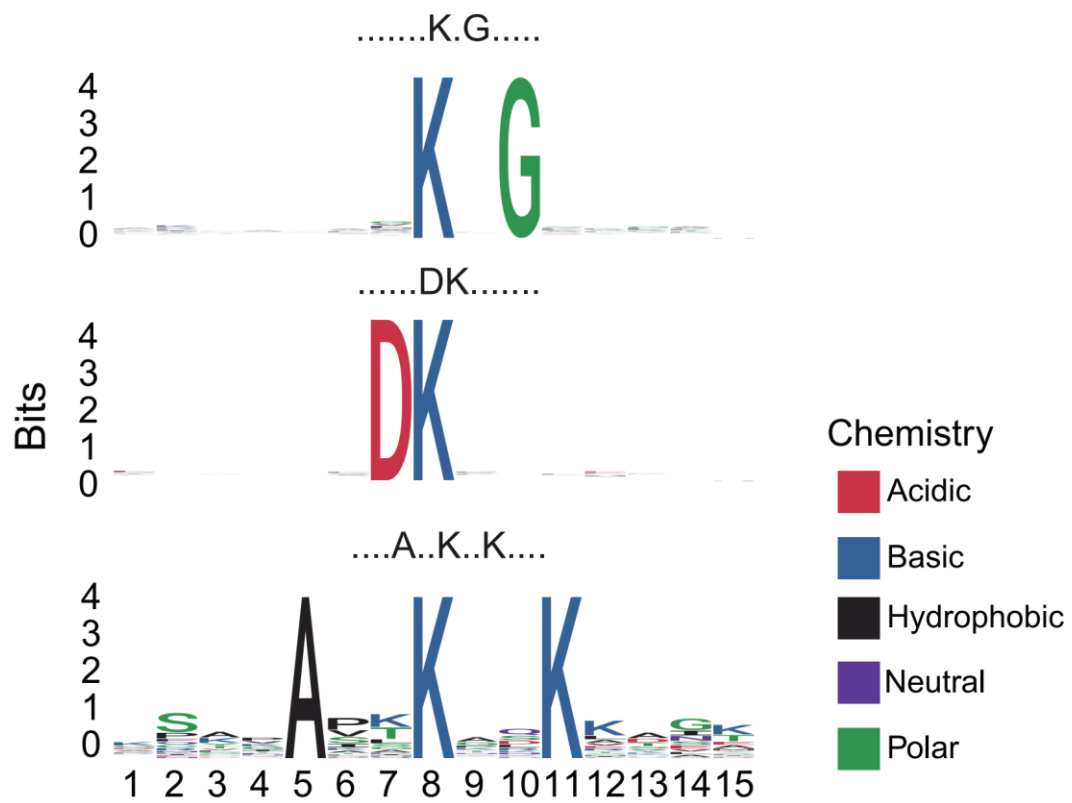
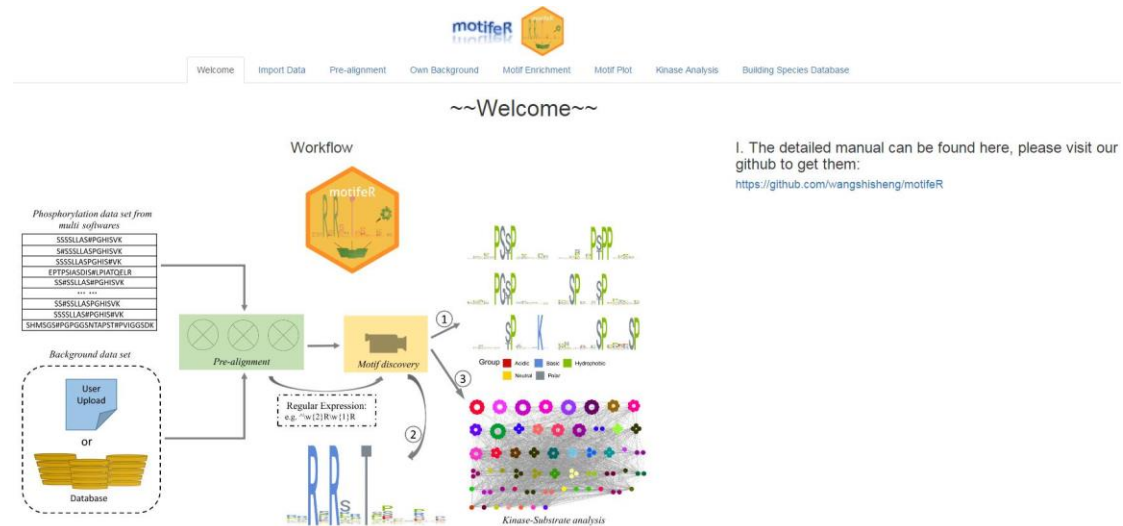


Figure S3. Acetylated peptide motif visualization after enrichment (data from Gu *et al.*²)

II. Supplementary notes:

motifeR, a powerful and comprehensive web server, can provide three main functions (sequence pre-alignment, motif discovery, kinase-substrate analysis) to help users find the patterns of residues along the short span of a protein or polypeptide and process kinase-substrate analysis subsequently. Here we present the detailed introduction and operation of motifeR, users can follow this manuscript to analyze their own data freely and conveniently.

Users can visit this site: <https://www.omicsolution.org/wukong/motifeR>. Then the website homepage can be shown like this:



1. Data Preparation

1.1 Foreground data set, which can be obtained from many softwares, such as MaxQuant, Proteome Discovery, Spectronaut and so on. Then users just prepare the peptide sequences like this:

	A
1	Pep
2	TQVPASVESQK#PR
3	LIDGK#DK#AAR
4	VSSPPPTITQQGK#K
5	VAHPQDPHHSSEK#PVIHCHK
6	DLTDYLMK#ILTR
7	DIK#EKLCYVALDFENEMATAASSSSLEK
8	EK#LCYVALDFENEMATAASSSSLEK
9	EITALAPSTMK#IK
10	EITALAPSTMKIK#
11	DSYVGDEAQSK#R
12	RGILTLK#YPIEHGIITNWDDMEK
13	LNKDDPIGNINLAMEIAEK#HLDIPK
14	TPEK#TMQAMQK
15	QSILAIQNEVEK#VIQSYSIR
16	ELPPDQAQYCIK#R
17	SEFYK#HIVLSGGSTMYPGLPSRLER
18	SALSCHLETVILGLLK#TPAQYDASELK
19	SVCHLQK#VFER
20	QTFK#SHFGR
21	LILGLMMPFAHYDAKQLK#
22	GACTDEK#TLTR
23	GSYNIK#SR
24	LHTEGDK#AFVEFLTDEIKEEK

The length of these sequences can be different, and the standard sequences can be obtained in the pre-alignment step. On the other hand, users should mark those modified residues with some label they like (such as “#” or “@”).

1.2 Database/Background data set. motiferR supports users to upload their own database, if they don't want to use the default data. Herein the database formats can be .fasta, .xlsx, .xls, .csv, or .txt. Particularly, if users want to upload the .fasta file, it may contain all protein sequences of the species they study, otherwise, the other formats (.xlsx, .xls, .csv, .txt.) should contain standard sequences like below:

Windows
____MSKVFKKTS
SKVFKKTSSNGKLSI
KVFKKTSSNGKLSIY
TSSNGKLSIYLKGRD
LQVVPAAESSSPQGGL
QVVPAAESSSPQGGLT
VVPAESSSPQGGLTV
MVTNLPCSVTLQPGP
GIDFEVKSFCAENPE
ENPEETVSKRDYVRL
PEAGPGPSAQTIIRRF

2. Import data.

2.1 Uploading data. When users prepare their data (Foreground and Background data set), they can upload these data from here:

The screenshot displays the 'Import Sequence Data' interface of the MotifTool web application. The top navigation bar includes links for 'Welcome', 'Import Data', 'Pre-alignment', 'Own Background', 'Motif Enrichment', 'Motif Plot', 'Kinase Analysis', and 'Building Species Database'. The 'Import Data' tab is active.

The interface is divided into two main sections: 'Import Sequence Data' on the left and a table of imported sequences on the right.

Import Sequence Data Section:

- 1. Upload:** Includes a 'File format' dropdown (set to 'xlsx'), a 'Browse...' button, and a 'No file selected' message.
- 2. Paste:** Includes a 'Paste your data here:' text area containing the following sequences:
INSWAPSSWPIK
MLISAVSPEIR
KINSWAPSSWPIK
KINSWAPSSWPIK
INSWAPSSWPIK
- Central amino acid:** A text input field with 'S' entered.
- Label of modification:** A text input field with '#' entered.
- Width:** A text input field with '7' entered.

Table of Imported Sequences:

The table shows 10 entries. The first column is 'Input_ID' and the second column is 'Peptide'.

Input_ID	Peptide
1	INSWAPSSWPIK
2	MLISAVSPEIR
3	KINSWAPSSWPIK
4	KINSWAPSSWPIK
5	INSWAPSSWPIK
6	EGSQGEPWTPTANLK
7	EGSQGEPWTPTANLK
8	SHMSGSPGPGGNTAPSTPVGGSDKPGMEEK
9	SHMSGSPGPGGNTAPSTPVGGSDKPGMEEK
10	SSWSSILLASPGHISVK

At the bottom of the table, it says 'Showing 1 to 10 of 3,185 entries'. There are pagination controls with 'Previous', '1', '2', '3', '4', '5', '...', '319', and 'Next'.

There are two choices for users:

1. Upload, users choose the right format and then click "Browse" button to import there data;
Header: this means whether the first row is column names. If true, you should choose this parameter.

First column: this means whether the first column is row names. If true, you should choose this parameter.

2. Paste, users can also paste their sequences in the box.

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

The screenshot shows a web-based parameter configuration form. It includes several input fields: 'Central amino acid' with the value 'S', 'Label of modification' with the value '#', 'Width' with the value '7', 'Minimum number' with the value '20', and 'P-value threshold' with the value '0.000001'. Below these fields are two radio buttons: '1. Select' (unselected) and '2. Upload' (selected). At the bottom, there is a section titled 'Please upload your fasta file:' with a 'Browse...' button and a status indicator 'No file selected'.

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”.

Label of modification: the label represents modification, users can use some label they like, such as “#”, “@”, where “#” is recommended. Here is an example:

The diagram illustrates the transformation of a motif label. On the left, the text 'EGSQGEPWT[Phospho (STY)]PTANLK' is shown. A red arrow points to the right, where the transformed version 'EGSQGEPWT#PTANLK' is displayed. The transformation replaces the descriptive text '[Phospho (STY)]' with the shorthand symbol '#'. Vertical lines are used to align the characters between the two strings.

Width: it is the number of left/right side characters of the central residue. The default is “7” but can be changed by the user. (The N and C terminal sequences, when not sufficient, will be replaced by “_” as MaxQuant).

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

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

Select or Upload fasta file as background data set: if users want to use the default database, they may just select relative species. Please note, the default database only have Human with 15 length sequences now, more species data will be implemented in the future. Optionally, users can also upload their own fasta file (no species limits here), but the calculation time would be longer.

3. 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.

Pre-alignment

☐ 1. Pre-aligned or not ?

☒ 2. Classical multiple sites analysis or not ?

☐ 3. Check if containing some regular sequence ?

Calculate

Alignment results Sites number distribution plot

Download

Show 10 entries

Pep.upload	Stripped.pep	Pep.main.index	Pep.all.index	Seqwindows	PRO.from.Database
1 LFLDGEKEKWAFFESK	LFLDGEKEKWAFFESK	16	16	KEVAFEEKDEHIF	Q9NRN7
2 ALVADEPEDLDTEDEGLISFEEER	ALVADEPEDLDTEDEGLISFEEER	12	12	DEPEDLDTEDEGLIS	Q9BZC7
3 AAKLSWEGSQPAEEEDQETPSR	AAKLSWEGSQPAEEEDQETPSR	5.8	5.8	QARAALKSEGSQPAEAAKLSWEGSQPAEEEDQETPSR	Q9UKV3
4 AAKLSWEGSQPAEEEDQETPSR	AAKLSWEGSQPAEEEDQETPSR	5.8	5.8	QARAALKSEGSQPAEAAKLSWEGSQPAEEEDQETPSR	Q9UKV3
5 QEPTQEHKQEGGQKQEGGQEGGQEGGK	QEPTQEHKQEGGQKQEGGQEGGQEGGK	4	4	VEHROEPTQEHKQEE	Q8NEB7
6 NIGFKVNSK	NIGFKVNSK	8	8	NIGFKVNSKMLGKY	Q5FYE4
7 SWEDQVEGDLGETQSR	SEDQVEGDLGETQSR	1	1	KIKPRLKSEDSQVEGD	Q09666
8 MYFPDFVDKSPK	MYFPDFVDKSPK	12	12	DVEFDKSPKAE	Q09666
9 ISMPDLDLHLKSPK	ISMPDLDLHLKSPK	12	12	DLDLHLKSPKAEV	Q09666
10 VSMPOVELNLKSPK	VSMPOVELNLKSPK	12	12	DVELNLKSPKVGDL	Q09666

Showing 1 to 10 of 6,065 entries

Previous 1 2 3 4 5 ... 607

3.1 Parameters

Pre-alignment

☒ 1. Pre-aligned or not ?

☒ 2. Classical multiple sites analysis or not ?

☐ 3. Check if containing some regular sequence ?

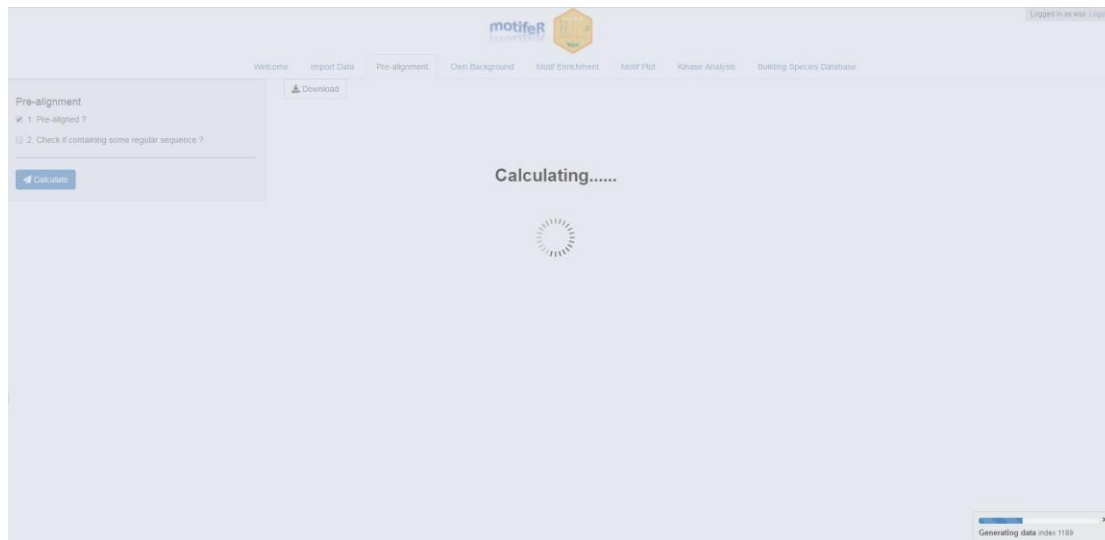
Calculate

Pre-aligned or not: ask users whether to pre-align their sequences, if your sequences are standard (e.g. 15 length amino acids), you can unselect this parameter. Default is true (i.e., start from tryptic peptide sequences from proteomic database search results).

Classical multiple sites analysis or not: ask users whether to process classical analysis. Classical analysis means not replacing the other modified sites with letter “Z” after pre-alignment, for example ‘TSLWNPT#Y#GSWFTEK’ to ‘TSLWNPTYGSWFTEK’, not to ‘TSLWNPZYGSWFTEK’. If true, do not process transformation, otherwise, transformation.

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 ‘ $\wedge\{2\}R\wedge\{1\}R$ ’ (more details can be found here: 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.

3.2 results

There are two results here.

3.2.1 Alignment results

Pep.upload	Pep.no	Pep.main.index	Pep.all.index	Seqwindows	PRO.from.Database	PROindex.from.Database	Contain.if
1 INS#APSS#PIK	INSAPSSPIK	3.7	3.7	SDRRKINSAPSSPIK KINSAPSSPIKTNKA	ADA/K6	413,417	No
2 MLISA/S#PEIR	MLISA/SPEIR	7	7	KMLISA/SPEIRNRD	ADA/K6	71	No
3 KINS#APSS#PIK	KINSAPSSPIK	4.8	4.8	SDRRKINSAPSSPIK KINSAPSSPIKTNKA	ADA/K6	413,417	No
4 KINS#APSS#PIK	KINSAPSSPIK	4.7	4.7	SDRRKINSAPSSPIK KINSAPSSPIKTNKA	ADA/K6	413,416	No
5 INSAPSS#PIK	INSAPSSPIK	7	7	KINSAPSSPIKTNKA	ADA/K6	417	No
6 EGSQGEPTWTPTANLK	EGSQGEPTWTPTANLK	9	9	GSQGEPTWTPTANLKM	ADA/K6	58	No
7 EGSQGEPTWTPTANLK	EGSQGEPTWTPTANLK	11	11	QGEPTWTPTANLKM	ADA/K6	60	No
8 SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEK	SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEK	6	6	IKSHMSGSPGPGGSN	ADFGR8	693	No
9 SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEK	SHMSGSPGPGGSNTAPSTPVIGGSDKPGMEK	14	14	PGPGGSNTAPSTPV	ADFGR8	701	No
10 SSSSLASPGHISVK	SSSLASPGHISVK	2.4	2.4	LHDLGRSSSSLLASPDGRSSSSLLASPOH	ADFGR8	737,739	No

Showing 1 to 10 of 3,185 entries

Previous 1 2 3 4 5 ... 319 Next

Pep.upload: this column contains those peptides users upload.

Stripped.pep: the peptide skeleton.

Pep.index: the position of modified amino acid in the peptide.

Seqwindows: the aligned standard peptides. Note for multiple modification sites or types, the column provides peptides with all the sites respectively centered.

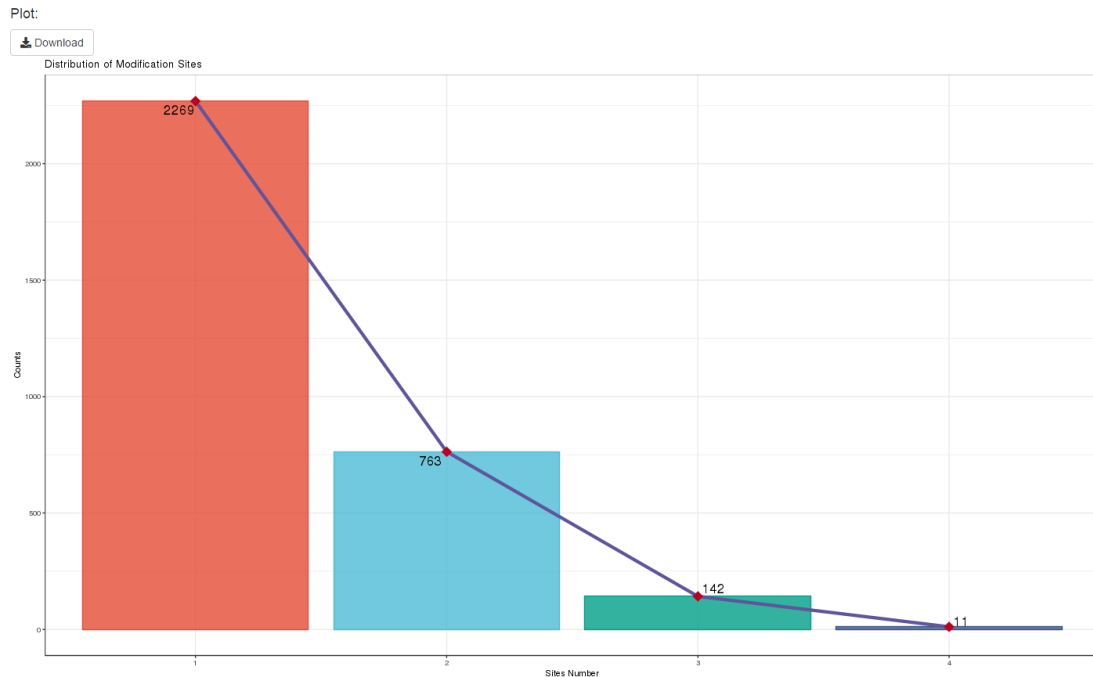
PRO.from.Database: provide the protein name containing this peptide from the fasta file the user uploaded.

PROindex.from.Database: the position of modified amino acid in the protein sequence.

Contain.if: 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*.

3.2.2 Sites number distribution plot

First, this software counts the number of modified sites and plot it:



Second, this tool also extracts those peptides with multi-modified sites so that the user can choose to perform additional analysis to them or just treat them in the conventional manner as previous tools (the result is similar to that from pre-alignment except the last column).

Multi-Sites Data:

Download

Show 10 entries

Search:

Pep.upload	Pep.no	Pep.index	Seqwindows	PRO.from.Database	PROindex.from.Database	Seqwindows_MultiSites
1	INSAPSSPIK	3.7	SDRRKINSAPSSPIK.KINSAPSSPIKTKA	ADA/K6	413,417	SDRRKINSAPSSPIK.KINSAPSSPIKTKA
3	KINSAPSSPIK	4.8	SDRRKINSAPSSPIK.KINSAPSSPIKTKA	ADA/K6	413,417	SDRRKINSAPSSPIK.KINSAPSSPIKTKA
4	KINSAPSSPIK	4.7	SDRRKINSAPSSPIK.KINSAPSSPIKTKA	ADA/K6	413,416	SDRRKINSAPSSPIK.KINSAPSSPIKTKA
10	SSSSLLASPGHISVK	2.4	LHDLGRSSSSLLASPGHISVK	ADFGR8	737,739	LHDLGRSSSSLLASPGHISVK
11	SHMSGSPGGGNTAPSTPVIGGSDKPGMEEK	14.18	PGPGGNTAPSTPVIGGSDKPGMEEK	ADFGR8	701,705	PGPGGNTAPSTPVIGGSDKPGMEEK
13	SHMSGSPGGGNTAPSTPVIGGSDKPGMEEK	4.18	TSIKSHMSGSPGGGNTAPSTPVIGGSD	ADFGR8	691,705	TSIKSHMSGSPGGGNTAPSTPVIGGSD
14	SHMSGSPGGGNTAPSTPVIGGSDKPGMEEK	1.2	GLHDLGRSSSSLLASPGHISVK	ADFGR8	736,737	GLHDLGRSSSSLLASPGHISVK
15	SHMSGSPGGGNTAPSTPVIGGSDKPGMEEK	1.3	GLHDLGRSSSSLLASPGHISVK	ADFGR8	736,738	GLHDLGRSSSSLLASPGHISVK
16	SHMSGSPGGGNTAPSTPVIGGSDKPGMEEK	4.6	TSIKSHMSGSPGGGNTAPSTPVIGGSD	ADFGR8	691,693	TSIKSHMSGSPGGGNTAPSTPVIGGSD
17	SHMSGSPGGGNTAPSTPVIGGSDKPGMEEK	12.18	GSPGGGNTAPSTPVIGGSDKPGMEEK	ADFGR8	699,705	GSPGGGNTAPSTPVIGGSDKPGMEEK

Showing 1 to 10 of 916 entries

Previous 1 2 3 4 5 ... 92 Next

Seqwindows_MultiSites: there are two situations here: First, the modified amino acid will be replaced with "X" if it is not the central residue, for example, 'NKPTSLWNPT(0.832)Y(0.168)GSWFTEK' has two phosphosites, one is the 10th amino acid with 0.832 location probability, the other is the 11th amino acid with 0.168 location probability, thus if we transform it like 'NKPTSLWNPT#Y@GSWFTEK' (high probability is replaced with '#', while low probability is replaced with '@'). Then in motifer, the 10th amino acid will be considered as central residue, the 11th amino acid will be replaced with "X", thus the standard sequence is 'PTSLWNPTYGSWFTE', correspondingly, the Seqwindows_MultiSites should be 'PTSLWNPTXGSWFTE'. Second, if we transform this peptide like 'NKPTSLWNPT#Y#GSWFTEK', the

two amino acids will be both considered as central residue, thus the standard sequence is 'PTSLWNPTYGSWFTE;TSLWNPTYGSWFTEK', correspondingly, the Seqwindows_MultiSites is still 'PTSLWNPTYGSWFTE;TSLWNPTYGSWFTEK'.

4. Own Background

Users can upload their own background database, but it is noteworthy that the database must contain peptide sequences with standard length (for example, peptide generated in previous steps by motifeR), not protein sequences.

motifeR

Welcome Import Data Pre-alignment **Own Background** Motif Enrichment Motif Plot

☒ Upload your own background data ?

File format :

☒ .xlsx ☐ .xls ☐ .csv/txt

Import your data :

Browse... No file selected

☐ Header ?

☐ First column ?

Sheet:

1

Users can click “Browse” button and import their data, like this:

Welcome Import Data Pre-alignment **Own Background**

Show 10 entries

☒ Upload your own background data ?

File format :

☐ .xlsx ☐ .xls ☒ .csv/txt

Import your data :

Browse... Seqwindows_database.csv

Upload complete

☒ Header ?

☐ First column ?

Separator :

☒ Comma ☐ Semicolon ☐ Tab ☐ BlankSpace

	Seqwindows
1	SDRRKINSAPSSPIK
2	KINSAPSSPIKTNKA
3	KMLISAVSPEIRNRD
4	SDRRKINSAPSSPIK
5	KINSAPSSPIKTNKA
6	SDRRKINSAPSSPIK
7	RKINSAPSSPIKTNK
8	KINSAPSSPIKTNKA
9	GSQGEPWTP TANLKM
10	QGEPWTP TANLKMLI

Showing 1 to 10 of 75,547 entries

Please note, if you upload you own fasta file as background database, you should unselect this parameter. If you choose this parameter stubbornly, this software will take the data in this step as background database and ignore the fasta file that you upload before.

5. Motif Enrichment

This step will find overrepresented sequence motifs.

5.1 Parameters

☒ Species data as background ?

☐ Only use multi-site data ?

Calculate

Species data as background: if you upload your own fasta file as background database in the 'Import Data' step, you can ignore this parameter (select or unselect is same). Otherwise, if you choose the database in our system (i.e., human) in the 'Import Data' step, selecting this parameter means this software will take the database in our system as background database. If you don't choose, the software will take the foreground data as background database.

Only use multi-site data: if selected, this tool will only take the peptides with multi modification sites as foreground data, that is, it will use the sequences in the *Seqwindows_MultiSites* column obtain from 'Pre-alignment' step as foreground data.

5.2 Results

The enrichment results like this:

Welcome Import Data Pre-alignment Own Background Motif Enrichment Motif Plot Kinase Analysis Building Species Database

Multiple motifs Regular sequence motif

Download

Show 10 entries Search:

	motif	score	fg.matches	fg.size	bg.matches	bg.size	fold.increase	Enrich.seq
1P.SP.....	317.89810602509	120	1726	7499	902328	8.36568552902457	KINSAPSSPIKTNKA:NVGSPPKSPTHASPQ:NSAVTLPSPGSSPFP:VQWLNQSPSTTTSSN:PKKI
2SP.....	307.652655568589	451	1606	57796	894629	4.34783714946362	KIMLISAVSPEIRNRD:IKSHMSGSPGPGGSN:SSSSLLASPGHISVK:KKKKNRHSDPHGMG:VKC
3SSP.....	615.305311137178	70	1155	7774	837033	6.52550459573871	RKINSAPSSPIKTNK:VTAEADSSSPTGILA:TSSLDLSSPSPVTT:FSKERSPSSPVVVK:SKETC
4S.SP.....	318.541773931106	47	1085	6581	829259	5.45841337686974	TSIKSHMSGSPGPGG:LKSPVSESVSPVVPD:FPEPTCLASAPPNAP:PGTPYKVCSCPTSGA:SGI
5	...R..S.....	307.652655568589	160	1038	47281	822678	2.68204375928299	HDLGRSSSSLLASPG:PPQARTSSLDNEGPH:TAGCRGSSAVLNVT:VPLRRRHSEQVANGP:WQ
6SD.D.....	24.6017271480482	29	878	2832	775397	9.0434543287042	PVPPETPSDSOHKKK:SAGYEEISDPDMEEK:LKRRLSYSDSLKRA:EDGEEDSDSDYEIS:TLQ
7SD.E.....	22.3583226679643	30	849	4154	772565	6.57176615497685	EGEEDSDSDYEISAK:GEAPEPDSDAEVAEA:SRFFTTGSDSESESS:FHYRTLHSDDEGTVL:GEI
8S.S.....	10.4709411486104	150	819	81386	768411	1.72922388844046	DLGRSSSSLLASPGH:PPETPSDSOHKKKKK:GLVYKSGSGEIGSET:LATSEKSMFVLGSV:PKLI
9S.S.....	10.4914122398175	130	669	73679	687025	1.81194934573683	EPTPSIASDISLP:IA:GRKTSIKSHMSGSPG:PSIASDISLP:ATQE:NTLKSVPVSESV:PVV:TPTSSLC
10S...D...	8.09910955740197	58	539	28674	613346	2.30174116480277	DLEVFRNSLYAPDYS:KNSMPTVSFLDQDQS:GKQPLLLSEDEEDTK:RLHGGFDSDCSEDEGE:GI

Showing 1 to 10 of 13 entries

Previous 1 2 Next

motif: the overrepresented motif.

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.

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

fg.size: total number of foreground sequences

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

bg.size: total number of background sequences

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

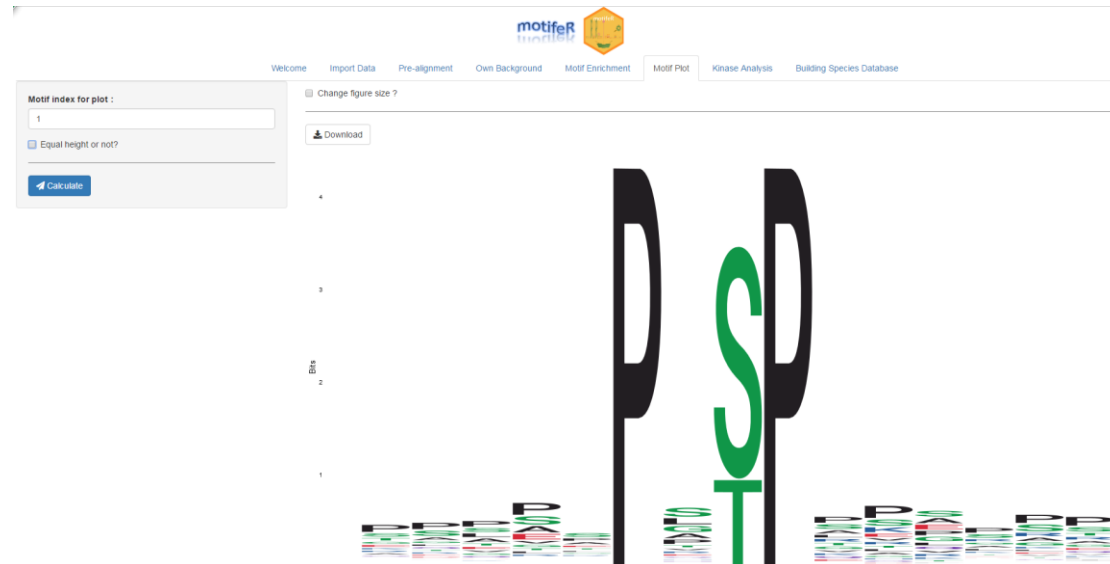
Enrich.seq: those peptides are overrepresented in this motif.

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

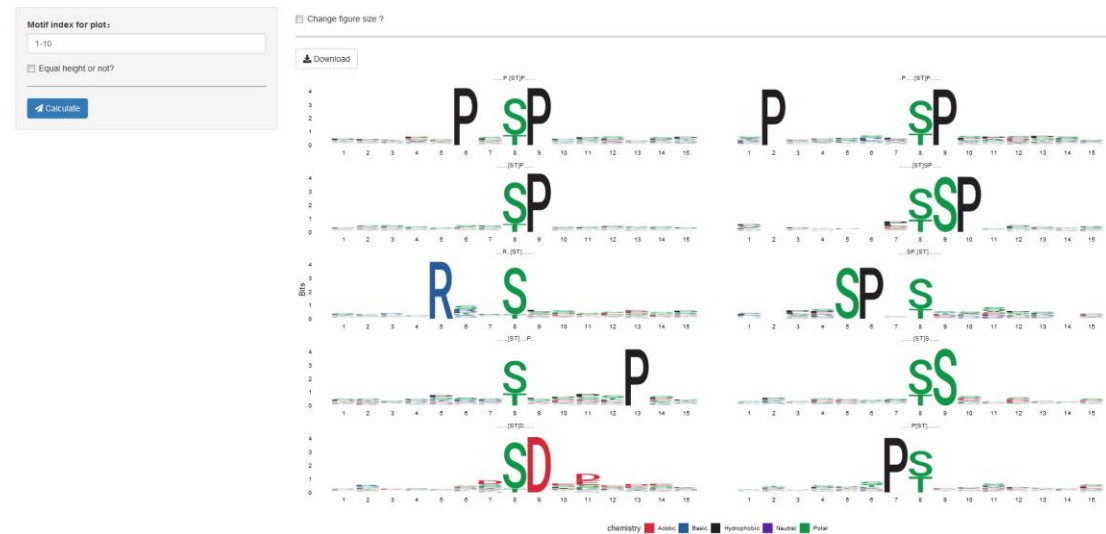
6. Motif Plot

This step will plot the motifs from the enrichment results.

If users only input one number in the 'Motif index for plot' parameter, it will plot the relative motif, shown as below:



If users type in '1-10', it will plot the 1th to 10th motifs, like this:



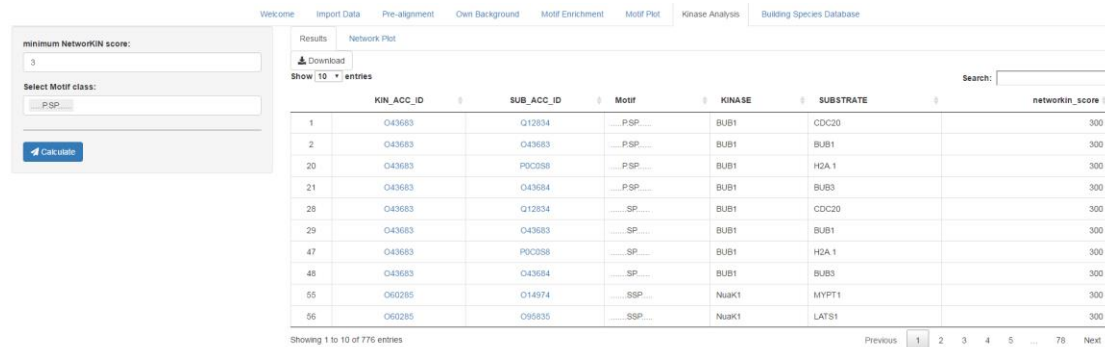
In addition, if users select ‘Equal height or not?’ parameter, which means whether all residues in the figure have equal height. This plot will be like:



7. Kinase-Substrate Analysis

This step will take a formatted phosphoproteomics data input and perform kinase-substrate analysis calculations to infer relative kinase activities. Users should note here:

- There are only Human database for this analysis in this system currently.
- This is only for phosphoproteomics data, other modification data are not inappropriate.



The interface includes a sidebar on the left with the following controls:

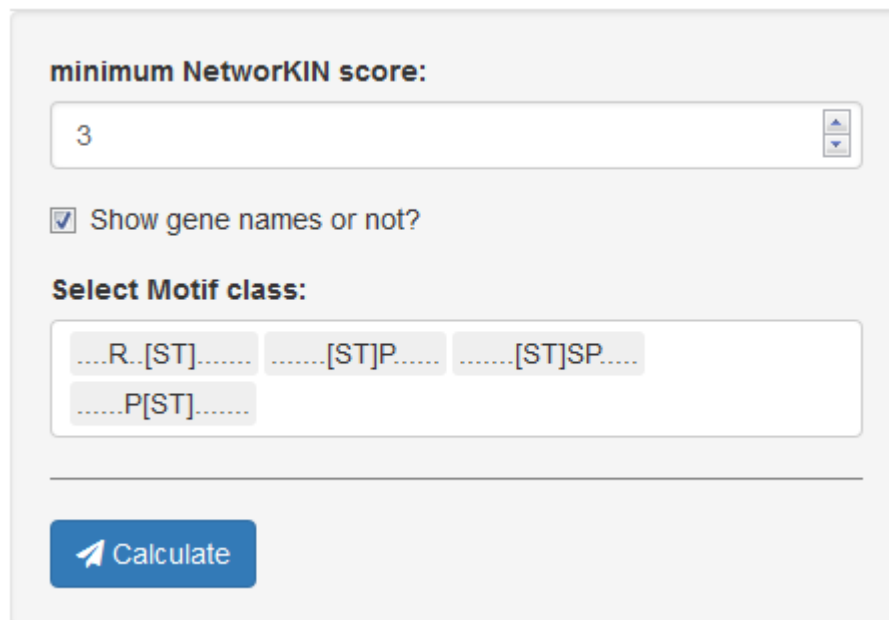
- minimum NetworkKIN score:** A text input field containing the value '3'.
- Select Motif class:** A dropdown menu currently showing 'PSP'.
- Calculate** button.

The main area displays a table of results with the following columns: KIN_ACC_ID, SUB_ACC_ID, Motif, KINASE, SUBSTRATE, and networkkin_score. The table shows 10 entries, with a 'Show 10 entries' link and a search bar. The data is as follows:

	KIN_ACC_ID	SUB_ACC_ID	Motif	KINASE	SUBSTRATE	networkkin_score
1	O43683	Q12834	...PSP...	BUB1	CDC20	300
2	O43683	O43683	...PSP...	BUB1	BUB1	300
20	O43683	P0C058	...PSP...	BUB1	H2A 1	300
21	O43683	O43684	...PSP...	BUB1	BUB3	300
28	O43683	Q12834	...SP...	BUB1	CDC20	300
29	O43683	O43683	...SP...	BUB1	BUB1	300
47	O43683	P0C058	...SP...	BUB1	H2A 1	300
48	O43683	O43684	...SP...	BUB1	BUB3	300
55	O60285	O14974	...SSP...	Nuak1	MYPT1	300
56	O60285	O55835	...SSP...	Nuak1	LATS1	300

At the bottom, it indicates 'Showing 1 to 10 of 776 entries' and a pagination control showing 'Previous 1 2 3 4 5 ... 78 Next'.

7.1 Parameters



The parameter form includes the following sections:

- minimum NetworkKIN score:** A text input field with the value '3' and a spinner control.
- Show gene names or not:** A checkbox that is checked.
- Select Motif class:** A container with four buttons: '...R..[ST]...', '.....[ST]P...', '.....[ST]SP...', and '.....P[ST]...'.
- Calculate** button.

minimum NetworkKIN score: a numeric value between 1 and infinity setting the minimum NetworkKIN score.³

Show gene names or not: if true, the gene names will be appeared in the network plot, otherwise, the uniprot ids will be shown.

Select Motif class: select those motifs you want to analyze, these motifs are obtained from 'Motif Enrichment' step:

Select Motif class:

.....PSP.....SP..... |

.....SSP.....

.....S...E.....

.....S.S.....

.....R..S.....

.....S.SP.....

.....SD.E.....

.....S...P.....

.....S.S.....

7.2 Results

First, this software give a result table containing kinases and substrates information.

Welcome Import Data Pre-alignment Own Background Motif Enrichment Motif Plot Kinase Analysis Building Species Database

Results Network Plot

Download

Show 10 entries Search:

	KIN_ACC_ID	SUB_ACC_ID	Motif	KINASE	SUBSTRATE	networkin_score
1	O43683	Q12834PSP.....	BUB1	CDC20	300
2	O43683	O43683PSP.....	BUB1	BUB1	300
20	O43683	P0C0S8PSP.....	BUB1	H2A.1	300
21	O43683	O43684PSP.....	BUB1	BUB3	300
28	O43683	Q12834SP.....	BUB1	CDC20	300
29	O43683	O43683SP.....	BUB1	BUB1	300
47	O43683	P0C0S8SP.....	BUB1	H2A.1	300
48	O43683	O43684SP.....	BUB1	BUB3	300
55	O60285	O14974SSP.....	NuaK1	MYPT1	300
56	O60285	O95835SSP.....	NuaK1	LATS1	300

Showing 1 to 10 of 776 entries Previous 1 2 3 4 5 ... 78 Next

KIN_ACC_ID: kinase uniprot id.

SUB_ACC_ID: substrate uniprot id.

Motif: the overrepresented motif.

KINASE: kinase gene name.

SUBSTRATE: substrate name.

networkin_score: the prediction score from networkIN (<https://networkin.info/>).

Second, this tool will plot the kinase-substrate network:

minimum NetworkKIN score:

3

☒ Show gene names or not?

Select Motif class:

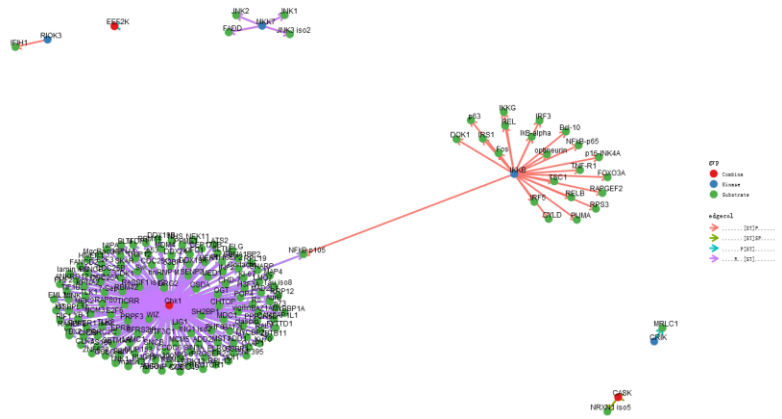
Calculate

Results

Network Plot

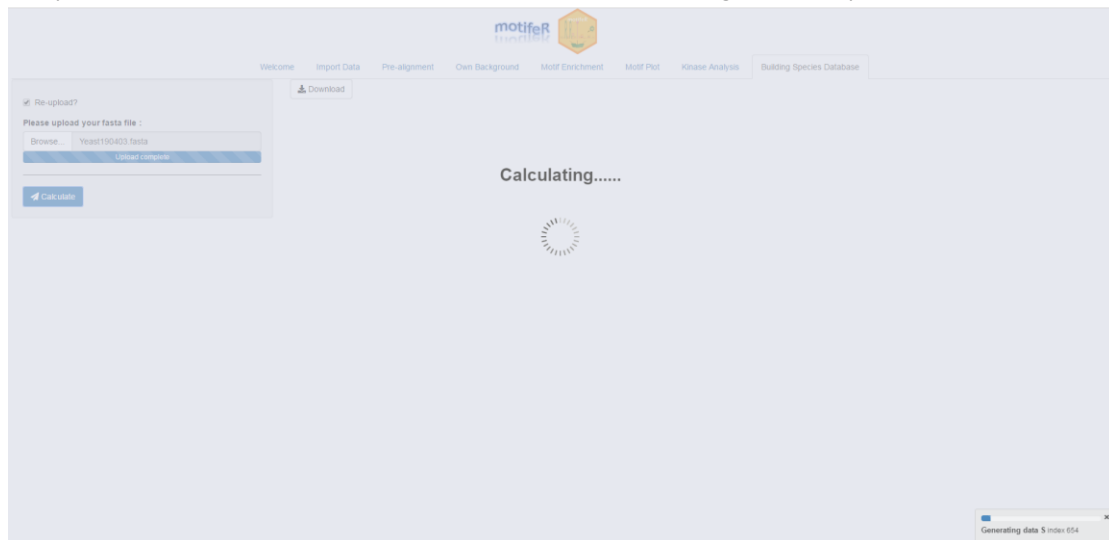
☐ Change figure size?

Download



8. Building Species Database

This step can build the standard database based on the fasta file that users upload, herein there is no species limit. And this results can also be used in 'Own background' step.



If users want to build their own database, they can select the 'Re-upload' parameter and then upload a fasta file, the results will be shown as below:

Welcome

Import Data

Pre-alignment

Own Background

Motif Enrichment

Motif Plot

Kinase Analysis

Building Species Database

Download

Show

10

▼

entries

Search:

	ID	Windows	Center
1	P33755	_MLIRFRSKNGTHRV	S
2	P33755	KNGTHRVSCQENDLF	S
3	P33755	GQGIHVAISELADRTV	S
4	P33755	DMLILNYSDKPFANEK	S
5	P33755	GVNVEIGSVGIDSKG	S
6	P33755	IGSVGIDSKGIQHR	S
7	P33755	GLIPRQKSKLCKHGD	S
8	P33755	RGMCEYCSPLPPWDK	S
9	P33755	KNKIKHISFHSYLKK	S
10	P33755	IKHISFHSYLKKLNE	S

Showing 1 to 10 of 274,123 entries

Previous

1

2

3

4

5

...

27413

Next

ID: uniprot ids.

Windows: the standard peptides.

Center: Center residue.

Those basic parameters in 'Import Data' step are also usable here. Therefore, if you want to get different results, just change those parameters.

All results can be saved to .pdf or .csv files by clicking corresponding "Download" button.

III. Case Study

The detailed parameters introduction can be found in supplementary notes above. Herein we mainly analyze a published data as a real example for users to operate this software better.

Step 1, data preparation

Here we just take a recently published data as an example, users can prepare their own data in this way. The example data can be downloaded from here (doi: 10.1126/scisignal.aap9752)¹:

https://stke.sciencemag.org/highwire/filestream/214464/field_highwire_adjunct_files/2/aap9752

[2 Data File S3.xlsx](#)

Download this data and then open it in Excel, the data is like this:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
id	Uniprot ID	Gene names	Protein names	Position +/- 6AA	Sequence window	Localization probability (Class I)	Phospho (STY) Probabilities	Charge	Mass error [ppm]	Score	Ratio (DMSO/cr isotinib) Exp1	Ratio (DMSO/cr isotinib) Exp3	Ratio (DMSO/cr isotinib) Exp4	Ratio (DMSO/cr isotinib) Exp2	Ratio (DMSO/LD K378) Exp2	Ratio (DMSO/LD K378) Exp3
15454	Q98B87	AAC3BP1	L-aminoadipate-semialdehyde dehydrogenase	2233	EWAFESIDEHHR	1.00	LFLDGEKEEKEWAFES(1)K	4	0.42	33.25						
176	A0A087VW41	AATF		549	GEDEGEDGEGSYKLC	1.00	FDEGEDGEGS(0.996)NY(0.004)KKLC	3	-0.36	31.88	6.24	61.96	19.77	51.84	51.84	98.02
18369	Q98DCT	ARCA2	ATP-binding cassette sub-family A member 2	12412	EDGLDTEDEGLI	1.00	ALVADEPEDLDT(1)EDEGLISFEEER	3	-1.70	137.28						
15675	Q9H7B9	AB12	Abi1 interactor 2	2326	Q9H7YSSSGSSG	0.98	TYS(0.98)S(0.02)SGSSGSHFPSSR	2	-0.97	123.22		4.47				4.47
969	Q9CAP0	ACD	Adrenocortical dysplasia protein homolog	5110	LGSETFSSPFAAGQ	0.78	ELILGS(0.002)ET(0.052)PS(0.167)PR	2	-0.73	51.53				0.19	3.27	
2261	ETBQ74	ACIN1	Apoptotic chromatin condensation inducer	2348	EPKESPSFPRTE	1.00	S(0.008)KS(0.992)PS(0.999)PPRLT(0.001)EDR	3	0.52	98.18	0.80	1.92	0.72	1.43	2.08	1.75
2250	ETBQ74	ACIN1	Apoptotic chromatin condensation inducer	2303	AKLSEGSQFAEE	1.00	AAKLS(1)EGS(1)QPAEEEEEDQETPSSR	3	1.49	122.81	1.24	2.47	1.16	1.09	1.47	2.56
2249	ETBQ74	ACIN1	Apoptotic chromatin condensation inducer	2200	AAAKLSEGSQFA	1.00	AAKLS(1)EGS(1)QPAEEEEEDQETPSSR	3	1.49	122.81	1.24	2.47	1.16	1.09	1.47	2.56
18815	Q98B87	ACRBP	Acrosin-binding protein	1208	ENRQETQEHQEQ	1.00	QEPT(1)QEHKQEEGQKQEEQEEQEEGK	4	-0.65	27.54				0.04	2.26	
9967	Q9F7E4	ACS2	Long-chain-fatty-acid-CoA ligase II	3358	IGFVNSQKHLGR	1.00	NIGFKVNS(1)K	2	2.48	71.03	4.49	495.07	0.77			1525.50
15724	Q98D32	ACTR10	Actin-related protein 10	2414	LMPHAPSTSE	1.00	RAFS(0.995)T(0.005)EK	2	0.47	107.21	1.32	2.93	2.30	1.85	2.14	1.67
7774	P61163	ACTR1A	Alpha-actinin	570	ENRGLLSIRVPMK	0.96	AAEHRGLLS(0.959)IRY(0.041)PMEHGIIVK	4	-0.69	25.25						
12227	Q98H55	AFAP1L2	Actin filament-associated protein 1-like 2	3353	LEPVERSLSTSSY	0.91	S(0.91)LET(0.078)S(0.004)Y(0.002)LNVLVNS(0.001)QWK	3	-1.28	33.25	171.76			171.76	74.50	66.93
16163	Q98B87	AF4	AF4/FMR2 family member 4	2836	SRKRTISQSSSLK	0.99	RT(0.012)IS(0.986)QS(0.003)SSLK	2	0.17	97.77	4.15	2.84	4.87	4.45	2.16	2.33
20919	Q9ULP2	AFTF	Aftaphilin	1617	NIDTPTPTSTSV	0.88	TDENIDT(0.124)PGT(0.876)PK	2	-0.90	105.39	1.58			1.54	1.63	2.06
8457	Q98666	ABNAK	Neuroblast differentiation-associated protein 1	25780	YFPRNSPFSDE	1.00	HRS(0.998)NS(0.002)FSDER	2	-1.16	268.33	3.68	2.18	2.24	3.06	1.74	2.54
8468	Q98666	ABNAK	Neuroblast differentiation-associated protein 1	2212	IRLPSGSGAASPT	1.00	LPS(0.002)GS(0.998)GAAS(0.996)PT(0.004)GSAVDIR	2	0.24	295.96	1.81	2.46	1.90	1.31	1.50	2.28
8466	Q98666	ABNAK	Neuroblast differentiation-associated protein 1	2135	IKFLKSEADGVEG	1.00	S(1)EDGVEGDLGETQSR	2	-0.51	272.63	2.91	2.48	2.06	2.57	1.64	2.33
8473	Q98666	ABNAK	Neuroblast differentiation-associated protein 1	35099	VEFDIKSPKFAE	1.00	MYFPDVEFDIKS(1)PK	3	-0.82	133.93	5.96	8.45	9.15	9.01	7.81	5.97
8461	Q98666	ABNAK	Neuroblast differentiation-associated protein 1	32397	LGHLKSPKFAE	1.00	ISMPDLHLKS(1)PK	3	-0.22	114.55	4.00	3.69	3.17	3.57	2.40	2.84
8475	Q98666	ABNAK	Neuroblast differentiation-associated protein 1	33426	VELMLKSPKFAE	1.00	VSMFDVELNLKS(1)PK	3	-0.66	114.55	4.29	6.61	3.86	4.67	4.28	4.95
11854	Q81YF2	ABNAK2	Protein ABNAK2	1160	SQRRLSNFPQK	1.00	ERLS(1)WPK	2	0.15	76.28	1.95	2.13	2.13	2.36	1.88	2.13
16313	Q98H44	AKAP11	A-kinase anchor protein 11	5422	TPRFESFYGLK	0.99	KPES(0.991)PY(0.003)GNLCDAFDS(0.006)PRPVK	3	1.16	92.41	1.90	10.26	2.75	1.54	1.31	6.64
8540	Q12802	AKAP13	A-kinase anchor protein 13	51876	FPRPRSAVLIVD	1.00	ERPRS(1)AVLLVDETATTPIFANR	3	-0.41	154.71	1.46	1.29	1.85	1.82	2.02	1.81
16039	Q97D6	AKAP2	A-kinase anchor protein 2	5951	RQRTLNITEEL	1.00	TLS(1)MIEEEIR	2	-0.49	105.65	2.93	1.01	2.41			2.40
21435	Q98536	AKT1S1	Proline-rich Akt1 substrate 1	1246	PRPRNTSDPGK	1.00	LMT(0.999)S(0.001)DPGK	2	0.12	139.54	11.13	2.95	9.45	11.63	6.22	6.77

We can extract the column “Phospho (STY) Probabilities” data in a new file, as this:

	A	B	C
1	Phospho_STY_Probabilities		
2	LFLDGEKEEKEWAFES(1)K		
3	FDEGEDGEGS(0.996)NY(0.004)KKLC		
4	ALVADEPEDLDT(1)EDEGLISFEEER		
5	TYS(0.98)S(0.02)SGSSGSHFPSSR		
6	ELILGS(0.002)ET(0.052)PS(0.779)S(0.167)PR		
7	S(0.008)KS(0.992)PS(0.999)PPRLT(0.001)EDR		
8	AAKLS(1)EGS(1)QPAEEEEEDQETPSSR		
9	AAKLS(1)EGS(1)QPAEEEEEDQETPSSR		
10	QEPT(1)QEHKQEEGQKQEEQEEQEEGK		
11	NIGFKVNS(1)K		
12	RAFS(0.995)T(0.005)EK		
13	AAEHRGLLS(0.959)IRY(0.041)PMEHGIIVK		
14	S(0.91)LET(0.078)S(0.004)S(0.004)Y(0.002)LNVLVNS(0.001)QWK		
15	RT(0.012)IS(0.986)QS(0.003)SSLK		
16	TDENIDT(0.124)PGT(0.876)PK		
17	HRS(0.998)NS(0.002)FSDER		
18	LPS(0.002)GS(0.998)GAAS(0.996)PT(0.004)GSAVDIR		
19	S(1)EDGVEGDLGETQSR		
20	MYFPDVEFDIKS(1)PK		
21	ISMPDLHLKS(1)PK		
22	VSMFDVELNLKS(1)PK		
23	ERLS(1)WPK		
24	KPES(0.991)PY(0.003)GNLCDAFDS(0.006)PRPVK		
25	ERPRS(1)AVLLVDETATTPIFANR		
26	TLS(1)MIEEEIR		

Then we need convert these this location probability into a specific label, so that motifer can recognize. For example, here we use ‘#’ as the specific label:

‘FDEGEDGEGS(0.996)NY(0.004)KKLC’ to ‘FDEGEDGEGS#NY#KKLC’

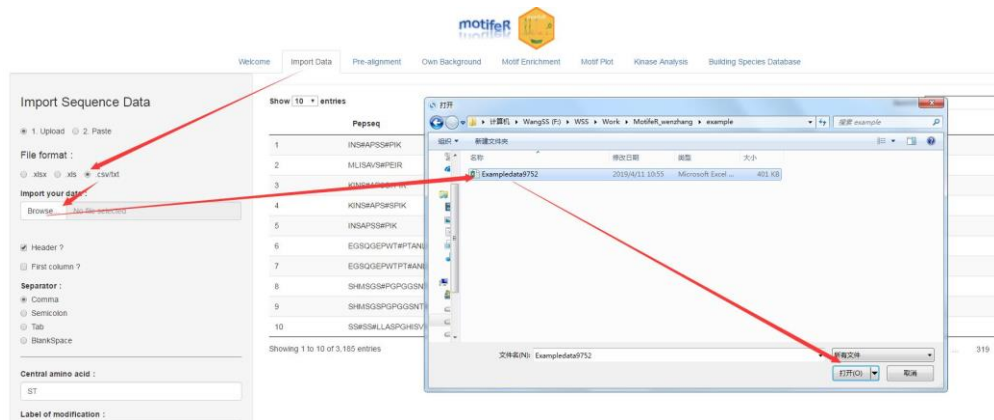
No matter what method you use (e.g. writing some R/Python codes to convert in bulk, or process it manually in Excel, and so on), your final data should be like this:

	A
1	Phospho_STY_Probabilities
2	LFLDGEEEEWAFES#K
3	FDEGEDGEGS#NY#KKLC
4	ALVADEPEDLD#EDEGLISFEER
5	TYS#S#SGSSGSHPSR
6	ELILGS#ET#PS#S#PR
7	S#KS#PS#PPRLT#EDR
8	AAKLS#EGS#QPAEEEDQETPSR
9	AAKLS#EGS#QPAEEEDQETPSR
10	QEPT#QEHKQEEGQKQEEQEEEGK
11	NICFKVNS#K
12	RAFS#T#EK
13	ABEHRGLLS#IRY#PMEHGIVK
14	S#LET#S#S#Y#LNVLVNS#QWK
15	RT#IS#QS#SSLK
16	TDENIDT#PGT#PK
17	HRS#WS#FSDER
18	LPS#GS#CAAS#PT#CSAVIDR
19	S#EDGVEGDLGETQSR
20	MYFPDVEFDIKS#PK
21	ISMFDLDLHLKS#PK
22	VSMFDVELNLKS#PK
23	ERLS#WPK
24	KPES#PY#GNLCDAPDS#PRPVK
25	ERPRS#AVLLVDETATTPIFANR

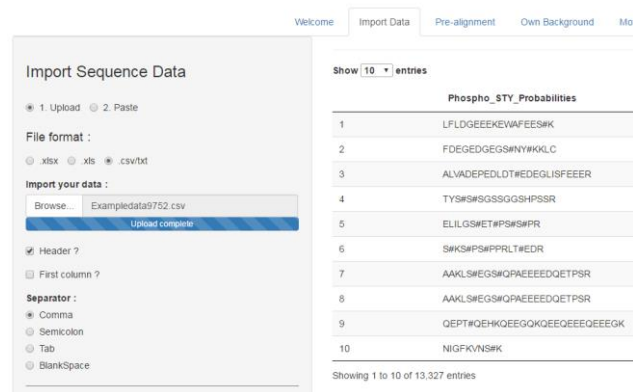
Save this data in a .xlsx, .xls, .csv or .txt file, here we save it in the “Exempladata9752.csv” file.

Step 2, Analysis in motifeR

First, open the software: <https://www.omicsolution.org/wukong/motifeR>. Click “Import Data”, find our data and then import:



Users must set right parameters based on their own data. Now we import the example data:



Second, set the basic parameters:

Central amino acid :

ST

Label of modification :

#

Width :

7

Minimum number :

20

P-value threshold :

0.000001

☐ 1. Select ☒ 2. Upload

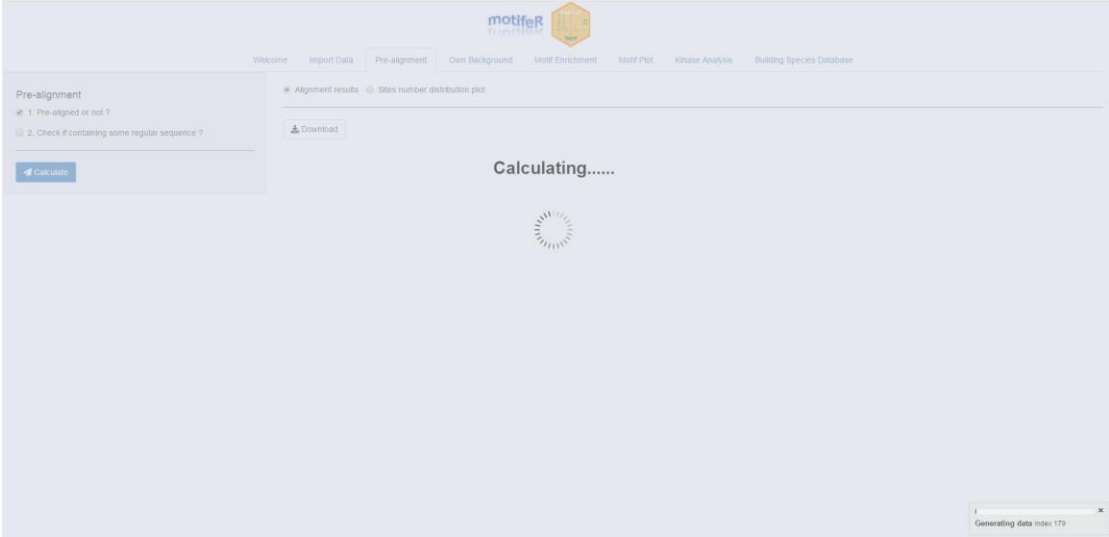
Please upload your fasta file :

Browse... 9606.fasta

Upload complete

Base on the example data, we just leave basic parameter as default, whereas, we upload a new fasta file as the background database.

Third, Click “Pre-alignment”. Herein we don’t want to check if the example data contain some regular sequences, but the peptides in example data are not standard, we need pre-align them with the background database. Then click “Calculate” button:



After a while (normally in a few seconds), this step can be done. The total calculation time is directly proportional to the data size. We can get the alignment results like this:

Alignment results | Sites number distribution plot

Download

Show 10 entries

Search:

Pep.upload	Pep.no	Pep.main.index	Pep.all.index	Seqwindows	PRO.from.Database	PROindex.from.Database
LFLDGEEKEVAFEESSK	LFLDGEEKEVAFEESSK	16	16	KEVAFEESSKDEHMF	Q9NRN7	233
ALVADEPEDLTWDEGLISFEER	ALVADEPEDLTWDEGLISFEER	12	12	DEPEDLTWDEGLIS	Q9BZC7	2412
TYSSSGSGSGGSHPPSSR	TYSSSGSGSGGSHPPSSR	3,4	3,4	NORNRITYSSSGSGG,QRNRTYSSSGSGGS	Q9NYB9	242,243
ELILGSETWPSHPR	ELILGSETPSR	6,8,10,11	6,8,10,11	IRELILGSETPSR,ELILGSETPSR,AGILGSETPSR,AGQLLGSETPSR,AGQLL	Q96AP0	106,108,110,111
SKSPSPRPRLTWEDR	SKSPSPRPRLTEDR	1,3,5,10	1,3,5,10	SQLKEKSKSPSPR,QLKEKSKSPSPRLT,KEKSKSPSPRLTED,SPSPRPRLTED,SKAS	Q9UKV3	384,386,388,393
AAKLSEGSQPAEEEDQETPSR	AAKLSEGSQPAEEEDQETPSR	5,8	5,8	QARAALSEGSQPAE,AAKLSEGSQPAEEEE	Q9UKV3	240,243
AAKLSEGSQPAEEEDQETPSR	AAKLSEGSQPAEEEDQETPSR	5,8	5,8	QARAALSEGSQPAE,AAKLSEGSQPAEEEE	Q9UKV3	240,243
QEPHQEHKQEGGQKQEGQEEQEEGK	QEPHQEHKQEGGQKQEGQEEQEEGK	4	4	VEHQEPHQEHKQEE	Q9NEB7	208
NIGFKVNSK	NIGFKVNSK	8	8	NIGFKVNSK,ILGKY	Q9PVE4	358
RAFSTWTEK	RAFSTWTEK	4,5	4,5	PLMKRAFSTWTEK,LMKRAFSTWTEK	Q9NZ32	414,415

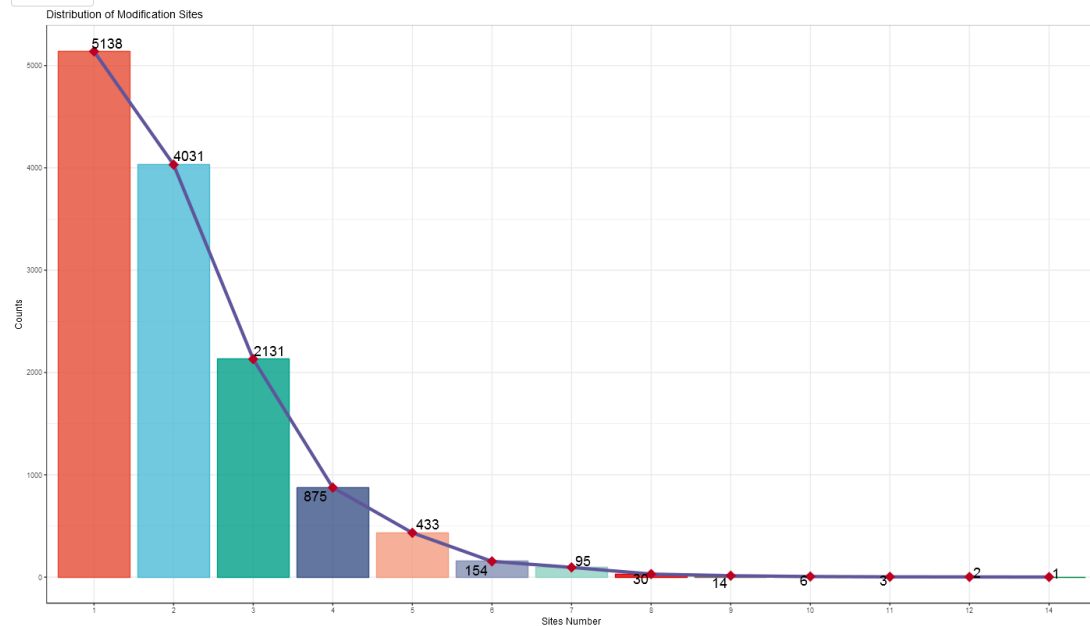
Showing 1 to 10 of 12,962 entries

Previous 1 2 3 4 5 ... 1297 Next

Users can also click the “Download” button to download the results into their own computer. Then we can click “Sites number distribution plot” to obtain multiple sites information:

Plot:

Download



Multi-Sites Data:

Download

Show 10 entries

Search:

Pep.upload	Pep.no	Pep.main.index	Pep.all.index	Seqwindows
4 TYSS#SGSSGGSHPSR	TYSSGSSGGSHPSR	3,4	3,4	NQRNRTYSSGSSGG;QRNRTYSSGSSGGS
5 ELILGS#ET#PS#S#PR	ELILGETPSSPR	6,8;10;11	6,8;10;11	IRELILGETPSSPR;ELILGETPSSPRAG;ILGETPSSPRAGQL;LGSETPSSPRAGQLL
6 S#KS#PS#PPRLT#EDR	SKSPSPRLTEDR	1;3;5;10	1;3;5;10	SOGLKEKSKSPSPR;GLKEKSKSPSPRLT;KEKSKSPSPRLTED;SPSPRLTEDRKAS
7 AAKLS#EGS#QPAEEEDQETPSR	AAKLEGSQPAEEEDQETPSR	5,8	5,8	QARAALKLEGSQPAE;AAKLEGSQPAEEEE
8 AAKLS#EGS#QPAEEEDQETPSR	AAKLEGSQPAEEEDQETPSR	5,8	5,8	QARAALKLEGSQPAE;AAKLEGSQPAEEEE
11 RAFS#T#EK	RAFSTEK	4,5	4,5	PLMKRAFSTEK____;LMKRAFSTEK____
12 AEEHRLLS#IRY#PMEHGIVK	AEEHRLLSIRYPMEHGIVK	9;12	9;12	EEHRLLSIRYPMEH;RGLLSIRYPMEHGIV
13 S#LET#S#S#Y#LNLVNS#QWK	SLETSSYLNVLNSQWK	1,4;5;6;7;14	1,4;5;6;7;14	SLEPVERSLETSSYL;PVERSLETSSYLNVL;VERSLETSSYLNVL;RSLETSSYLNVLNS
14 RT#IS#QS#SSLK	RTISQSSSLK	2,4;6	2,4;6	EHGSRKRRTISQSSSL;GSRKRRTISQSSSLKS;RKRTISQSSSLKSSS
15 TDENIDT#PGT#PK	TDENIDTPGTPK	7;10	7;10	RTDENIDTPGTPKTH;ENIDTPGTPKTHSVP

Showing 1 to 10 of 7,824 entries

Previous 1 2 3 4 5 ... 783 Next

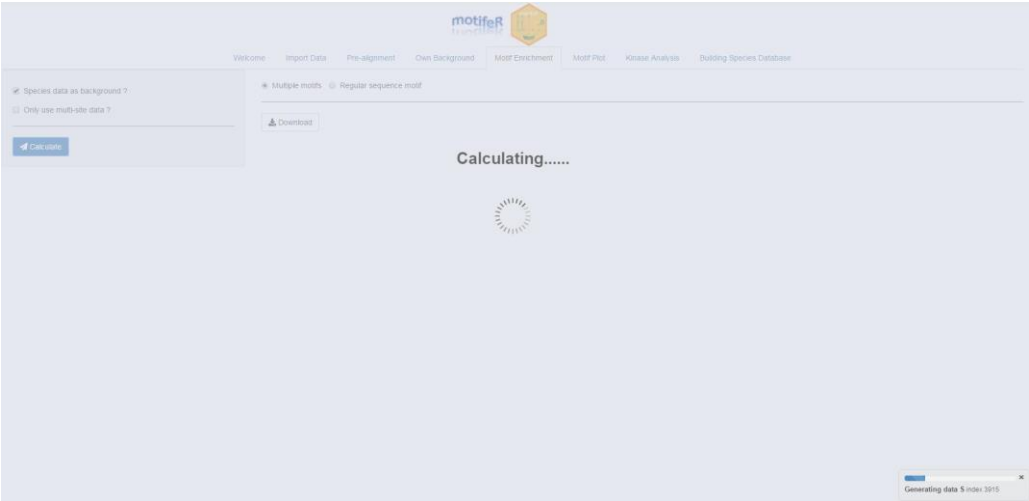
If there are multiple modifications in peptides, this part can give users necessary information. From here we can know there are total 7824 peptides with >2 phosphorylation sites.

Fourth, click “Own Background”. As we have uploaded fasta file as background database, here we needn’t upload background data again, just keep it as default.

[Welcome](#)
[Import Data](#)
[Pre-alignment](#)
[Own Background](#)
[Motif Enrichment](#)
[Motif Plot](#)
[Kinase Analysis](#)
[Building Species Database](#)

☐ Upload your own background data ?

Fifth, click “Motif Enrichment”. As we upload fasta file as background database and don’t check these peptides if containing some regular sequence, here the default parameter set is ok. Then click “Calculate” button.



Then motifeR would generate the standard background database based on the fasta file we upload and then calculate the significant motifs. This step will take relatively a long time, especially when we analyze >10000 peptides. Now we can take a cup of coffee and wait some minutes.

Multiple motifs Regular sequence motif

Download

Show 10 entries Search:

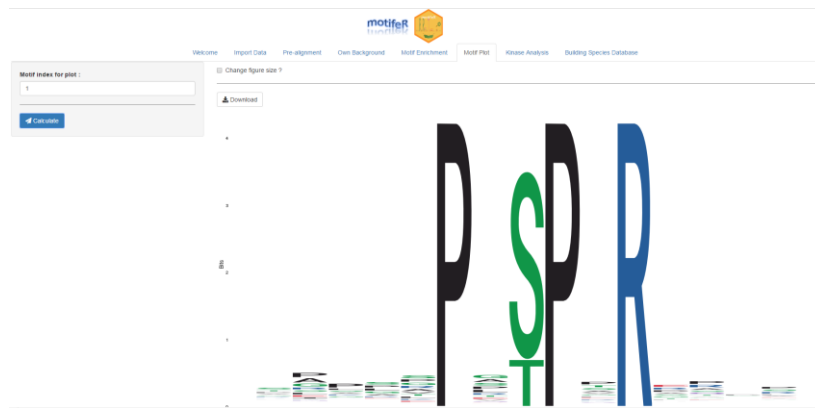
motif	score	fg.matches	fg.size	bg.matches	bg.size	fold.increase	Enrich.seq
1 ...P [ST]P.R	625.448696013719	135	20869	697	1478013	13.7175832736192	TPGPGPGSPORRAQQ.GEAGGLTPRRVSSD.MATEPPSLRVEAP.PISPGGLTPMREKDV.LLPTPLSPSRSSQL.LLPTPLSPSRGFAE.EFWLLRSPSPRRL.DF
2 ...RP [ST]P...	623.101621850051	112	20734	630	1477316	12.6668252896477	ESMERPESPKEFLDL.TTSPRPASVPQSER.RGNSRPGTSAEGGS.SLSRPASPQAQWPO.DGPAPASPEAGNTL.VRLRRPLSPETRRRR.KSSERLSPKAWKGF.F
3 ...P [ST]PK...	623.321612780932	86	20622	455	1476686	13.5345689709379	EDDTVPLSPLKYMAQ.QAQSPTTPEKTDLTA.VPKPPLSPKLSV.KKKKDPFSPDKKKPV.ECRSDPESPKKTL.SLKPTVYKPSPEKSKPD.TTMLPASPAKAPET.PNKEL
4 ...R. [ST]P.P...	922.957966705766	150	20536	644	1476231	16.7434078799962	SKMQRSHSPVAAAP.RERYQRSPSPAPAP.RYRQRSPSPAPAP.PYKRRRSTPAKKEE.ENARRRGTPEEAG.GESRPTLSPTTSAEG.KSDSRSESPGYYV.LI
5 ...P [ST]P.K...	622.233410206199	77	20386	432	1475587	12.9014872906606	NLCDAPDSRPV.KAS.SITSPPLSPALPKYK.SQSRGPASPLEAKKK.DLPKEPTTPEQKDW.DPSKPVTPSPSKLV.ETAPPPSPVSEKLI.EEEDPSGIEKID.TSPSKP
6 ...P [ST]PK...	621.747127649934	66	20309	377	1475155	12.7160346127137	ENIDTPTGKTHSVF.PLSPPLTPKATRTL.DGVEGRTPKYMINN.KGEKTPKPKGPSSV.GSLTPPSSPKTQAG.PKTQSPHSKPEESER.KASEXPVSPKSGTLK.GRI
7 ...P [ST]P.K...	621.402592239037	60	20243	351	1474778	12.4536285836488	SSPLSPLSGIKSPT.APREELTPRLKEG.TEKHLVSPSGKTEK.PYAGAPGSPRTKRKL.DWRGPPNSPOLKSLD.VPVL.PASPPKDSL.MEQDQPKSPSHKSR.VGI
8 ...K. [ST]PP...	621.569083259662	60	20183	322	1474427	13.6123445627785	KEKSKSPSPRLTED.POKGLISPPASPPV.PWKGLISPPASPPV.GQKVGSLTPPSSPKT.QNKFRLDSPRYDSL.DDKLNKSPPLVKAC.YHKKPGTTPPSALP.EDKLI
9 ...R. [ST]PP...	622.281423513503	76	20123	447	1474105	12.4549434845944	NKRRRSTPPEEQE.RRRRSRTPPVTRRR.RRREAPGSPPLSPRG.IKRRREVSPPGARTR.EERIVQSSPPHSGED.RLRGRKSPSPYARR.RSRGAASPPRELTE.V
10 ... [ST]PP...	615.305311137178	635	20047	7544	1473658	6.18755911677547	NRAPRRATPAHPPP.KETDPVKSPLPEHQ.PQSTPLSPPLTPK.SSGGSTSPPLSPAL.MASPPESDGF.VAAGTLSPPGEEA.KKFELLTPPLSPSR.VGPH

Showing 1 to 10 of 132 entries

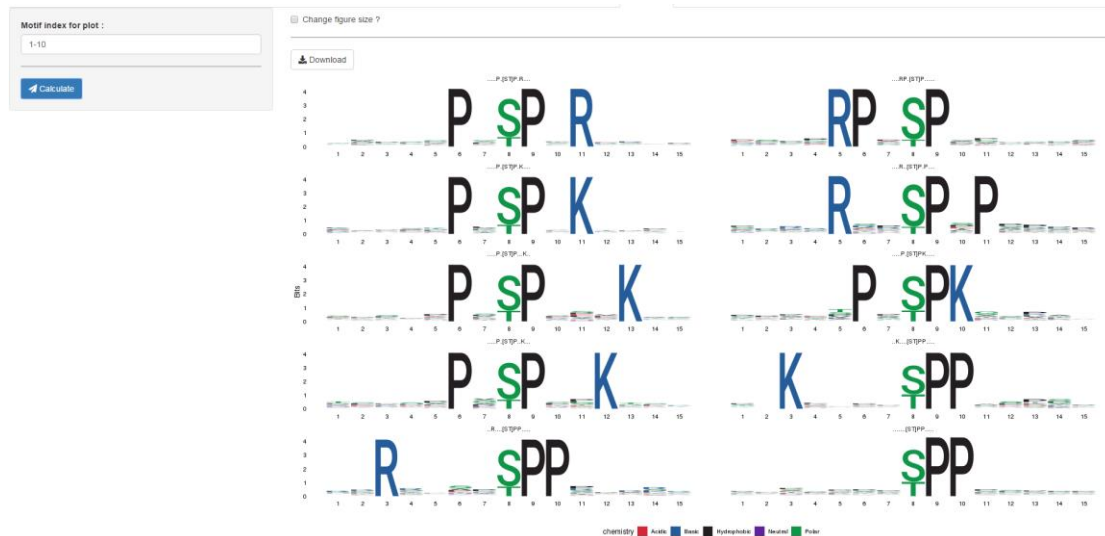
Previous 1 2 3 4 5 14 Next

From this results, we can find there are 132 significant motifs and users can click “Download” button to save this results.

Sixth, click “Motif Plot” to get the visualization of these motifs. By default, the first motif will be plotted:



We can change the “Motif index for plot” parameter to plot more motifs:



Seventh, as this example data were obtained from Human samples, here we can continue. Click “Kinase Analysis”. (Warning: if your data are not obtained from Human samples, you need not continue.)

minimum NetworkKIN score: 3

Select Motif class: [ST]ED [ST]PK [ST]SP

Calculate

Results Network Plot

Download

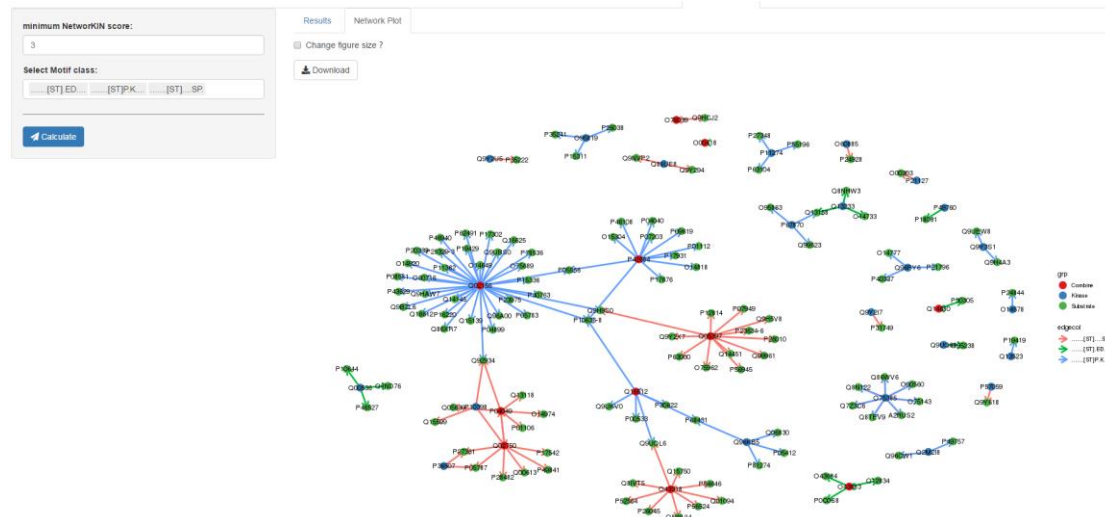
Show 10 entries

	KIN_ACC_ID	SUB_ACC_ID	Motif	KINASE	SUBSTRATE	networkkin_score
1	P21127	O00303	[ST]ED	CDK11B	eIF3-epsilon	300
2	P21127	O00303	[ST]PK	CDK11B	eIF3-epsilon	300
3	P21127	O00303	[ST]SP	CDK11B	eIF3-epsilon	300
4	P21127	O00303	R[ST]P	CDK11B	eIF3-epsilon	300
5	P21127	O00303	[ST]SP	CDK11B	eIF3-epsilon	300
6	P21127	O00303	[ST]SPR	CDK11B	eIF3-epsilon	300
7	P21127	O00303	[ST]PP	CDK11B	eIF3-epsilon	300
8	P21127	O00303	A[ST]P	CDK11B	eIF3-epsilon	300
9	P21127	O00303	[ST]SE	CDK11B	eIF3-epsilon	300
10	P21127	O00303	[ST]PR	CDK11B	eIF3-epsilon	300

Showing 1 to 10 of 9,654 entries

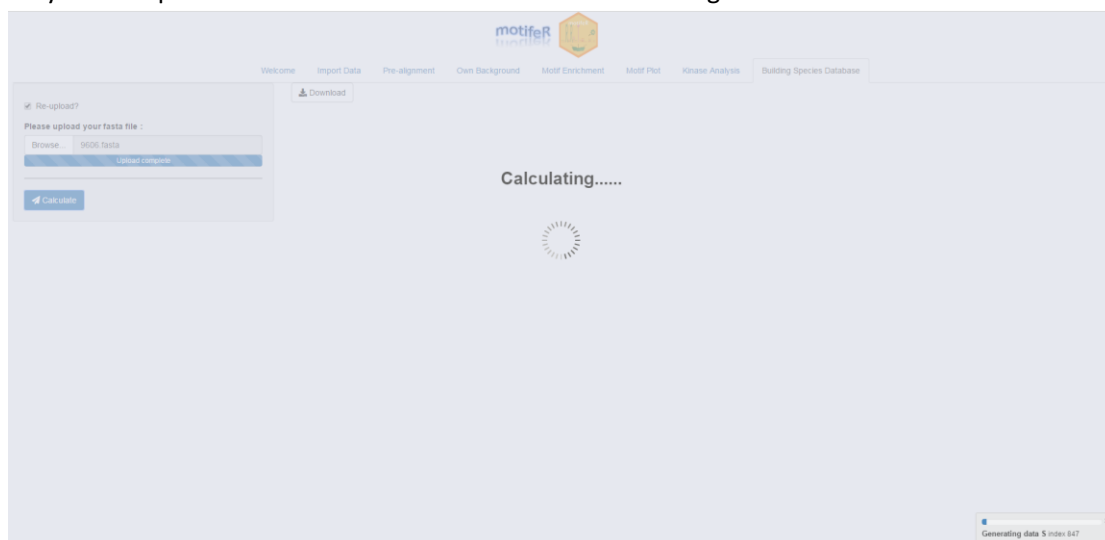
Previous 1 2 3 4 5 ... 966 Next

This part will give users some kinase-substrate information and then plot the network:



Herein we choose three motifs and this software plots their network based on the matched results. Perhaps, these information are useful and helpful for some users, thus we provide them here as an optional function.

Eighth, here we don't want to build the background database and download it, so we can ignore the "Building Species Database". However, if users are interested in the background database, they can re-upload the fasta file and obtain the standard background database as below:



ID	Motif	Center
1	P06575	MSKVFKKTS
2	P06575	SKVPKTTSSNGKLSI
3	P06575	KVFKKTTSSNGKLSIY
4	P06575	TSSNGKLSIYLKGRD
5	P06575	LQVPAESSSPQGPL
6	P06575	QVVPRESSPQGPLT
7	P06575	VVPRESSPQGPLTV
8	P06575	MVTNLPCSVTLQPGP
9	P06575	GDFEVKSCAENPIE
10	P06575	ENPEETYSKRDYRL

This result may be huge based on the species fasta file. If users think this information is also useful for them, just click "Download" button to save it in the local computer. Note: this result

can be regulated based on the parameter in “Import Data” part.

Reference

- (1) Emdal, K. B.; Pedersen, A.-K.; Bekker-Jensen, D. B.; Lundby, A.; Claeys, S.; De Preter, K.; Speleman, F.; Francavilla, C.; Olsen, J. V. *Sci. Signal.* **2018**, *11*, eaap9752.
- (2) Gu, H.; Stokes, M. P.; Silva, J. C. In *Analysis of Post-Translational Modifications and Proteolysis in Neuroscience*; Springer, 2015, pp 1-29.
- (3) Linding, R.; Jensen, L. J.; Pasculescu, A.; Olhovsky, M.; Colwill, K.; Bork, P.; Yaffe, M. B.; Pawson, T. *Nucleic acids research* **2007**, *36*, D695-D699.